Contextual and Hierarchical Classification of Satellite Images Based on Cellular Automata

Moisés Espínola, José A. Piedra-Fernández, Rosa Ayala, Luis Iribarne and James Z. Wang

***Abstract*—Satellite image classification is an important technique used in remote sensing for the computerized analysis and pattern recognition of satellite data, which facilitates the automated interpretation of a large amount of information. Today there exist many types of classification algorithms, such as parallelepiped and minimum distance classifiers, but it is still necessary to improve their performance in terms of accuracy rate. On the other hand, over the last few decades cellular automata have been used in remote sensing to implement processes related to simulations. Although there is little previous research of cellular automata related to satellite image classification, they offer many advantages that can improve the results of classical classification algorithms. This paper discusses the development of a new classification *Algorithm based on Cellular Automata* (ACA) which not only improves the classification accuracy rate in satellite imag es by using contextual techniques, but also offers a hierarchical classification of pixels divided into levels of membership degree to each class and includes a spatial edge detection method of classes in the satellite image.**

***Index Terms*—Image classification, pattern recognition, remote sensing, cellular automata.**

I. Introduction

REMOTE sensing has been used in countless environmental applications with the aim of solving and improving all sorts of problems: soil quality studies, water resources research, meteorology simulations, environmental protection, among others [14]. To solve all of these problems, one must collect and process huge amounts of satellite data, which creates one of the most difficult problems facing remote sensing [8]. Among all the techniques used in remote sensing to help analyst experts interpret the data gathered, classification algorithms are the most useful and promising. These classification algorithms for satellite images group together image pixels into a finite number of classes, which helps to interpret a great deal of data contained in the spectral bands [44]. When applying a classification algorithm to a satellite image, the data obtained by the satellite sensors as digital levels are changed into a categorical scale that is ea sily interpreted by analyst experts. The resulting classified image is a thematic map of the original satellite image, and pixels belonging to the same class share similar spectral characteristics.

The results provided by classification algorithms of satellite images have many political, social and environmental applications. These results are very important for any problem that requires the use of GIS (Geographic Information Systems) for critical information, such as calculating the growth of urban land in cities in a given time interval, monitoring environmental quality after natural disasters, creating GPS (Global Positioning System) maps automatically, preventing natural disasters like fires from spreading, avoiding snow avalanches, evaluating risk management of natural resources, or studying climate change evolution in risk areas. Classification algorithms are the basis of GIS and, for that, it is very important to provide optimal results not only in the classification accuracy rate obtained but also in the amount of useful information offered after the classification process. Therefore, if the classification process is improved by reducing the error margin in the labeling of each pixel and by increasing the amount of useful information offered, the GIS performance could also be improved because a very important feature for the correct behavior of a GIS is the correct identification of each pixel with its corresponding class. There are many classification algorithms of satellite images, and the use of a particular one depends on the analyst expert´s knowledge about the study zone. However, most classical algorithms present three major problems.

First, although there are a large number of satellite image classification algorithms, it is still necessary to improve their performance in terms of accuracy rate [26]. In general, classification algorithms work acceptably well if the spectral properties of the pixels determine the classes well enough or if the images are not noisy. However, if there are some classes with a high degree of heterogeneity grouping pixels with different characteristics that may belong to several classes (uncertain pixels), or the images are altered with a Gaussian impulse-type noise (noisy pixels), the resulting image may have many tiny areas (often a pixel) that are misclassified. All these problems cause a loss of classification accuracy rate. To solve these problems, we can apply contextual post-classification algorithms that use contextual data in addition to spectral data. Several contextual post-classification algorithms use average values or texture description to improve the spe ctral classification. However, such approaches generally require the use of three algorithms: a pre-classification algorithm to eliminate noisy pixels, a classification algorithm, and a post-classification algorithm to improve the classification of uncertain pixels. Grouping the three algorithms into one could improve the classification accuracy rate based on spectral-contextual data.

A second problem in regard to the current classification algorithms of satellite images is that the results are too rigid because each pixel is labeled in its corresponding class regardless of its membership degree. Only the fuzzy classification

Moisés Espínola, José A. Piedra-Fernández, Rosa Ayala and Luis Iribarne are with the Department of Informatics, University of Almería, Spain.

James Z. Wang is with the College of Information Sciences and Technology, The Pennsylvania State University, USA.

algorithms provide such information, not the classical ones. It would be helpful if we use an algorithm that offers a hierarchical classification divided into levels of quality that analyst experts could use to determine which pixels are closer to their classes and which are more distant in the feature space in order to detect the doubtful pixels.

Third, in some studies, it would be desirable to obtain additional information such as edge detection from the classification process. It may also be desirable to locate the satellite image pixels that cause the most problems in the classification process. It is therefore important to customize the classification process to obtain the largest number of possible results.

This paper reports on the development of ACA, which solves the three problems described, providing improved spectral-contextual results divided into levels of membership degree for each class. This allows the analyst experts to have as much information as possible to improve the subsequent interpretation of the results. Thus ACA optimizes the functionality of any GIS that uses spectral-contextual results because it improves the classification results of satellite images. Cellular automata have been widely used by the scientific community to simulate the behavior of complex systems [45] and, in the field of remote sensing, to implement simulation of environmental and weather processes in satellite images [30], but they have been used very little to implement classification algorithms of satellite images [9] [46].

The rest of the paper is structured as follows. Section II describes the classification problems of satellite images in classical algorithms. Section III describes the mathematical description of cellular automata and their applications in remote sensing simulation processes. In Section IV, the paper focuses on the use of cellular automata to classify satellite images (ACA), improving the characteristics and solving the problems of the classical classification algorithms. Section V describes the experimental features and Section VI describes the results of the work. Finally, section VII shows some conclusions and future work.

II. Problems of Classical Classification Algorithms

Classification algorithms of satellite images can be divided into two main categories: supervised and unsupervised algorithms [47]. The use of supervised or unsupervised algorithms in the classification process depends on the analyst expert´s knowledge of the satellite image study area [4] [41]. Despite the large number of classification algorithms of satellite images, all algorithms have limitations that prevent them from being fully reliable in terms of classification accuracy rate [38]. These limitations are increased when some classes have a high degree of heterogeneity because it complicates the grouping of pixels with different characteristics that may belong to several classes (uncertain pixels) or when the images are altered with a Gaussian impulse-type noise (noisy pixels), causing the resulting image to have lots of tiny areas (often a pixel) which are misclassified. In this paper, we focus on solving the classification problems of two classical supervised classification algorithms of satellite images: parallelepiped and minimum distance. Figure 1 shows a graphical representation of the classification process with these two classical supervised algorithms from the viewpoint of feature space and taking into account three different classes in the satellite image.

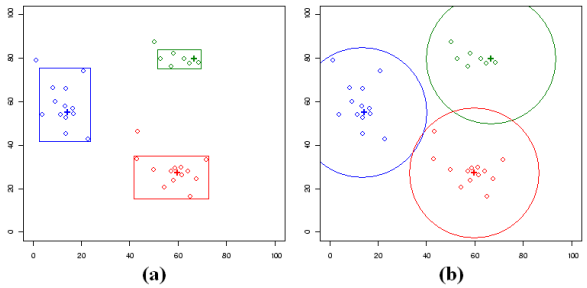


Fig. 1: (a) Parallelepiped (b) Minimum distance.

The parallelepiped algorithm assigns the pixel x to the class A if its *SV* (spectral values) are included in the domain area of that class in the different N bands while considering their centroid value and dispersion range, as shown in the following formula:

$$\overline{CV}_{A,n} - DR_{A,n} \leq SV_{x,n} \leq \overline{CV}_{A,n} + DR_{A,n} \quad (1)$$

where:

- $\overline{CV}_{A,n}$: centroid value of the class A in the band n.
- $DR_{A,n}$: dispersion range of the class A in the band n.
- $SV_{x,n}$: spectral value of pixel x in the band n.
- $n = 1, 2, ..., N$ are the bands of the satellite image.

The parallelepiped algorithm has the disadvantage that some pixels may be unclassified after the process because their digital values are not within a range of any class (as you can see in Figure 1). It can also happen that a pixel is wrongly classified into several classes.

The minimum distance algorithm assigns the pixel x to the class A with which there is less spectral euclidean distance with respect to its centroid taking into account the different N bands involved in the classification process, as shown in the following formula:

$$d_{x,A} = \sqrt{\sum_{n=1}^{N_{bands}} (SV_{x,n} - \overline{CV}_{A,n})^2} \quad (2)$$

where:

- $d_{x,A}$: distance between pixel x and class A.
- $SV_{x,n}$: spectral value of pixel x in the band n.
- $\overline{CV}_{A,n}$: centroid value of the class A in the band n.
- $n = 1, 2, ..., N$ are the bands of the satellite image.

Once all the distances between the pixel and the classes have been calculated, the algorithm assigns the pixel to the nearest class using the following formula:

$$class(x) = \{A | d_{x,A} = minimum\} \quad (3)$$

The minimum distance algorithm has the disadvantage of being prone to commission errors (assigning a pixel to a wrong

class) because the variance of each one of the classes is not considered in the classification process.

If we add some classes with a high degree of heterogeneity and the presence of Gaussian impulse-type noise in the satellite images, the classification accuracy rate obtained by these classical supervised classification algorithms decreases considerably.

Most of these disadvantages can be overcome with the use of cellular automata that use contextual data taking into account not only the pixel's spectral values but also its surrounding pixels. This paper presents some research on how these two supervised classification algorithms (parallelepiped and minimum distance) can be improved using techniques of cellular automata modifying the mathematical formulas shown in this section.

III. Cellular automata

A cellular automaton is a mathematical model which consists of a set of cells usually distributed in a matrix form [28]. In recent years, cellular automata have become a powerful tool applied in remote sensing especially to implement any kind of simulation process in satellite images. From a mathematical point of view a cellular automaton is a set of six components, as shown in the following expression:

$$CA = (d, r, Q, \#, V, f) \tag{4}$$

where:

- $d|d > 0$: spatial dimension of the cellular automaton. The position of each cell is shown by a vector of Z^d. Given $d = 1$ it is a one-dimensional cellular automaton with cells positioned in Z; given $d = 2$ it is a bi-dimensional cellular automaton with cells positioned in ZxZ; given $d = 3$ it is a tri-dimensional cellular automaton with cells positioned in $ZxZxZ$, and so on.
- r: an index that shows the neighborhood dimension; that is, how many neighbors interact with each cell of the cellular automaton.
- Q: a set of states per cell. The set of states is finite, equal for all the cells of the cellular automaton, and it cannot be changed during the cellular automaton application process.
- $\#$: state called quiescent. This state shows inactivity in the cells of the cellular automaton and it is often used as the initial state of the cells.
- V: neighborhood vector which has r different elements from Z^d. The most common types of neighborhoods in a cellular automaton are 4 neighbors (von Neumann neighborhood), 8 neighbors (Moore neighborhood) and 24 neighbors (extended Moore neighborhood). Figure 2 shows the most common types of neighborhoods that we can find in a regular cellular automata. The neighborhood vector V is a subset of Z^d, as shown in the following expression:

$$V \subset (Z^d)^r \tag{5}$$

- f: cellular automaton transition function. It takes as input arguments the states of the current cell and its neighborhood, and returns a new state for the current cell. The transition function f uses a set of rules that specify the changes of the cellular automaton cell states, and it is applied to each cell through a finite number of iterations, as shown in the following expression:

$$f : Q^{r+1} \rightarrow Q \tag{6}$$

$$q_i(t) = f(q_{i-r}(t-1), q_{i-r+1}(t-1), ..., q_{i+r}(t-1))$$

where $q_i(t)$ is the state of the cell i at time t. The changes in cells states of the cellular automaton occur in discrete time form. In each iteration, the whole cells stored in Z^d are checked and the rules are applied through the transition function f to each cell, taking into account the neighborhood V to change its state Q to its corresponding state.

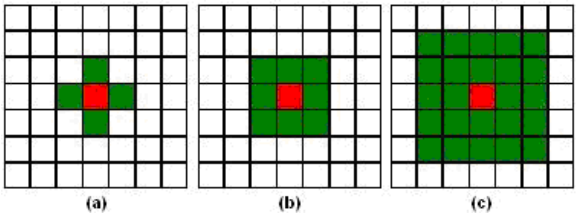


Fig. 2: (a) von Neumann neighborhood (b) Moore neighborhood (c) extended Moore neighborhood.

When we work with satellite images, we usually consider each pixel of the image as a cell of a bi-dimensional cellular automaton ($d = 2$), we normally use the Moore neighborhood ($r = 8$), we assign to each cell a defined set of states to perform the corresponding simulation ($Q = q_1, q_2, ..., q_n$), and we apply the different rules ($R = r_1, r_2, ..., r_m$) in each iteration i through the transition function f.

Therefore, cellular automata have an evolution process because the cells are always changing their states through the different iterations [50]. From this point of view, cellular automata have become a powerful tool to simulate environmental processes in satellite images.

Cellular automata have been widely used for environmental simulations and models like simulating snow-cover dynamics [33] or snow avalanches [3], modeling vegetation systems dynamics [6] or land use dynamics [32], simulating lava flows [49], simulating forest fire spread for the prediction of disasters [27] [40], and modeling species competition and evolution [13].

There is also cellular automata research related to urban and complex social phenomena simulations: a system for the understanding of urban growth [12], an advanced urban mesh generation model in urban design [11], a model of vehicular traffic in a city [2], and a traffic noise control system [43].

As far as medicine is concerned, we can find the creation of a cellular automata-based model for therapies against HIV infection [48], a model that simulates how people are infected by a periodic plague [19], the detecting of Vibrio cholerae by indirect measurement of climate and infectious disease [35], and a model for simulating cancer growth [25].

Cellular automata have been applied not only to implement simulation processes, but also to sort out a wide range of problems of different types: new cryptographic systems based on the cellular automata approach [17] [24], new modular illumination systems [7], new image processing techniques like image enhancement (noise-reduction filters) and edge detection with cellular automata [42], and new texture characterization systems in images [31].

However, in the ambit of satellite image classification there is little previous research [37] [39] related to cellular automata despite the many advantages they can offer, like getting contextual techniques or a hierarchical classification using some properties of cellular automata, such as the use of iterations.

In the next section we propose a new application of cellular automata: a novel supervised classification algorithm based on cellular automata which not only improves the classification accuracy rate but also gives us information on the membership degree of each pixel to its class through a hierarchical classification divided into levels of reliability.

IV. Classification with cellular automata

This paper presents a new methodology for implementing a supervised classification algorithm of satellite images (ACA) [20] [21] [22] [23] that classifies the image pixels based on both spectral and contextual data of each pixel, and it also provides hierarchical classification results divided into hierarchical levels of membership degree to each class. Thus, ACA improves the results obtained by other classical supervised classification algorithms as described in current literature.

ACA is a supervised classification algorithm, implemented with Visual C++ and Erdas Imagine Toolkit, based on a multistate cellular automaton that allows analyst experts to introduce new states and rules to the cellular automaton in order to customize as much as possible the classification process of satellite images. In this sense, there is also research related to cellular automata combined with artificial neural networks that define the rules with a higher degree of objectivity [34] [36]. We must take into account the following associations between a cellular automaton and the basic elements of a generic process of supervised classification of satellite images:

(a) Each pixel of the satellite image corresponds to a particular cell of the cellular automata grid.
(b) Each different class of the supervised classification process is represented by a particular state of the cellular automaton.
(c) The neighborhood of each cell may consist of the 4 surrounding pixels (von Neumann neighborhood), the 8 nearest cells (Moore neighborhood) or even the 24 surrounding pixels (extended Moore neighborhood) in order to customize the final classification process.
(d) The transition function f must correctly classify each pixel of the image based on the features of the current cell and its neighborhood, using mixed spectral and contextual data to improve the results obtained by the classical supervised classification algorithms.

The analyst expert of satellite images must establish the desired behavior of ACA through the states and rules definition of the cellular automaton to adjust to the classification process in order to customize the final classification process. In this paper, we have implemented a version of ACA (ACA v1.0) whose main goals are the following:

- Objective #1. Improve the classification accuracy rate obtained by the classical parallelepiped and minimum distance supervised classification algorithms by means of contextual information to avoid misclassifying the uncertain or noisy pixels. ACA must classify the problematic pixels, taking into account not only their spectral data (ambiguous for uncertain pixels, wrong for noisy pixels) but also their neighbor´s contextual data in order to improve the final classification accuracy rate. With this objective, ACA must merge the following three techniques into one algorithm: pre-classification process (noisy pixels elimination), supervised classification process and post-classification process (uncertain pixel refinement).
- Objective #2. Obtain a hierarchical classification divided into hierarchical layers of membership degree to each class. ACA must classify only those pixels which are within a maximum spectral distance in the featured space with regards to the center of their corresponding class, and such distance must increase in each iteration. Thus ACA will get a hierarchical classification divided into hierarchical layers of membership degree to each class, where the first layers offer more reliability than the last ones because the pixels of the first layers are closer spectrally to their classes and consequently they have a higher membership degree. These results can be very useful for the subsequent interpretation of the results made by the analyst experts. Moreover, this objective helps the first one because the uncertain and noisy pixels, usually further from the center of their classes, must be classified in the last iterations of the cellular automaton and they use, as neighbors, pixels that are very likely to be classified in previous iterations, so they offer more reliability in terms of membership degree to their classes, improving the total accuracy rate of the optimized classification process.
- Objective #3. Get a detailed list of the uncertain and noisy pixels, which can be useful if the classification process fails even when using contextual techniques, and get a list of those pixels that determine the spatial edges of the image classes comparing the class of each pixel with the classes of its neighbors through the cellular automaton rules. In this way it is easy to determine the spatial edges of the image classes because if the class of a pixel is different from some of the classes of its neighbors, this pixel is a spatial edge of its class in the image. Conversely, if the class of a pixel is the same as all the classes of its neighbors, this pixel is called focus to differentiate from edge pixels. Therefore, we can analyze the results from two different perspectives: spectral level (uncertain and noisy pixels) and spatial level (edge and focus pixels). This second choice, spatial level,

is of interest because the spatial edges detection of classes is al so a problem in the scope of remote sensing.

ACA is based on the parallelepiped and minimum distance supervised classifiers. The cellular automaton selects the results of one classification algorithm and subsequently applies the rules of its transition function f. In each iteration of the cellular automaton, the permitted spectral radius distance of search in the feature space (called $threshold$) increases. In the first iterations, ACA classifies the pixels whose distance with respect to their class is very low. In the next iterations, the threshold increases again so that the majority of image pixels are classified. The uncertain and noisy pixels are classified in the last iterations. The transition function f takes into account the inputs in order to apply the cellular automaton rules:

(a) Possible classes offered by the selected supervised algorithm (parallelepiped or minimum distance): classes set of current pixel (maybe one class or several classes for uncertain pixels that are near two or more classes). The spectral classification classes are given by a parallelepiped or minimum distance algorithm modified with cellular automata techniques.
(b) Neighborhood states: states of current pixel neighborhood. This neighborhood can be von Neumann, Moore, or extended Moore type. The cellular automaton neighborhood is chosen by the user before undergoing the classification process in order to customize the results obtained as much as possible.
(c) Cellular automaton iteration: the iteration of the cellular automaton that specifies the current quality level of the classification process.

A. Mathematical Definition of ACA

ACA is based on a multi-state cellular automaton, and each cell of the grid has three independent and different states, namely *class*, *quality* and *type*, which correspond to the three objectives outlined. The first state, *class*, corresponds to the class in which each pixel of the satellite image is classified by using not only its spectral values but also contextual data. This state allows us to improve the classification accuracy rate (objective #1 of ACA v.1.0). The second state, *quality*, indicates the iteration number of the cellular automaton in which each image pixel is classified. This state allows us to obtain the hierarchical classification based on hierarchical levels of membership degree to each class (objective #2 of ACA v.1.0). The third state, *type*, provides additional information and corresponds to the pixel type: uncertain, noisy, edge, or focus. This state allows us to get a detailed list of uncertain, noisy, and class edge pixels (objective #3 of ACA v.1.0). These three states of the cellular automaton can take the following values:

- State #1. [*class*] = $spectralClass$ (defined by the training group) or $emptyClass$ (pixels that have not been classified yet).
- State #2. [*quality*] = $1..numIterations$ (iteration of the cellular automaton that determines the hierarchical layers of membership degree to each class; the first iterations are more reliable in terms of classification accuracy rate than the final ones).
- State #3. [*type*] = $uncertain$ (doubtful pixels), $noisy$ (noisy pixels), $edge$ (spatial border pixels of classes) and $focus$ (pixels that are not uncertain, noisy or edge).

The state #1 [*class*] can take any spectral class previously defined in the training group by the analyst expert or empty class (quiescent state) for pixels not classified yet. So, in the first iteration of the cellular automaton, all cells have empty class in this state. The state #2 [*quality*] takes the value of the cellular automaton iteration where the pixel is classified, a value between 1 and the maximum number of iterations. This information enables us to know in which iteration each pixel is classified, which allows us to calculate the membership degree to its class, as shown in the following formula:

$$md_{x,A} = \frac{iteration_{A,finish} - iteration_{x,classified(A)}}{iteration_{A,finish}} \quad (7)$$

where:

- $md_{x,A}$: membership degree of pixel x to class A.
- $iteration_{A,finish}$: iteration in which all the pixels of class A have been classified.
- $iteration_{x,classified(A)}$: iteration in which pixel x has been classified into class A.

The state #3 [*type*] can take the type of pixel: uncertain, noisy, edge or focus. The cellular automaton rules that achieve the three ACA objectives are the following:

- Rule #1. If the number of $spectralClass$ is 0 because the current pixel has wrong spectral values:
  [*class*][*quality*][*type*] = {*majority class of the neighborhood*, *iteration*, *noisy*}
- Rule #2. If the number of $spectralClass$ is 1 and all the neighborhood class states are $emptyClass$ or the same as current pixel then:
  [*class*][*quality*][*type*] = {*spectralClass*, *iteration*, *focus*}
- Rule #3. If the number of $spectralClass$ is 1 and any neighborhood class state is different from current pixel class then:
  [*class*][*quality*][*type*] = {*spectralClass*, *iteration*, *edge*}
- Rule #4. If the number of $spectralClass$ is bigger than 1 then:
  [*class*][*quality*][*type*] = {*majority class of the neighborhood among the dubious classes*, *iteration*, *uncertain*}

Each rule identifies one of the four types of pixels of the state #3: rule #1 for the noisy pixels, rule #2 for the focus type, rule #3 for the edge type and rule #4 for the uncertain pixels. By means of the rules #1 and #4, ACA improves the classification accuracy rate obtained through contextual techniques (objective #1) and by means of the rules #2 and #3, ACA gets a list of geographical-spatial focus and edges in the satellite image (objective #3). Through the iterative behavior of the cellular automaton, ACA offers a hierarchical classification divided into layers of membership degree for each class (objective #2). Moreover, the first rule is characterized by an error in the classification process, so the potential number of spectral classes is zero. In the second

and third rules, the first condition is the same: the number of spectral classes obtained in the supervised classification (through the modified parallelepiped or minimum dis tance algorithm) is equal to 1. Therefore, we are dealing with pixels which we can definitely classify (focus or edge certain pixels). This condition changes in the fourth rule because uncertain pixels are known for belonging to several classes.

The cellular automaton of ACA can be expressed mathematically, following the nomenclature of the expressions (4)(5)(6), by the following expression:

$$ACA = (d, r, Q, \#, V, f) = \tag{8}$$
$$(2 \times N, \{4|8|24\}, [q_{class}, q_{quality}, q_{type}], [\emptyset, \emptyset, \emptyset], V, f)$$

where:

- $d = 2 \times N$: the spatial dimension of the ACA cellular automaton is 2, so cells are distributed in a matrix form. However, as each pixel of the image has a total of N values stored (one for each satellite image band), we are actually working on a $d = 2 \times N$ dimension.
- $r = \{4|8|24\}$: the neighborhood dimension may consist of the 4, 8 or 24 surrounding pixels in order to customize the final classification process.
- $Q = [q_{class}, q_{quality}, q_{type}]$: the set of states per cell is formed by all sets of values that can take each of the three states, as shown in the following expression:

$$\begin{aligned}[q_{class}, q_{quality}, q_{type}] = \\ [\{emptyClass|spectralClass\}, \\ \{1|...|numIterations\}, \\ \{focus|edge|uncertain|noisy\}]\end{aligned} \tag{9}$$

- $\# = [\emptyset, \emptyset, \emptyset]$: the quiescent state, or initial value, is $\emptyset$ for the three states.
- V: the neighborhood vector is configurable to 4, 8 or 24 surrounding neighbors of each one of the cells in all the working dimensions, as shown in the following expression:

$$V \subset (Z^{2\times N})^{\{4|8|24\}} \tag{10}$$

- f: the transition function applies the four rules to each one of the cells along the different iterations in order to change their states taking into account the neighborhood chosen, as shown in the following expression:

$$f : Q^{\{4|8|24\}+1} \rightarrow Q \tag{11}$$

B. Main, Spectral and Contextual ACA Algorithms

ACA consists of three algorithms: main, spectral and contextual. The main ACA algorithm executes the iterations of the cellular automaton. In each iteration, we first make a spectral classification of all the pixels not classified yet in the satellite image and subsequently we make a contextual classification for pixels that have been classified in the current iteration in order to improve the results provided by the spectral classification. The threshold increases its value in each iteration of the classification process. Table 1 shows the main ACA algorithm that follows the nomenclature of all the formulas and expressions seen before.

Main ACA algorithm $(SV, \overline{CV}, DR, CA.r, CA.nIter, thr)$

Input:
SV: spectral values of all the pixels in all the bands
$\overline{CV}$: centroid value of all the classes in all the bands
DR: dispersion range of all the classes in all the bands
$CA.r$: neighborhood dimension of the cellular automaton
$CA.nIter$: number of iterations of the cellular automaton
thr: threshold for class membership in each iteration
Output:
$CA.Q = \{q_{class}, q_{quality}, q_{type}\}$: set of 3 states per all CA cells

```
01 for i ← 1 to CA.nIter do
02   foreach CA.Q_j|j ∈ {1..numPixels} do
03     if CA.Q_{j,class} = ∅ then
04       CA.Q_j ← spectralACA(SV_j, CV, DR, CA.Q_j, thr);
05       if CA.Q_{j,class} ≠ ∅ then
06         CA.Q_j ← contextualACA(CA.Q_j, CA.r, i);
07       endif
08     endif
09   end
10   thr ← thr + inc;
11 end
12 return CA.Q;
```

TABLE 1: Main ACA algorithm pseudocode.

The spectral ACA algorithm is based on classical supervised algorithms (parallelepiped and minimum distance) improved by means of cellular automata techniques dividing the classification process into several iterations through the threshold increasing. This division entails a hierarchical classification of different layers with a different level of reliability, each one in terms of membership degree to each class. In this part of the algorithm we achieve the objective #2 of ACA. Table 2 shows the spectral ACA algorithm that follows the nomenclature of all the formulas and expressions seen before.

Spectral ACA algorithm $(SV_j, \overline{CV}, DR, CA.Q_j, thr)$

Input:
SV_j: spectral values of pixel j in all the bands
$\overline{CV}$: centroid value of all the classes in all the bands
DR: dispersion range of all the classes in all the bands
$CA.Q_j = \{q_{class}, q_{quality}, q_{type}\}$: set of 3 states of cell j
thr: threshold for class membership in each iteration
Output:
$CA.Q_j = \{q_{class}, q_{quality}, q_{type}\}$: set of 3 states of cell j

```
01 foreach class_A|A ∈ {1..numClasses} do
02   if hierarchicalClass(SV_j, CV, DR, thr, class_A) ∈ thr
03     then CA.Q_{j,class} ← CA.Q_{j,class} + class_A;
04   endif
05 end
06 return CA.Q_j;
```

TABLE 2: Spectral ACA algorithm pseudocode.

In the function $hierarchicalClass()$ ACA uses a classical supervised algorithm modified by means of cellular automata. In the case of the parallelepiped algorithm, ACA uses the following formula modified from the formula (1):

$$\overline{CV}_{A,n} - DR^{thr}_{A,n} \leq SV_{j,n} \leq \overline{CV}_{A,n} + DR^{thr}_{A,n} \tag{12}$$

where the modified argument from the formula (1) is:

– $DR_{A,n}^{thr}$: dispersion range of the class A in the band n at iteration i of the classification process. This value is increased in each iteration taking into account thr.

For the minimum distance algorithm ACA uses the following formula modified from formula (3):

$$CA.Q_{j,class} = \{A, \forall\ A\ |d_{j,A} \leq thr\} \tag{13}$$

ACA modifies the behavior of these classical classification algorithms by means of formulas (12)(13), adjusting their behavior to the number of iterations of the cellular automaton through the parameter thr.

The contextual ACA algorithm applies the four rules of the cellular automaton. The other two objectives are achieved in this part of the algorithm. On the one hand, ACA improves the classification accuracy rate obtained through contextual techniques (objective #1 of ACA) by applying the rules #1 and #4 (for noisy and uncertain pixels). On the other hand, ACA defines the pixel type in the following order: noisy, focus, edge or uncertain (objective #3 of ACA). Table 3 shows the contextual ACA algorithm that follows the nomenclature of all the formulas and expressions seen before.

Contextual ACA algorithm $(CA.Q_j, CA.r, i)$	
Input:	
$CA.Q_j = \{q_{class}, q_{quality}, q_{type}\}$: set of 3 states of cell j	
$CA.r$: neighborhood dimension of the cellular automaton	
i: current iteration of the cellular automaton	
Output:	
$CA.Q_j = \{q_{class}, q_{quality}, q_{type}\}$: set of 3 states of cell j	
01	**if** $size(CA.Q_{j,class}) = \emptyset$ **then**
02	$CA.Q_{j,class} = majorityneighborhood(j, CA.r)$;
03	$CA.Q_{j,type} =$ "$noisy$";
04	**endif**
05	**if** $size(CA.Q_{j,class}) = 1$ **and**
06	$classesEqual(CA.Q_{j,class}, CA.Q_{CA.r,class}) = true$
07	**then** $CA.Q_{j,type} =$ "$focus$";
08	**endif**
09	**if** $size(CA.Q_{j,class}) = 1$ **and**
10	$classesEqual(CA.Q_{j,class}, CA.Q_{CA.r,class}) = false$
11	**then** $CA.Q_{j,type} =$ "$edge$";
12	**endif**
13	**if** $size(CA.Q_{j,class}) > 1$ **then**
14	$CA.Q_{j,class} =$
14	$= majorityneighborhoodClass(CA.Q_j, CA.r)$;
15	$CA.Q_{j,type} =$ "$uncertain$";
16	**endif**
17	$CA.Q_{j,quality} = i$;
18	**return** $CA.Q_j$;

TABLE 3: Contextual ACA algorithm pseudocode.

The function $classesEqual()$ checks whether the class of the current cell is the same as the classes of its neighborhood, the function $majorityneighborhood()$ obtains the majority neighborhood class, and the function $majorityneighborhoodClass()$ obtains the majority neighborhood class among the dubious classes.

C. General ACA architecture

The ACA architecture is composed of (a) the classification with cellular automata (ACA) and (b) the calculation of quality (classification accuracy rate), as shown in Figure 3:

(a) *Classification with cellular automata* (ACA). ACA has eight parameters as input arguments and produces a single output: the classified image. Of the eight input parameters, two are related to the original image: the image loading function, that loads the satellite image (*image.img*= SV), and the class samples loading function, that loads the samples of each class selected by the analyst experts (*samples.sig*=$\{\overline{CV}, DR\}$). With these two components, ACA is prepared to make a supervised classification, although it is based on the results previously obtained by a modified classical supervised classification algorithm (parallelepiped or minimum distance) through the threshold (thr). ACA changes the behavior of these classifiers by using a cellular automaton, adding the following parameters to the supervised classification process: states ($CA.Q$), rules, neighborhood ($CA.r$) and iterations ($CA.nIter$). The user can configure the neighborhood and iteration parameters of the cellular automaton before carrying out the classification process in order to adjust the cellular automata behavior to the study area and customize the final results of the classification process.

(b) *Calculation of quality*. This algorithm takes two parameters as input arguments: the classified image through ACA (*classified image*=$CA.Q$) and the classified image through expert field work (*expert classified image*). As a result, this algorithm produces the confusion matrix between these two images, it shows an index of the accuracy rate in the cellular automaton classification process, and it provides a list of wrongly classified pixels that relates the class to which it really belongs (expert field work) to the class where it has been classified (ACA classification results).

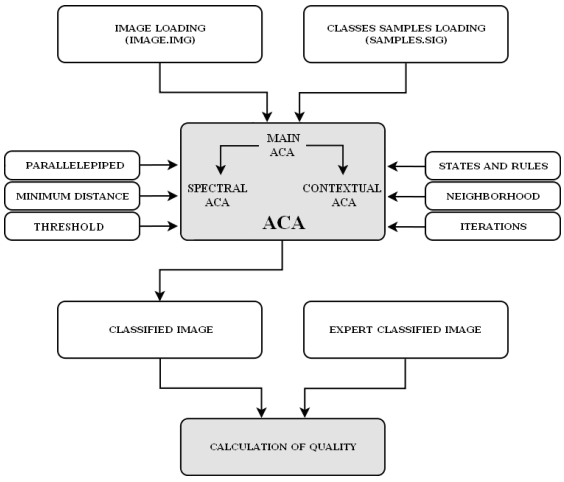


Fig. 3: General ACA architecture.

D. ACA improvements in classical algorithms

In Section II of this paper we discussed the most common classification problems that exist in the parallelepiped and minimum distance supervised algorithms. The advantages and

disadvantages of these classical classification algorithms are summarized in Table 4.

Algorithm	Advantages	Disadvantages
Parallelepiped	Considers dispersion Fast execution	Pixels not classified Pixels in several classes
Minimum distance	All pixels classified Fast execution	Prone to commission errors Does not consider variance

TABLE 4: Advantages and disadvantages of parallelepiped and minimum distance.

By using cellular automata, all these disadvantages are eliminated. In the parallelepiped algorithm all the unclassified pixels disappear because the dispersion range increases at each iteration of the cellular automaton until all image pixels are classified, and the uncertain pixels disappear because it uses contextual information like neighborhood of each pixel if it can belong to several classes. Furthermore, in the case of the minimum distance algorithm, the uncertain pixels misclassified by commission errors are greatly reduced with the contextual classification, and thus the variance is deemed to take into account the neighboring pixels. Figure 4 shows a graphical representation of the classification process of ACA parallelepiped and ACA minimum distance algorithms in 3 iterations.

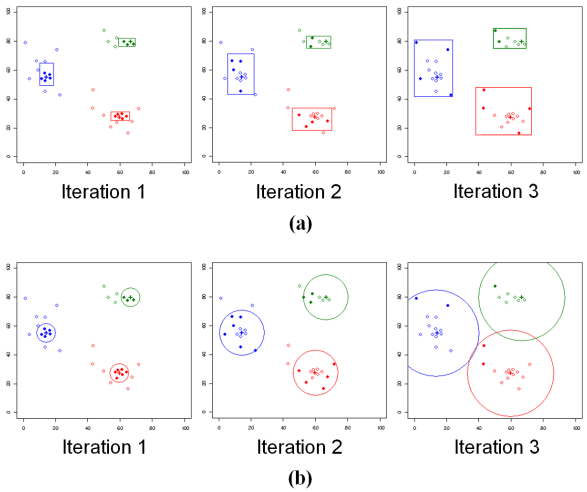


Fig. 4: (a) ACA parallelepiped (b) ACA minimum distance.

V. Experimental features

ACA was tested under the framework of the Soleres Project: "a spatiotemporal environmental management information system based on neural networks, agents and software components". The experiments were carried out on three 7-band multi-spectral Landsat TM satellite images for Níjar, west El Ejido and east El Ejido, three regions in Almería (southeast Spain). Figure 5 shows satellite images of Níjar, west El Ejido and east El Ejido (Bands 3, 2, 1) with the 400×400 window (total of 160000 pixels) used to test ACA performance, at a 30×30 m spatial resolution.

Vegetation and soil in southeast Spain are extremely diverse, complicating verification of any classification algorithm, as

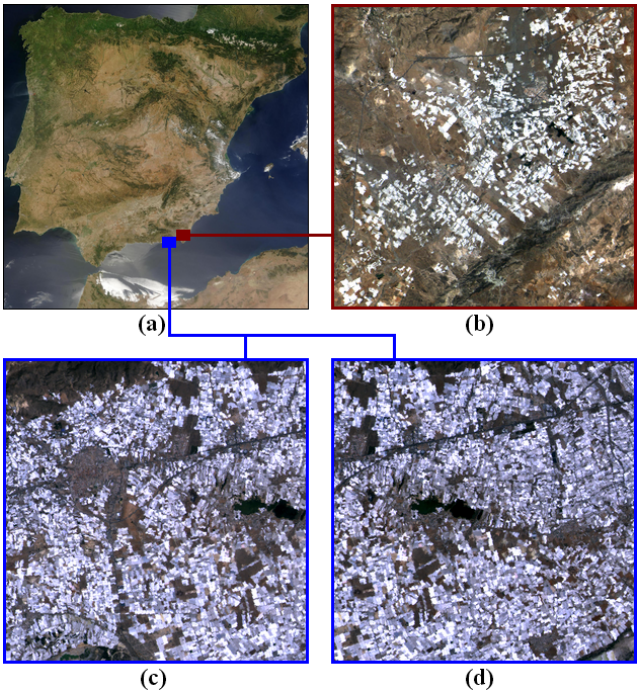
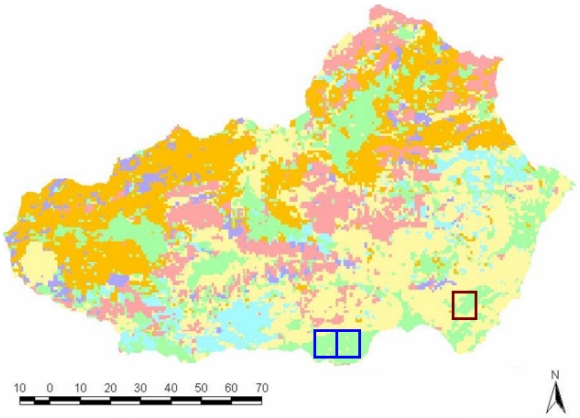


Fig. 5: (a) Areas of study in Almería, Spain (April 2003) (b) Níjar (c) West El Ejido (d) East El Ejido .

shown in previous Soleres Project experiments [5]. Figure 6 shows the six main vegetation areas in southeast Spain [15] [16] with the areas of study marked by the squares.



1		Sparse scrub grass, rock or soil and crop waste areas
2		Woody crops in rainfed, evidence of abandonment
3		Intensive agriculture: greenhouses, crop irrigation, constructed and altered zones
4		Conifer plantations, scrub and grass
5		Mediterranean mountains, meadows and pastures
6		Olivar and upland crops

Fig. 6: Vegetation areas of the southeast of Spain. Níjar (red square) and El Ejido (blue squares).

Specifically, the Níjar and El Ejido regions are characterized by vegetation areas 1 and 3, and can be classified into 6 classes, some of them highly diverse. Above all, El Ejido is characterized by the presence of an enormous number of

greenhouses. There are so many greenhouses in El Ejido that they can be seen from the ISS (International Space Station) in outer space, even in satellite images with low spatial resolution. Níjar also has greenhouses, but fewer of them. Pixel classification in the 'greenhouse' class is very complicated, because this class is so extremely diversified, and feature spaces in its pixels may have a wide spectral range due to the different construction materials used, like plastic or polycarbonate. So classic classification algorithms fail with this type of pixels [1] [10]. These regions are also characterized by the presence of the 'built-up and disturbed areas' class, which is also very diverse, beca use it groups pixels with different characteristics, such as buildings and soil. The classic classification algorithms therefore fail in this class as well. Níjar and El Ejido are also characterized by moderately diversified classes, like 'continuous pasture', 'scattered scrub with rock', and 'wetlands and open water'. Finally, Níjar and El Ejido are characterized by the presence of the 'paved road' class, with low-level diversity, as shown in Table 5.

Class	Area	Description	Heterogeneity
C1	3	Wetlands and open water	Medium
C2	3	Greenhouses	**Very high**
C3	3	Continuous pasture	Medium
C4	3	Built-up and disturbed areas	**High**
C5	1	Scattered scrub with rock	
C6	3	Paved road	Low

TABLE 5: Classes description of Níjar and El Ejido.

Summarizing, Níjar and El Ejido are characterized by one very highly diverse class, one highly diverse class, three moderately diverse classes and one less diverse class. Many classes are spectrally too close, as shown in Figure 7, with the following colour code: C1 (blue), C2 (white), C3 (green), C4 (light brown), C5 (dark brown) and C6 (gray).

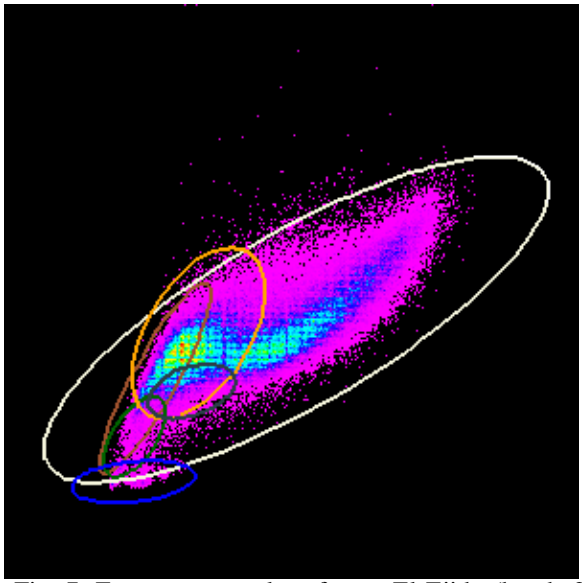


Fig. 7: Feature space plot of west El Ejido (bands 2-7).

However, the satellite images of El Ejido are more difficult to classify than Níjar, because there are more greenhouses in El Ejido than in Níjar, and the greenhouses are the class with the most diversity. In addition, east El Ejido has a higher concentration of greenhouses than west El Ejido. Specifically, Níjar has 18% greenhouses, west El Ejido has 55% and east El Ejido 71%. To further complicate classification in El Ejido, 1% and 2% noise was artificially added to west and east El Ejido satellite images respectively. The noise was random and entered before ACA was applied. These percentages were chosen because when the image pixels are noisy, the percentage of this type of pixels is usually very low, so there was no point in making the percentage any higher. However, ACA greatly improves this type of pixels, because the cellular automaton rules take the average of the noisy pixels neighbor into account, and almost always classifies it correctly. Therefor e, as noise increases, the ACA improvement percentage is higher, because classic algorithms always make many mistakes in labeling this type of pixels.

With these three images, we established three levels of classification difficulty: Níjar (low), west El Ejido (medium) and east El Ejido (high), in order to compare ACA performance with different levels of classification complexity. Table 6 summarizes the characteristics of the three satellite images.

Satellite image	Noise	C2	Classification complexity
Níjar	0%	18%	**Low**
West El Ejido	1%	55%	**Medium**
East El Ejido	2%	71%	**High**

TABLE 6: Characteristics of Níjar, west and east El Ejido.

VI. Results

The application of ACA to these three satellite images aims to achieve the objectives described in section IV: improve the classification accuracy rate using contextual techniques (objective #1), construct a hierarchical classification based on degree of class membership (objective #2), and select edge, uncertain, and noisy pixels (objective #3).

A. Objective #1: improve the classification accuracy rate using contextual techniques

As discussed in the section above, the highly diversified features of the study areas complicate their classification, and therefore, classification algorithm accuracy rates are disappointing. However, ACA accuracy is better than other supervised classification algorithm rates, since the surrounding pixels are used as the neighborhood of the transition function f in the classification of each pixel. These image pixel relationships provide an optimized contextual classification that improves the final results in uncertain and noisy pixels.

ACA improves the performance of classical parallelepiped and minimum distance algorithms. In the first iterations of the cellular automaton, there are only well classified pixels in the confusion matrix, because the threshold is very low, so most of the pixels classified in these first iterations are in the training set. If the algorithm continues to run, there is a point

at which all the pixels in low and moderate diversity classes have already been classified (C1, C3, C5 and C6), and from that point on, only some pixels in the high diversity classes remain unclassified (C2 and C4). This occurs at approximately iteration 40 during classification of the three satellite images. During the following 60 iterations, ACA refines the classification to further improve the results, correctly classifying the most difficult pixels: uncertain pixels in classes C2 and C4 and noisy pixels. By iteration 100, all image pixels have been classified. Therefore ACA has solved one of the main pr oblems of classic parallelepiped algorithm, since by using cellular automata there are no unclassified pixels at the end, because the threshold increases with each iteration. Moreover, ACA has grouped the pre-classification (noise reduction), classification, and post-classification (uncertain pixel refinement) processes. The use of cellular automata not only improves the performance of the classic algorithms, but also its classification accuracy rate: the ACA parallelepiped algorithm improves the accuracy rate of the classical parallelepiped algorithm (by 4.82% in Níjar, 8.10% in west El Ejido and 15.73% in east El Ejido) and the ACA minimum distance algorithm improves the accuracy rate of the classical minimum distance algorithm (by 3.31% in Níjar, 3.92% in west El Ejido and 9.71% in east El Ejido).

In a comparison of the ACA accuracy rate with five other widely used classification algorithms (C4.5, multilayer perceptron, Naive Bayes, k-NN and RBF network), ACA is observed to provide better results. A field image of each study area made by expert ecologists was used to calculate the accuracy rate of algorithms. All algorithms were evaluated using 10-fold cross-validation. Table 7 shows the accuracy rates of these classification algorithms for the classification of the three images and their computational complexity.

Comparison of the satellite image of Níjar (low classification complexity) shows that the C4.5 and multilayer perceptron algorithms outperform the ACA parallelepiped algorithm accuracy rate by 5.52% and 1.47%, respectively. However, although the C4.5 and multilayer perceptron algorithms are more accurate than ACA in satellite images with low classification complexity, they do not offer the additional information that the ACA does: hierarchical classification and edge detection. Thus, even with these accuracy rates, it is still better to use ACA instead of C4.5 and multilayer perceptron if more information is desired from the classification process. The ACA parallelepiped algorithm surpasses the accuracy rate of ACA minimum distance algorithm by 0.79%, and the ACA minimum distance algorithm is 0.49% more accurate than Naive Bayes, 2.41% better than k-NN with k=3, and 3.22% better than RBF network. This satellite image is the one with the lowest ACA scores, and therefore , two algorithms have better accuracy rates.

Comparison of the satellite image of west El Ejido (medium classification complexity) shows that only the C4.5 algorithm outperforms the accuracy rate of ACA parallelepiped algorithm by 3.39%. The ACA parallelepiped algorithm improves the accuracy rate of multilayer perceptron by 1.72%, but the accuracy rate of multilayer perceptron is better than the accuracy rate of ACA minimum distance algorithm. However, the accuracy rate of ACA minimum distance algorithm is 1.53% better than the Naive Bayes, 3.08% better than k-NN with k=3, and 3.79% better than the RBF network. This satellite image shows how the ACA accuracy rate is better than the other algorithms, except C4.5.

Comparison of the east El Ejido satellite image (high classification complexity) shows that the ACA parallelepiped algorithm outperforms all the other algorithms. The accuracy rate of ACA parallelepiped algorithm is 1.91% better than C4.5, 5.28% better than multilayer perceptron, 9.44% better than Naive Bayes, 10.81% better than k-NN with k=3, and 12.53% better than RBF network. This satellite image is the one in which ACA scores highest, because ACA provides better results with satellite images that have a large number of uncertain pixels belonging to highly diverse classes, images in which the other classification algorithms often fail.

Regarding the computational cost, ACA is moderately complex, because it is not as fast as k-NN, Naive Bayes or multilayer perceptron, but neither is it as slow as C4.5 and RBF network.

Figure 8 shows how the accuracy rates of the classification algorithms used in the three satellite images evolve. It may be seen that the ACA parallelepiped and minimum distance algorithm accuracy rates perform better than the others, since the ACA accuracy rate decreases more slowly as classification complexity increases.

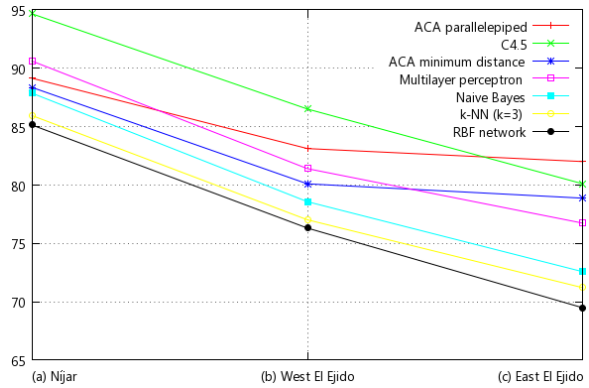


Fig. 8: Accuracy rate of classification algorithms: (a) Níjar (b) West El Ejido (c) East El Ejido.

B. Objective #2: construct a hierarchical classification based on degree of class membership

ACA produces a hierarchical classification divided into layers by degree of membership in each class based on spectral proximity of the pixels to their class within the feature space. In each iteration of the cellular automaton, ACA classifies image pixels farthest from the center of their corresponding class as indicated by the threshold. The range of membership distance permitted in each class increases with every iteration. The pixels spectrally closest to the classes are classified in the first iteration. The furthest pixels are classified in the following iterations. Therefore, pixels classified in a given iteration are more reliable in terms of classification accuracy rate than those classified in the following iteration and so on. Finally,

Algorithm	Níjar	West El Ejido	East El Ejido	Computational complexity
ACA parallelepiped	89.15%	83.12%	**82.01%**	$O(n \cdot i + log\ n)$
C4.5	**94.67%**	**86.51%**	80.10%	$O(m \cdot n^2)$
ACA minimum distance	88.36%	80.10%	78.87%	$O(n \cdot i + log\ n)$
Multilayer perceptron	**90.62%**	81.40%	76.73%	$O(n^2)$
Naive Bayes	87.87%	78.57%	72.57%	$O(m \cdot n)$
k-NN (k=3)	85.95%	77.02%	71.20%	$O(n \cdot log\ n)$
RBF network	85.14%	76.31%	69.48%	$O(n^3)$

TABLE 7: Accuracy rate and computational complexity of classification algorithms (where: n is the number of training instances, m is the number of attributes and i is the number of CA iterations).

the uncertain pixels are classified by contextual classification techniques, based on neighboring pixels correctly classified in previous iterations. Figure 9 shows classification in the three satellite images after 100 iterations of the ACA minimum distance algorithm with the following colour code: C1 (blue), C2 (white), C3 (green), C4 (light brown), C5 (dark brown), and C6 (gray). The ACA parallelepiped algorithms are similar.

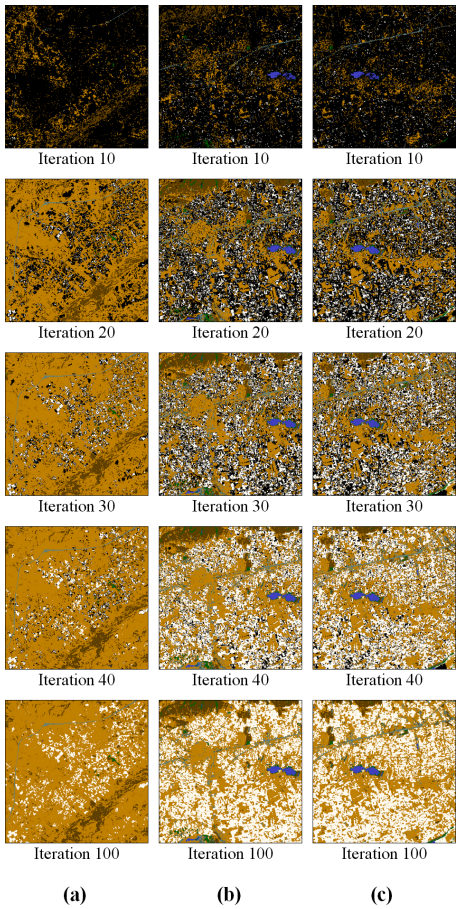


Fig. 9: ACA minimum distance classification process: (a) Níjar (b) West El Ejido (c) East El Ejido.

Most pixels are observed to be classified in the first 40 iterations, and the following refine the classification and optimize results using contextual techniques. As shown, in iteration 40, the satellite images of El Ejido have a higher percentage of unclassified pixels than the satellite image of Níjar, because the number of greenhouses pixels is much larger. In the remaining 60 iterations, the classification is improved using contextual techniques to classify the uncertain and noisy pixels. These results display where the most difficult pixels (uncertain and noisy) in the satellite image are.

C. Objective #3: edge, uncertain and noisy pixels detection

ACA assigns not only the class (state #1) and iteration number to each pixel when it is classified (state #2), but also the pixel type: uncertain, noisy, focus, or edge (state #3). This third state provides the expert analysts with more information from the classification process, enabling them to arrive at specific conclusions for each class.

Analysis of noisy pixels is important, because they contribute to the improvement of the classification accuracy rate. The noise entered in west (1% noise) and east El Ejido (2% noise) is random, and therefore, the noisy pixels are randomly distributed throughout the satellite images. The number of random pixels in each class is weighted by the sizes of the classes and consequently, the classes that have a higher percentage of total pixels in each image also have a higher percentage of noisy pixels. The noise introduced in each class in east El Ejido is not always double the noise in each class in west El Ejido because the two images have different numbers of pixels in each class. It is important to highlight the percentage of well classified noisy pixels, because they raise the classification accuracy rate. Table 8 shows the well classified noisy pixels in west and east El Ejido found with the ACA parallelepiped (ACA P) and ACA minimum distance (ACA MD) by clas s.

Class	West El Ejido		East El Ejido	
#	ACA P	ACA MD	ACA P	ACA MD
C1	0.007%	0.007%	0.013%	0.013%
C2	**0.454%**	**0.418%**	**1.036%**	**0,961%**
C3	0.008%	0.008%	0.006%	0.006%
C4	**0.310%**	**0.314%**	**0.535%**	**0.584%**
C5	0.048%	0.052%	0.036%	0.039%
C6	0.004%	0.004%	0.009%	0.009%
Total	**0.831%**	**0.803%**	**1.635%**	**1.612%**

TABLE 8: Correctly classified noisy pixels (West and East El Ejido).

The classes with the highest percentages of well classified noisy pixels are C2 and C4, because there are more pixels in them than in the rest of the classes in the satellite image. In west El Ejido, the ACA parallelepiped algorithm correctly classified 0.831% of the noisy pixels and the ACA minimum distance algorithm correctly classified 0.803% of noisy pixels out of a total 1%. In east El Ejido, the ACA parallelepiped

algorithm correctly classified 1.635% noisy pixels and the ACA minimum distance algorithm correctly classified 1.612% noisy pixels out of a total 2%. In both satellite images the percentage of well classified noisy pixels is quite high.

Analysis of uncertain pixels allows the expert analysts to specify the contribution of each class to the improvement in the classification accuracy rate, which is based primarily on refining classification of this type of pixels. The uncertain pixels may be either misclassified or well classified. The misclassified uncertain pixels in each class are equal to the sum of all values outside the main diagonal in the corresponding row of the confusion matrix. The misclassified uncertain pixels are subtracted from the total number of uncertain pixels to find the number of well classified uncertain pixels in each class. These pixels show the improvement of ACA over classical algorithms. Table 9 shows the uncertain pixels in Níjar, west El Ejido and east El Ejido properly classified by the ACA parallelepiped and ACA minimum distance by class.

C2 and C4 are again the classes with the highest percentages of well classified uncertain pixels, but unlike the previous case, it is not because there are more such pixels in the satellite image, but because these two classes are the most diversified. In the classes with low and moderate diversity, the uncertain pixels often appear at the space edges, and so their neighbors may be of any of several different classes and ACA fails. However, in the more diversified classes, uncertain pixels may appear in places where they are surrounded by other pixels of the same class, and therefore ACA classifies them successfully. Classes C2 and C4 coincide precisely with the two classes we had previously established as highly diverse and a great many iterations are required to classify all their pixels. In fact, the results of objective #1 showed that the last 60 ACA parallelepiped iterations were aimed solely at improving the classification of these two classes, whose pixels are furt hest away from the centers of their spectral classes. The other classes have a very low percentage of well classified uncertain pixels (below 1%), except C1 in west El Ejido and east El Ejido, so ACA barely improved the classification accuracy rate in these classes. Therefore, ACA improved the classification accuracy rate mainly due to uncertain pixel refinement in classes C2 and C4, which are highly diversified and classic classification algorithms fail. Although some percentages in Níjar are lower than in the other images, its accuracy rate is higher, because the satellite image has fewer uncertain pixels.

The analysis of focus and edge pixels allows the expert analysts to establish the spatial distribution of the classes throughout the satellite image. Table 10 shows the edge and focus pixels in Níjar and west and east El Ejido found with ACA parallelepiped and ACA minimum distance by class.

Class C1 has the highest percentage of edge pixels in Níjar and the lowest percentage in west El Ejido and east El Ejido. This is because in Níjar little ponds of water occupy a single pixel surrounded by pixels of other classes. In west El Ejido and east El Ejido, the regions of water are larger than in Níjar and spatially distributed in the ellipse (wetlands and open water). Class C3 has a high percentage of edge pixels in all three satellite images, since the vegetation in the study areas is distributed in many small groups of pixels, and therefore, most of these pixels are borders (continuous pasture). Class C6 has the highest percentage of edge pixels of almost all of the satellite images, which indicates that the pixels in this class are rather elongated (paved road) and spatially distributed in the images.

VII. Conclusions and future work

In conclusion, the results with ACA are very satisfying from several points of view. Cellular automata have been validated as a technique that improves the results of satellite image classification, successfully achieving the three objectives set.

First, the ACA classification accuracy rate was very high in satellite images with low, medium, and high classification complexity. Compared to other classifiers (C4.5, multilayer perceptron, Naive Bayes, k-NN and RBF network), the ACA accuracy rate was higher than most on low and medium complexity images and better than any of the rest on highly complicated satellite images. Furthermore, the more complicated classification is, the greater the improvement in the success rate compared to the other algorithms. ACA improves classification precisely in satellite images with great difficulties and very diversified classes with many uncertain and noisy pixels. Therefore, although ACA can be used for any type of satellite image, its performance is especially good in the more complicated ones. ACA achieves this objective by using the contextual information provided by the cellular automaton neighborhood, allowing it to make a mixed spectral-contextual classification. This process was improved because ACA could make use of a neighborhood of uncertain and noisy pixels classified in previous iterations, and therefore, pixels closer to spectral classes. This process enables classification of uncertain and noisy pixels to a higher degree of certainty. ACA not only improved the accuracy rate of classic parallelepiped and minimum distance algorithms, but also performance of the classic algorithms by solving the disadvantages described in section II of this paper. In the ACA parallelepiped algorithm classification results, there are no unclassified pixels, because the threshold is raised in each iteration until all pixels have been classified, and no pixels are identified in more than one class, because uncertain pixels are refined by contextual information. Moreover, in the ACA minimum distance algorithm classification results, commission errors are reduced by using contextual information, and although the variance is not taken into account in the ACA classification process, it is simulated for uncertain and noisy classes by the neighborhood functions. Thus ACA optimizes the general functionality of any GIS that uses it because it improves the classification results of satellite images. A very important feature in the proper behavior of a GIS is correct identification of each pixel with its corresponding class. Therefore, with respect to improvement in the accuracy rate, the results with low and moderate complexity satellite images are very good and with highly complex satellite images they are excellent.

Second, ACA hierarchical classification is based on degree of class membership layers, each layer corresponding to a cellular automaton iteration. Throughout ACA execution, the class spectral radius increases with each iteration. The first

Class	Níjar		West El Ejido		East El Ejido	
#	ACA parallelepiped	ACA minimum distance	ACA parallelepiped	ACA minimum distance	ACA parallelepiped	ACA minimum distance
C1	0.000%	0.000%	17.903%	2.848%	17.841%	2.794%
C2	**31.034%**	**17.612%**	**44.191%**	**23.773%**	**43.561%**	**22.393%**
C3	0.397%	0.373%	0.510%	0.400%	0.488%	0.376%
C4	**68.043%**	**23.474%**	**31.802%**	**11.913%**	**31.551%**	**11.676%**
C5	0.046%	0.224%	0.071%	0.309%	0.055%	0.249%
C6	0.002%	0.002%	0.000%	0.000%	0.001%	0.001%

TABLE 9: Well classified uncertain pixels of Níjar, west El Ejido, and east El Ejido.

Class	Níjar				West El Ejido				East El Ejido			
	ACA parallel.		ACA minimum dist.		ACA parallel.		ACA minimum dist.		ACA parallel.		ACA minimum dist.	
#	Edge	Focus	Edge	Focus	Edge	Focus	Edge	Focus	Edge	Focus	Edge	Focus
C1	**100.0%**	**0.00%**	**100.0%**	**0.00%**	**45.02%**	**54.98%**	**44.97%**	**55.03%**	**45.86%**	**54.14%**	**45.80%**	**54.20%**
C2	79.96%	20.04%	79.78%	20.22%	59.99%	40.01%	59.76%	40.24%	60.21%	39.79%	60.07%	39.93%
C3	**93.16%**	**6.84%**	**93.21%**	**6.79%**	**88.03%**	**11.97%**	**88.05%**	**11.95%**	**97.19%**	**2,81%**	**97.23%**	**2,77%**
C4	50.92%	49.08%	89.97%	10.03%	65.78%	34.22%	66.01%	33.99%	73.67%	26.33%	73.75%	26.25%
C5	82.06%	17.94%	82.12%	17.88%	61.00%	39.00%	61.01%	38.99%	63.47%	36.53%	63.49%	36.51%
C6	**98.94%**	**1.06%**	**98.99%**	**1.01%**	**96.01%**	**3.99%**	**96.04%**	**3.96%**	**96.73%**	**3.27%**	**96.75%**	**3.25%**

TABLE 10: Edge and focus pixels of Níjar, west El Ejido, and east El Ejido.

iterations provide the most reliable layers, because these layers contain the pixels spectrally closest to the centers of their classes, some even correspond to the set of samples selected by the expert analyst. As the cellular automaton iterations are run, spectral distance increases, and as a result, the classified pixels are less reliable in terms of classification accuracy, because they are further from the center of their corresponding classes. Finally, the uncertain pixels are classified in the last iterations with the assistance of contextual information. They require contextual information to count the classes of the neighboring pixels, already classified in previous iterations and therefore, more reliable in terms of classification accuracy, and thus improve the final result. Using this method, ACA simulates a pseudo-fuzzy classification in which the degree of class membership of each pixel is indicated by the iteration in which it has been classified, so it is very high in the first iteration, but is considerably lower in the last iterations. Therefore, presenting the results of the classification process by iterations provides the expert analyst with extra information about each pixel, who can determine the degree of class membership of each pixel. We can also determine how diversified the classes are, since all pixels in less diversified classes are classified in the first 40 iterations, and the uncertain pixels belonging to highly diversified classes are classified in the following 60 iterations. The longer it takes all the uncertain pixels in a class to be classified, the more diversified that class is.

Third, the algorithm also provides spatial edge detection of classes in the satellite image, which can be rather useful in later interpretation and analysis of the results, as well as a list of uncertain and noisy pixels, so experts can detect them easily. Analysis of uncertain and noisy pixels shows the contribution of classes to the improvement in classification accuracy rate, while the analysis of focus and edge pixels determines the spatial distribution of the pixels for each class in the image.

As a final advantage of using cellular automata for satellite image classification, we conclude that they allow expert analysts to configure a personalized classification of each particular satellite image in each specific study area, by modifying only some of the properties of cellular automata: neighborhood, number of iterations, rules and states. It is therefore a novel application of cellular automata which opens the way to additional research.

As future work we would like to develop a new version of ACA with a different configuration of states and rules of the cellular automaton to customize the classification process (i.e. develop a fuzzy ACA classification algorithm adding fuzzy rules and states to the cellular automata). It would also be of interest to create a new version of ACA to determine the number of classes of the image directly without knowing them before the classification process. Finally, we would like to add a new level of classification to ACA: textural classification (based on textures). Thus, we would have two different levels of classification: pixel level (spectral and contextual information) and regional level (textural data).

ACKNOWLEDGMENT

This work was funded by the EU ERDF and the Spanish Ministry of Economy and Competitiveness (MINECO) under Project TIN2010-15588, and the Andalusian Regional Government (Spain) under Project P10-TIC-6114. This work was also supported by the CEiA3 and CEIMAR consortia. J. Z. Wang has been funded by the National Science Foundation under Grant No. 1027854. We thank the anonymous reviewers for the constructive comments.

REFERENCES

[1] Agüera F., Aguilar M.A., Aguilar F.J., "Detecting greenhouse changes from QuickBird imagery on the Mediterranean coast", *International Journal of Remote Sensing*, 27 (21): 4751-4767, 2006.

[2] Aponte A., Moreno J.A., "Cellular automata and its application to the modelling of vehicular traffic in the city of Caracas", *7th International Conference on Cellular Automata for Research and Industry (ACRI 2006)*, 4173: 502-511, 2006.

[3] Avolio M.V., Errera A., Lupiano V., Mazzanti P., Di Gregorio S., "Development and calibration of a preliminary cellular automata model for snow avalanches", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, 6350: 83-94, 2010.

[4] Ayala R., Menenti M., Girolana D., "Evaluation methodology for classification process of digital images Igarss´02", *IEEE Int. Geoscience and Remote Sensing Symposium and the 24th Canadian Symposium on Remote Sensing*, 3363-3365, 2002.

[5] Ayala R., Becerra A., Flores I.M, Bienvenido J.F, Díaz J.R., "Evaluation of greenhouse covered extensions and required resources with satellite images and GIS. Almería´s case", *Second European Conference of the European Federation for Information Technology in Agriculture, Food and the Environment*, 27-30, 1999.

[6] Balzter H., Braun P., Kühler W., "Cellular automata models for vegetation dynamics", *Ecological Modelling*, 107: 113-125, 1998.

[7] Bandini S., Bonomi A., Vizzari G., "A cellular automata based modular illumination system", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, 6350: 334-344, 2010.

[8] Barret E.C., Curtis L.F., "Introduction to environmental remote sensing", *Cheltenham Stanley Thornes Publishers Ltd*, 1999.

[9] Bi Y., Zhang Y., Chen Y., "Image classification method based on cellular automata transforms", *WCICA 2006*, 10058-10062, 2006.

[10] Carvajal F., Crisanto E., Aguilar F.J., Agüera F., Aguilar M.A., "Greenhouses detection using an artificial neural network with a very high resolution satellite image", *Proceedings of the ISPRS Vienna 2006 Symposium Technical Commission II*, 37-42, 2006.

[11] Castro A., Gómez N., "Self-organizing map and cellular automata combined technique for advanced mesh generation in urban and architectural design", *International Journal of Information Technologies and Knowledge*, 2: 354-360, 2008.

[12] Cheng J., Masser I., "Cellular automata based temporal process understanding of urban growth", *5th International Conference on Cellular Automata for Research and Industry (ACRI 2002)*, 2493: 325-336, 2002.

[13] Chopard B., Lagrava D., "A cellular automata model for species competition and evolution", *7th International Conference on Cellular Automata for Research and Industry (ACRI 2006)*, 4173: 277-286, 2006.

[14] Chuvieco E., Huete A., "Fundamentals of satellite remote sensing", *CRC Press*, 2010.

[15] Cruz M., Espínola M., Iribarne L., Ayala R., Peralta M., Torres J.A., "How can neural networks to speed up ecological regionalization friendly? Replacement of field studies by satellite data using RBFs", *International Conference on Neural Computation (ICNC 2010)*, 2010.

[16] Cruz M., Espínola M., Iribarne L., Ayala R., Peralta M., Torres J.A., "Ecological sectorization process improvement through neural networks: synthesis of vegetation data from satellite images using RBFs", *9th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2010)*, 113-116, 2010.

[17] Das S., Chowdhury D.R., "Generating cryptographically suitable non linear maximum length cellular automata", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, 6350: 241-250, 2010.

[18] Doi T., "Quantum cellular automaton for simulating static magnetic fields", *IEEE Transactions on Magnetics*, 49(5): 1617-1620, 2013.

[19] Dzwinel W., "A cellular automata model of population infected by periodic plague", *6th International Conference on Cellular Automata for Research and Industry (ACRI 2004)*, 3305: 464-473, 2004.

[20] Espínola M., Piedra-Fernández J.A., Ayala R., Iribarne L., Leguizamón S., Menenti M., "ACA multiagent system for satellite image classification", *(PAAMS 2012)*, AISC 157: 93-100, 2012.

[21] Espínola M., Piedra-Fernández J.A., Ayala R., Iribarne L., Leguizamón S., Menenti M., "A hierarchical and contextual algorithm based on cellular automata for satellite image classification", *International Conference on Computational Intelligence and Software Engineering (CiSE 2011)*, 1-4, 2011.

[22] Espínola M., Ayala R., Leguizamón S., Iribarne L., Menenti M., "Cellular automata applied in remote sensing to implement contextual pseudo-fuzzy classification", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, LNCS 6350: 312-321, 2010.

[23] Espínola M., Ayala R., Leguizamón S., Menenti M., "Classification of satellite images using the cellular automata approach", *First World Summit on the Knowledge Society (WSKS 2008)*, CCIS 19: 521-526, 2008.

[24] Fúster-Sabater A., Caballero-Gil P., "Chaotic cellular automata with cryptographic application", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, 6350: 251-260, 2010.

[25] Ghaemi M., Naderi O., Zabihinpour Z., "A novel method for simulating cancer growth", *The Ninth International Conference on Cellular Automata for Research and Industry (ACRI 2010)*, 6350: 142-148, 2010.

[26] Guzman O.F., Gómez I.D., "Fundamentos físicos de teledetección", *Instituto Geográfico Agustin Codazzi IGAC*, 2007.

[27] Karafyllidis I., Thanailakis A., "A model for predicting forest fire spreading using cellular automata", *Ecological Modelling*, 99: 87-97, 1997.

[28] Kari J., "Theory of cellular automata: a survey", *Theoretical Computer Science*, 334: 3-33, 2005.

[29] Karim F., Walus K., "Efficient simulation of correlated dynamics in quantum-dot cellular automata (QCA)", *IEEE Transactions on Nanotechnology*, 13(2): 294-307, 2014.

[30] Kowsuwan N., Kanongchaiyos P., "3D cloud animation using CA based method", *ISPACS 2009*, 387-392, 2009.

[31] Leguizamón S., Espínola M., Ayala R., Iribarne L., Menenti M., "Characterization of texture in images by using a cellular automata approach", *Third World Summit on the Knowledge Society (WSKS 2010)*, CCIS 112(2): 522-533, 2010.

[32] Leguizamón S., "Modelling land features dynamics by using cellular automata techniques", *Proceedings of the ISPR Technical Comission 7 Mid-Term Symposium "From pixels to Processes"*, 497-501, 2006.

[33] Leguizamón S., "Simulation of snow-cover dynamics using the cellular automata approach", *Proceedings of the 8th International Symposium on High Mountain Remote Sensing Cartography*, 87-91, 2005.

[34] Li X., Gar-on Yeh A., "Neural-network-based cellular automata for simulating multiple land use changes using GIS", *International Journal of Geographical Information Science*, 16(4): 323-343, 2002.

[35] Lobitz B., Beck L., Huq A., Woods B., Fuchs G., Faruque A., Colwell R., "Climate and infectious disease: use of remote sensing for detection of Vibrio cholerae by indirect measurement", *Proceedings of the National Academic of Sciences of the USA*, 97(4): 1438-43, 2000.

[36] Mahajan Y., Venkatachalam P., "Neural network based cellular automata model for dynamic spatial modeling in GIS", *Computational Science ant Its Applications (ICCSA 2009)*, LNCS 5592: 341-352, 2009.

[37] Maji P., Shaw C., Ganguly N., Sikdar B., Chaudhuri P., "Theory and application of cellular automata for pattern classification", *Fundamenta Informaticae*, 58(3-4): 321-354, 2003.

[38] Mather P., Tso B., "Classification methods for remotely sensed data", *CRC Press 2 edition*, 2009.

[39] Mojaradi B., Lucas C., Varshosaz M., "Using learning cellular automata for post classification satellite imagery", *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 35(4): 991-995, 2004.

[40] Muzy A., Innocenti E., Aiello A., Santucci J.F., Santonio P.A., Hill D., "Modelling and simulation of ecological propagation processes: application to fire spread", *Environmental Modelling and Software*, 20: 827-842, 2005.

[41] Piedra-Fernández J.A., Cantón-Garbín M., Wang J.Z., "Feature selection in AVHRR ocean satellite images by means of filter methods", *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, 48: 4193-4203, 2010.

[42] Popovici A., Popovici D., "Cellular automata in image processing", *Proceedings of the 15th International Symposium on the Mathematical Theory of Networks and Systems*, 2002.

[43] Quartieri J., Mastorakis N.E., Iannone G., Guarnaccia C., "Cellular automata application to traffic noise control", *Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling and Simulation (ACMOS10)*, 29-31, 2010.

[44] Rees W.G., "Physical principles of remote sensing", *Cambridge University Press 2nd Edition*, 2001.

[45] Riener A., Fullerton M., Maag C., Mark C., Beltran Ruiz C., Minguez Rubio J.J., Zia K., "Modular simulation-based physical and emotional assessment of ambient intelligence in traffic", *IEEE Transactions on Human-Machine Systems*, 44(2): 286-292, 2014.

[46] Sarhan E., Khalifa E., Nabil A.M., "Post classification using cellular automata for Landsat images in developing countries", *ICIIP 2011*, 1-4, 2011.

[47] Schowengerdt R.A., "Techniques for image processing and classification in remote sensing", *Academic Press*, 1985.

[48] Sloot P., Chen F., Boucher C., "Cellular automata model of drug therapy for HIV infection", *5th International Conference on Cellular Automata for Research and Industry (ACRI 2002)*, 2493: 282-293, 2002.

[49] Spataro W., D. Ambrosio D., Rongo R., Trunfio G.A., "An evolutionary approach for modelling lava flows through cellular automata", *6th International Conference on Cellular Automata for Research and Industry (ACRI 2004)*, 3305: 725-734, 2004.

[50] Wolfram S., "A new kind of science", *Wolfram Media Inc.*, 2002.

Moisés Espínola received the B.S. and M.S. degrees in Computer Science from the University of Almería in 2002 and 2005, respectively. In 2007 he joined the Applied Computing Group and he worked in several national and international research projects. He is currently pursuing the Ph.D. degree and his research work is focused on remote sensing, image processing, and cellular automata. He is mainly involved in the classification of satellite images using Algorithms based on Cellular Automata (ACA). Other major research interests include soft computing and computer graphics applied in the field of psychology and education.

Jose A. Piedra-Fernández received a degree in Computer Science from Almería University, Spain, in 2001, and M. S. and Ph.D. degrees from Almería University, Spain, in 2003 and 2005, respectively. He is an Assistant Professor at that University. He was visiting the College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802 (2008-2009). He has designed a hybrid system applied to remote sensing problems. His main research interests include image processing, soft computing, and image retrieval.

Rosa Ayala received the B.S. and M.S. degrees in Computer Science from the University of Granada, and the Ph.D. degree in Computer Science from the University of Almería, Spain. From 1992 to 1998, she worked as a researcher in the Foundation for Agricultural Research in the Province of Almería (FIAPA). She worked in several national and international research projects on distributed simulation and geographic information systems (GIS). She became Associate Professor at the University of Almería in 2000. In 2007 she joined the Applied Computing Group. Her research interests include satellite classification, GIS, and cellular automata.

Luis Iribarne received the B.S. and M.S. degrees in Computer Science from the University of Granada, and the Ph.D. degree in Computer Science from the University of Almería, Spain. From 1991 to 1993, he worked as a Lecturer at the University of Granada, and collaborated as IT Service Analyst at the University School of Almería. Since 1993, he has been a Lecturer in the Advanced College of Engineering at the University of Almería. From 1993 to 1999, he worked in several national and international research projects on distributed simulation and geographic information systems. He has been the coordinator of several projects founded by the Spanish Ministry of Science and Technology. In 2007, he founded the Applied Computing Group. His research interests include software engineering, model-driven engineering, ontology-driven engineering, component-based software development, modeling, and simulation.

James Z. Wang received the bachelor's degree in mathematics and computer science summa cum laude from the University of Minnesota, the MS degree in mathematics and the MS degree in computer science, both from Stanford University, and the PhD degree in medical information sciences from Stanford University. He has been a faculty member at The Pennsylvania State University since 2000 where he is a Professor in the College of Information Sciences and Technology. His main research interests are automatic image tagging, image retrieval, computational aesthetics, and computerized analysis of paintings. He was a visiting professor at the Robotics Institute at Carnegie Mellon University (2007-2008) and a program manager at the National Science Foundation (2011-2012). He has been a recipient of a US National Science Foundation Career award and the endowed PNC Technologies Career Development Professorship.