Federation University ResearchOnline

https://researchonline.federation.edu.au

Copyright Notice

This is the published version of:

See this record in Federation ResearchOnline at:
http://researchonline.federation.edu.au/vital/access/HandleResolver/1959.17/180940

MDPI

Article

Using Vector Agents to Implement an Unsupervised Image Classification Algorithm

Kambiz Borna [1,*], **Antoni B. Moore** [2], **Azadeh Noori Hoshyar** [3] **and Pascal Sirguey** [2]

1 Civil Engineering and Land Surveying Department, Unitec Institute of Technology, Auckland 1025, New Zealand

2 School of Surveying, University of Otago, Dunedin 9054, New Zealand; tony.moore@otago.ac.nz (A.B.M.); pascal.sirguey@otago.ac.nz (P.S.)

3 School of Engineering, IT and Physical Sciences, Federation University Australia, Brisbane, QLD 4000, Australia; a.noorihoshyar@federation.edu.au

* Correspondence: kborna@unitec.ac.nz

Abstract: Unsupervised image classification methods conventionally use the spatial information of pixels to reduce the effect of speckled noise in the classified map. To extract this spatial information, they employ a predefined geometry, i.e., a fixed-size window or segmentation map. However, this coding of geometry lacks the necessary complexity to accurately reflect the spatial connectivity within objects in a scene. Additionally, there is no unique mathematical formula to determine the shape and scale applied to the geometry, being parameters that are usually estimated by expert users. In this paper, a novel geometry-led approach using Vector Agents (VAs) is proposed to address the above drawbacks in unsupervised classification algorithms. Our proposed method has two primary steps: (1) creating reliable training samples and (2) constructing the VA model. In the first step, the method applies the statistical information of a classified image by k-means to select a set of reliable training samples. Then, in the second step, the VAs are trained and constructed to classify the image. The model is tested for classification on three high spatial resolution images. The results show the enhanced capability of the VA model to reduce noise in images that have complex features, e.g., streets, buildings.

Keywords: Vector Agents; unsupervised image classification; dynamic geometry

1. Introduction

In the remote sensing context, the purpose of image classification is to extract meaningful information (land cover categories) from an image [1]. This process is generally performed via a supervised or unsupervised method [2]. In the supervised classification, the algorithm is first trained using ground data and then applied to classify the image pixels. Unsupervised algorithms only use the information that is contained in the image to classify pixels without requiring any training data [3,4]. As the algorithm does not require training samples, it is easy to perform with minimum human intervention and cost [5,6]. The algorithms use the pixel values through a set of statistical rules or cost functions to label pixels in the feature space [7]. The process is typically performed via an iterative mechanism by minimising the spectral distances (similarities) between pixel values and the calculated cluster centers in the feature space, regardless of their locations in the image. This causes a speckled appearance (also known as the salt-and-pepper noise) in which there are isolated pixels or small regions of pixels in the scene.

The above drawback is typically addressed by incorporating the spatial information between pixels into the clustering process [1]. To determine the neighbouring pixels, the algorithms use a fixed-size window or segmented objects [8]. In the fixed-size window structure, the algorithm, e.g., hidden Markov model [1], or textural information [9,10] applies the spectral–spatial information extracted from a fixed-size window that is centred

on each pixel to extract the spatial information and label image pixels. The methods show better classification outcomes in comparison to the conventional clustering algorithms. However, the clustering results are not always adequate because there is no specific rule to determine the optimum size and shape of these local windows [8,11]. As a result, the classified maps may contain significant amounts of misclassifications, especially when there are heterogeneous or complex features in a scene, e.g., roads or buildings. The main assumption for these methods is that objects have similar geometric structures. Thus, a fixed neighbourhood distance can model all possible spatial interactions between and within objects in a scene.

To tackle this limitation, the algorithms usually utilise image segmentation for extracting spatial information, which is specified by the geometry of segmented objects. These methods employ different techniques, such as graph-cut-based segmentation [3], edge features [12], statistical region merging (SRM) [13], or robust fuzzy c-means (RFCM) [14] to create segments and extract the spatial information from images. The structure of these algorithms allows them to increase the accuracy of output maps without setting any fixed neighbourhood distances. Despite the advantages that these methods offer, such as flexible neighbourhood distances, there is no unique solution for the image segmentation problem [15,16]. This is usually solved based on parameters, e.g., scale and colour weight, typically determined based on trial and error [17]. This can lead to poor results when the algorithm uses an over-or under- segmented map to extract the spatial information.

The issue is typically addressed via a hierarchical network of segmented or classified images generated at different scales in the image or feature space [18–22]. For example, Gençtav et al. [19] proposed a hierarchical segmentation algorithm to determine segments. They used homogeneity and circularity of segments to formulate the spatial and spectral relationship between small segments in the lower level of the hierarchy and meaningful segments in the higher level of the hierarchy. Hesheng and Baowei [20] updated the fuzzy C-means (FCM) function to integrate the spatial information at different scales into spectral information to label pixels in the feature space via an iterative process. Kurtz et al. [18] used segmented images at different resolutions to classify images at different levels of spatial detail in an urban area. Fang et al. [21] implemented a multi-model deep learning framework formulated based on an over-segmentation map and semantic labelling to classify an image. The method utilised an iterative mechanism via a fixed geometry to merge segments into each other and relabel pixels to produce a real land cover map. In contrast, Yin et al. [22] proposed a top-bottom structure using graphs for an unsupervised hierarchical segmentation process. The spatial connectivity between pixels was formulated via nodes and edges, and the algorithm used the average intensity of each region to tune the weight of each edge. These methods showed that they can generate better clustering maps even when there are complex features in an image. However, there is no unique setting to formulate the concept of the scale and hierarchical structure between objects at different scales [23]. In other words, spatial relationships between objects in the scene can be subjective to the parameters defined by a human expert that reduces the flexibility of the neighbourhood system.

Some advanced classification algorithms use a cyclic mechanism of image segmentation and classification to address a flexible neighbourhood system to extract spatial information and reduce noises. These approaches use an iterative process of image segmentation and classification to integrate spatial and spectral information, use expert knowledge, and reduce noise within objects in an image [24–27]. The approach allows the algorithms to alter the geometry of objects during the classification process. For example, Baatz et al. [24] used a set of geometric operators to enable segments to change their geometry during a classification process. Hofmann et al. [27] proposed a classification method that enables objects to negotiate at object and pixel levels to change their geometry. This structure is mainly applied by supervised approaches as they require training samples and expert knowledge.

All the above methods have two main geometric characteristics in common to create a flexible neighbourhood system. First, they construct the internal and external boundaries

of an object in an image separately. This is because the geometry applied by these methods lacks the necessary complexity to take advantage of topological relationships between and within objects in a scene, e.g., car and street, or chimney and roof. For example, to address a street object which includes cars, the algorithm segments the street into multiple parts and then the street object is formed by merging the segments, regardless of the spatial relationship between cars and streets. Second, they use a segmentation process to define the initial geometry of objects. Thus, the geometric changes are restricted to the object level, not the pixel. For example, when the initial geometry of forest and shadow objects are formed in a scene, the forest object cannot capture a shadow pixel located in the boundary between two classes, and vice versa.

To overcome the above drawback, we propose a dynamic and unified geometry constructed on the Vector Agents (VAs). The VAs are a distinctive type of Geographic Automata (GA) [28] that can draw and find their geometry and state and interact with each other and their environment in a dynamic fashion [29]. The dynamic structure of the VAs enables them to support a flexible neighbourhood system where objects determine their neighbourhood distances, rather than a human expert. The method also applies a unified geometric structure that allows the interior (holes) and exterior boundaries of objects to simultaneously be modelled in a geographical area, i.e., a scene is represented by remote sensing images. This geometry gives the power to the VAs to automatically identify and remove isolated pixels and regions in the image when they lie within objects. The proposed method distinguishes itself from other classification methods based on the following spatial capabilities:

i.   Construct and change the interior and exterior geometry of objects in an image simultaneously;
ii.   Describe the topological relationships between objects in the image;
iii. Support geometric changes of objects at the pixel and object level with minimum human intervention;
iv.   Remove salt-and-pepper noise using the geometry of objects in the image.

The remainder of this paper is organised as follows: Section 2 demonstrates the structure of the VA model, and Section 3 presents the clustering results of the proposed method. Experimental results are discussed in Section 4. Finally, Section 5 concludes this paper.

2. Proposed Method

The proposed approach works in two main steps: unsupervised creation of training samples and construction of the VA model. The method first selects a collection of reliable samples from the image that is clustered by the k-means algorithm in feature space. In the construction step, the proposed method applies the selected samples to train the classifier set inside the VA model. The VAs are then automatically created and added to the image to model the cluster objects in the image.

2.1. Creation of Training Samples

Let $X = (x_{1,1}, x_{1,2}, x_{1,3}, \cdots, x_{R,C}) \in \mathbb{R}^d$ denote a multispectral d-dimensional image with $R \times C$ pixels and $x_{r,c}$ is the feature vector of size d at position (r, c) in the image (Figure 1a). Let $Y = (y_1, y_2, y_3, \cdots, y_m) \in \mathbb{C}$ denote the set of land-cover labels in the image X; where $y_i \in k$, $k = \{1, 2, \cdots, K\}$ is a set of K class labels and K is already known. Let $Z = (z_{1,1}, z_{1,2}, z_{1,3}, \cdots, z_{R,C}) \in \mathbb{R}^2$, $z_{r,c} \in k$, be the classified image (Figure 1b) that represents the ground data.
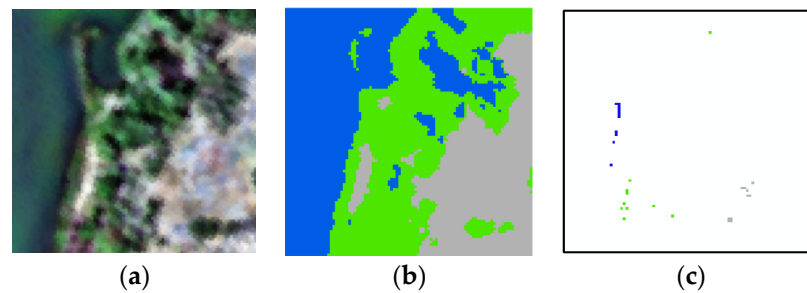
Figure 1. (**a**) RGB (red, green, blue) image of Samson, California, USA (downloaded from https: //rslab.ut.ac.ir/data, 10 March 2021) contains three targets in the scene, i.e., soil (grey), water (blue) and tree (green). The original image includes 156 bands and the principal component analysis (PCA) function in MATLAB is applied to reduce the number of bands to four, (**b**) classified image via the k-means and (**c**) the selected samples via Equation (1).

The clusters are first constructed through k-means using the Euclidean distance with 100 iterations. The number of clusters, *K*, and the dimension of feature space, *d*, are set to 3 and 4 in Figure 1, respectively. The Euclidean distance is calculated based on a vector of spectral values specified by the location of pixels and cluster centres, which are randomly selected in feature space. The algorithm calculates the spectral distance at each iteration, reassigns the pixels to the nearest cluster centre, and then updates the vector. This structure allows the algorithm to minimise the variance within each cluster based on parameters, such as the number of clusters and iterations specified by an expert user. The following rule is then applied to identify reliable training samples for each cluster from *X*.

$$\overline{p_l} - \lambda \times \sigma_{p,i} < p_i < \overline{p_l} + \lambda \times \sigma_{p,i} \tag{1}$$

where $\overline{p_l}$, $\sigma_{p,i}$, and p_i are the mean reflectance, standard deviation in band *i* of all pixels in each cluster, and the value of pixel $p \in X$ in band *i*, respectively. λ is a constant that can be set between the range of 1–3. λ in Equation (1) determines how close or far the selected samples can be from the centroid of the cluster. For example, if λ is set to 1, the selected pixels are close to the cluster centroid. Considering Equation (1), $X_L = \{(x_i, z_i)\}_{i=1}^{l}$, reliable training samples are determined, where *l* is the number of initial training samples, $z_i \in Z$ and $x_i \in X$ (Figure 1c).

VAs use these samples to train the Support Vector Machines (SVMs) within the VA model to implement transition rules. The LIBSVM classification library for support vector machines developed by [30] was applied. In our case, the SVM classifier is trained according to the Radial Basis Function (RBF) kernel. C, regularisation parameter and G, the bandwidth, of the RBF kernel are selected using the n-fold cross-validated grid search algorithm where *n* is set to 10.

2.2. Construction of the VA Model

In general, VAs are a distinctive type of geographic automata (GA) [28] that can change their geometry and interact with other objects in a simulation space [29]. The main elements of the VA model are geometry (**L**) and geometry rules (**M$_L$**), state (**S**) and transition rules (**T$_S$**) and neighbourhood (**N**) and neighbourhood rules (**R$_N$**) defined by [31]. In the construction step, we update these components to classify the image. Figure 2 demonstrates the elements of the VA model for image classification. In each iteration, the VAs use their sensors to capture information from feature and image space simultaneously. This information includes the location and geometric structure of VAs in the image and the pixels' spectral information in the feature space.
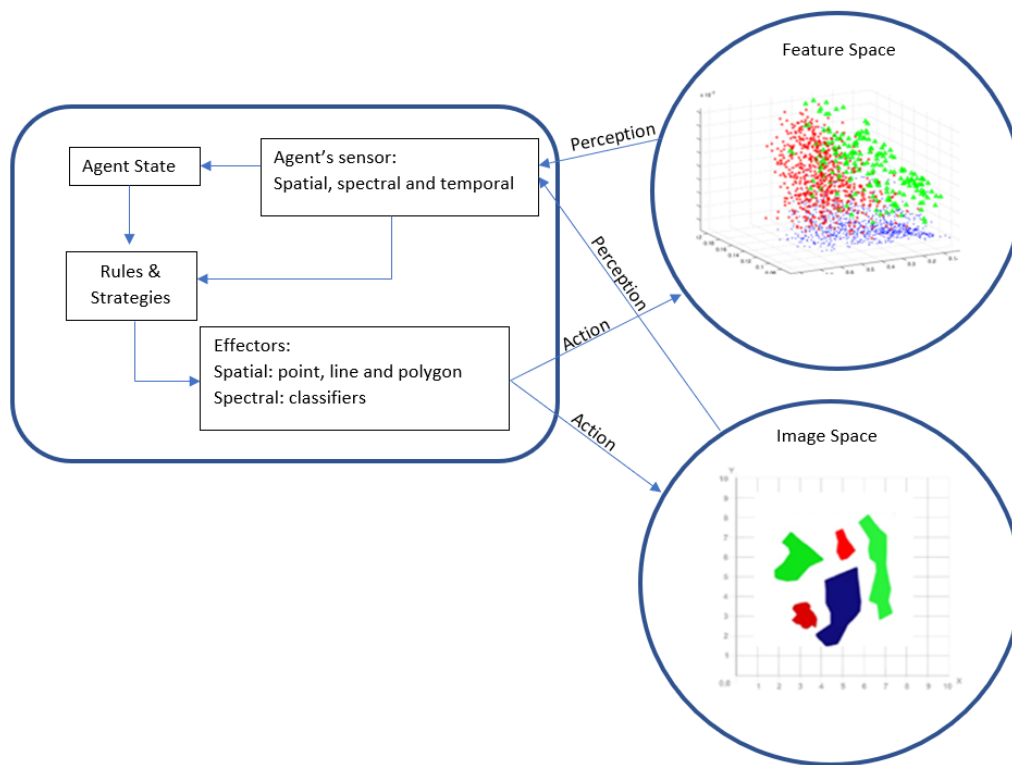
Figure 2. Schematic diagram of the model architecture of the VA including state, sensor, rules and strategies, and effectors.

The VAs then apply this information to execute rules and strategies in the image via their effectors, i.e., point, line, and polygon. The spatial relationship between these elements is defined and formulated in the following section. We used a Java implementation of the Repast V.2.5 (Recursive Porous Agent Simulation Toolkit, the University of Chicago, IL, USA) modelling framework [32] along with a generic Vector Agent library developed by [30] to implement the proposed method.

2.2.1. Geometry and Geometry Methods

In the context of image classification, the geometry component (**L**) of a VA stores the vertices that define the boundary ∂X_{VA} of the VA. X_{VA} is the connected subset of the image space formed by the pixels belonging to the VA. In contrast to the conventional agent-based classification approaches that use a predefined geometry to classify an image [33,34], the VAs can automatically construct their geometry. The geometric methods $\mathbf{M_L}$ enable VAs to change the boundary ∂X_{VA} and interact with other VAs in the simulation domain. These methods can be summarised as follows:

- Vertex displacement: This places a new vertex and connects two vertices together by two half-edges, specified according to a single direction (Figure 3a,b);
- Converging vertex displacement: Two new edges are constructed to a single vertex form two existing neighbouring vertices (Figure 3c);
- Half-edge joining: This constructs a new edge based on a twin or bidirectional edge that is formed by two half-edges (Figure 3d);
- Edge remove: This forms a new polygon by merging two polygons (Figure 3f).

Figure 4 illustrates how the VA uses the above rules to create a polygon that includes a hole. These holes can be applied to define the geometry of isolated pixels or regions (connected pixels) that may exist inside the information classes in the image, e.g., the blue pixels within the green area in Figure 1b.
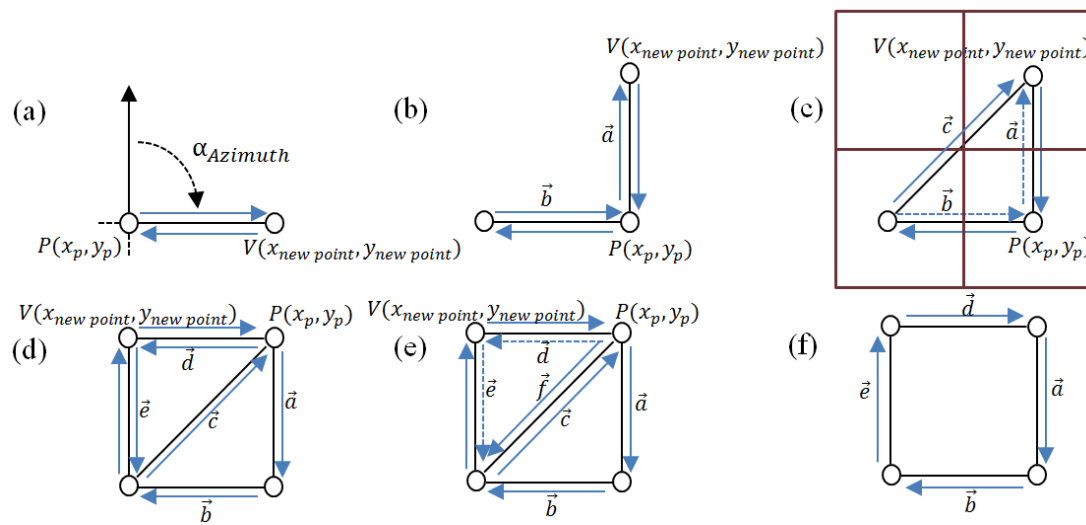
Figure 3. Four elementary operations are required to change the image objects' geometry: (**a,b**) vertex displacement, (**c**) converging vertex displacement, (**d**) vertex displacement to create $\vec{d}$ and converging vertex displacement to construct $\vec{e}$, (**e**) edge joining, and (**f**) edge remove (from [33]).

$t = 60$ $\qquad$ $t = 65$ $\qquad$ $t = 72$ $\qquad$ $t = 76$ $\qquad$ $t = 84$
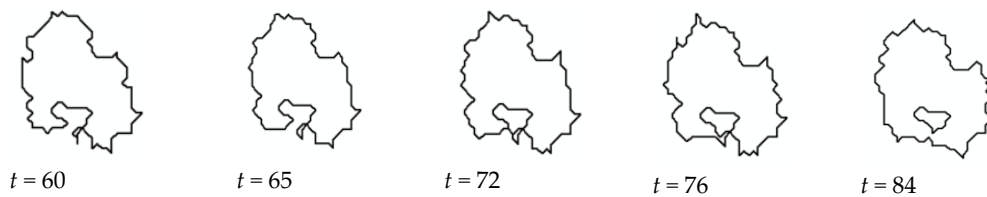
Figure 4. Simulation result for 84-time steps in the agent modelling Repast Simphony representing how the VA uses the four operations above to transform its geometry, including an interior and exterior boundary.

Through this geometry, the VAs can reduce the salt-and-pepper noise in the image and increase the quality of the classified outputs.

2.2.2. State and Transition Rules

The state **S** of a VA is the VA class that is determined via the SVM classifier trained on the selected samples generated in the creation step. $\mathbf{T_S}$ rules allow the VAs to find and update their classes and evaluate pixels in the image. A VA is added to the image as a point if it can satisfy the following rule:

- The candidate pixel x_c and its immediate neighbours in the image must be members of the same class. VAs use the SVM classifier to evaluate such membership.

The VA points then start evolving geometrically by capturing nearby pixels. In each iteration, the VA adds a new pixel as a point to its geometry if the new pixel and VA are in the same class. In the event that a new point is added to the VA, it updates its geometry (**L**) and state (**S**) using $\mathbf{M_L}$ and $\mathbf{T_S}$ rules. If the candidate pixel belongs to another VA, and these VAs are in the same class, then the VA applies the joining/removing rule to create a new geometry. In this event, all attributes of the removed VA are transferred to the active VA in the image. Figure 5a shows different sizes and shapes for the exterior boundaries of the information classes, while the size of holes is restricted by the large holes (Figure 5c).
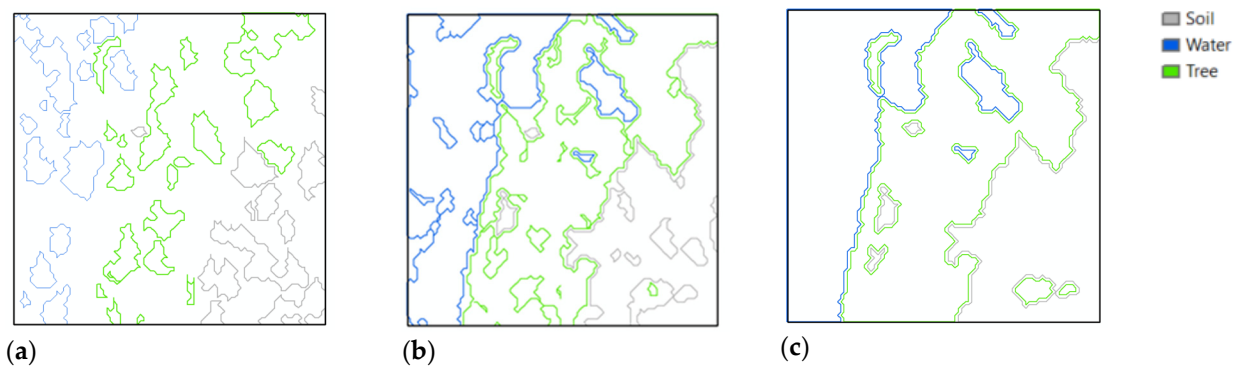
Figure 5. (**a**–**c**) display the evolving process of the soil VA, water VA, and tree VA at different successive time steps in the image by using the transition and geometric rules.

2.2.3. Neighbourhood and Neighbourhood Rules

The neighbourhood component **N** for a VA is a collection of VAs that are located within the neighbour distance of the VA in the image. If they can stratify the neighbourhood rule (**R$_N$**) in Equation (2):

$$\left\{ VA_{j,t}, j \in \mathbb{N} : d\left(VA_{i,t}, VA_{j,t} \right) \leq r\sqrt{2} \right\} \tag{2}$$

where $d\left(VA_{i,t}, VA_{j,t} \right)$ is the Euclidean distance in simulation space between VA_i and VA_j at time step, t. VAs use this rule to interact with each other geometrically through a set of advanced geometric methods, such as joining/removing or growing/shrinking, defined and implemented by [33]. As the spectral information of pixels is the only available information, VAs solely use the joining/removing method to geometrically interact with each other in the image. Figure 6 shows an example of the joining/removing method in which two VAs implement **R$_N$** to merge into each other.
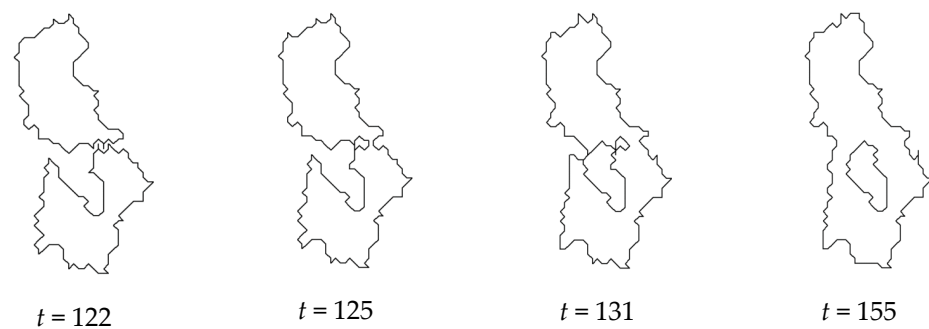
Figure 6. The interaction between two VAs that creates a new VA by geometric integration.

2.2.4. Implementation of VA for Unsupervised Classification

The flexibility of the VA approach to both segment and classify objects dynamically allows a comprehensive image classification approach whereby VAs are born, evolve, and sometimes die throughout the image space until the objects are extracted from the image. The process starts by seeding a desired number of VAs as points in the image whose coordinates correspond to the centre of image pixels. This seeding process can obey a specific sampling scheme (e.g., fully random, stratified random, systematic unaligned random, etc.). In this paper, a systematic unaligned random sampling scheme is chosen so that VAs are seeded for every class at various locations throughout the image as prescribed in the previous section. These points or VAs will evolve via an iterative process into polygons by capturing nearby pixels based on transition and neighbourhood rules via their geometry methods. In the event that all VAs are passive, the algorithm only considers the geometry location of pixels as spatial criteria to capture unclassified pixels. Figure 7 displays the classified map that is generated by the VA model. A comparison between

Figures 1b and 7a shows the effectiveness of the proposed method to reduce the noise in the classified map by using the geometry of features in the image (Figure 7b). In this example, the VAs are only applied to remove the small holes.

Soil
Water
Tree

(**a**)   (**b**)

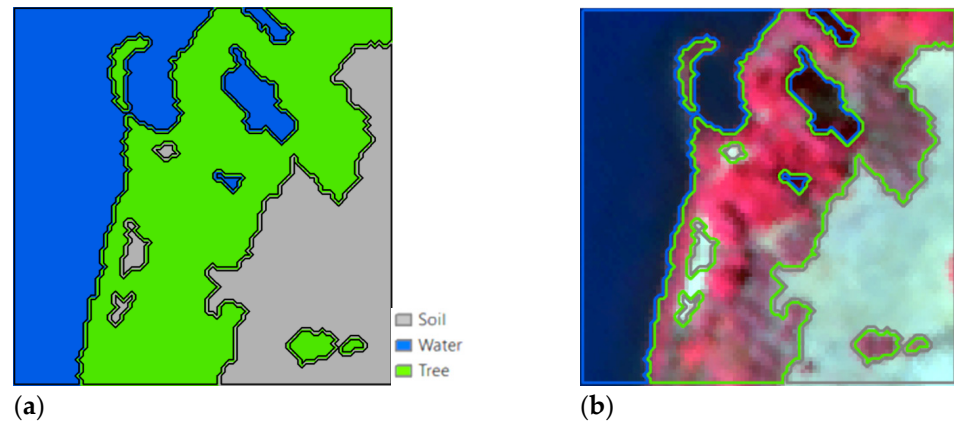**Figure 7.** (**a**) Displays the classified map that includes three objects and (**b**) displays the classified image superimposed on the false colour image.

3. Experiments and Results

To study the performance of the VA model, three high-resolution satellite remotely sensed images have been used. They are described below along with the details of clustering methods and the metrics applied to measure the effectiveness of the proposed method and the outputs of the proposed method.

3.1. Datasets

The proposed approach was experimentally tested on three hyperspectral images: Pavia Centre and Pavia University datasets collected by a ROSIS sensor and an AVIRIS scene of Salinas Valley, California (Figure 8). The number of spectral bands is 102 for Pavia Centre, 103 for Pavia University, and 224 for Salinas Valley. Considering the computational efficiency, four bands for each image are selected using the PCA function in MATLAB. In the first experiment, a subset of 199×199 is cut from the Pavia Centre image with a pixel size of 1.3-metres (Figure 8a). The image includes four information classes: water (C1), tree (C2), bare soil (C3), and bridge (C4) (Figure 8b).

For the second experiment, we used a subset of 199×199 cut from the University of Pavia image with a pixel size of 1.3-m (Figure 8c). It covers an urban area including five information classes: buildings and asphalt (C1), shadow (C2), bare soil (C3), meadows (C4), and painted metal sheets (C5) (Figure 8d). In the third experiment, a subset of hyperspectral AVIRIS with the size of 198×198 pixels was applied to test the VA method (Figure 8e). The geometric resolution is 3.7 m. The scene covers an agricultural zone in California, containing seven information classes: Vinyard_untrained (C1), Brocoli_green_weeds_2 (C2), Grapes_untrained (C3), Fallow_rough_plow (C4), Fallow_smooth (C5), Stubble (C6), and Celery (C7). (Figure 8f).

3.2. Image Clustering

To evaluate the performance of the proposed approach for removing noises from the classified images, we compared the results of the VA model with the conventional classification approaches in two modes.
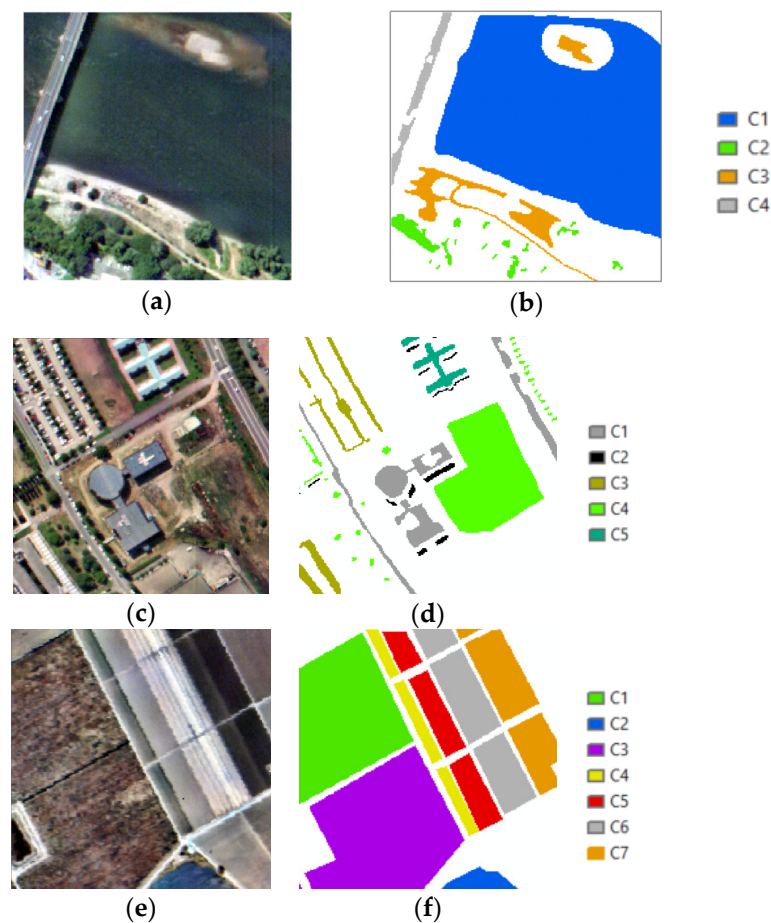
Figure 8. (**a**) Pavia Center image and its ground truth data (**b**); (**c**) the University of Pavia image and (**d**) displays the ground truth data for the University of Pavia; (**e**) Salinas Valley image and its ground truth data (**f**).

i.   Unsupervised

We assume that the only available information is the number of clusters and there is no semantic information. In this setting, a classical k-means algorithm is applied, and the spatial information is imposed to the algorithm at two different stages: pre-and post-classification as follows:

- Spectral-Spatial classification (SSC)

The algorithm utilises the density function first to group pixels together in the feature space, using a combination of spectral similarity and spatial proximity between pixels. It then applies the k-means algorithm to label pixels, based on a new vector determined by the pixel values and the features associated with its corresponding regions in the segmented image.

- Majority Filtering (MF)

The algorithm uses a three-by-three fixed window to replace labels in the classified image by k-means based on the majority of their contiguous neighbouring pixels. The number of neighbours is set to eight to retain edges.

ii.   Semi-supervised

The algorithms use a limited number of training samples at two levels to train the SVM model and classify the images: pixel and object.

- SVM:

We used the SVM method described in Section 2 to label the pixels without using spatial information. Training samples were manually selected from the ground truth data to train the SVM.

- Mean Shift Segmentation (MSS):

The method first uses the MSS algorithm to segment the image. The SVM model is then applied to label segments. In this scenario, we manually selected the training objects.

- Multiresolution Segmentation (MRS):

The algorithm employs the MRS function, which is formulated based on a combination of spectral homogeneity and shape homogeneity to segment the image. Like the MSS method, the training objects are manually selected to train the SVM classifier.

In the above methods, the segmentation parameters are manually defined. Since there are no rules to determine those parameters, we conducted different segmentation experiments to initialise the segmentation parameters to generate segments with less speckle.

3.3. Evaluation Metrics

To evaluate the results quantitatively, the VA objects are compared with their corresponding reference objects in the ground truth maps. We use the following metrics to assess the accuracy of the VA maps [35]:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Jaccard\ index = \frac{TP}{TP + FN + FP} \tag{4}$$

$$F - score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{5}$$

$$recall = \frac{TP}{TP + FN} \tag{6}$$

True Positive (TP), False Positive (FP), and False Negative (FN) are correctly detected pixels, wrongly detected pixels, and unrecognised pixels, respectively.

We also evaluate the spatial connectivity and fragmentation of the objects in the classified maps. The Perimeter/Area (P/A) ratio of classified objects is applied to assess the local neighbourhood connectivity between pixels within clusters in the classified maps.

3.4. Results

We first used the k-means algorithm to cluster the images. From the classified maps, 15 pixels are randomly selected for each cluster using Equation (1), where λ is set to 1. The VA model then applies the selected samples to train its SVM classifier. The parameters of the k-means and SVM algorithm were set as defined in Section 2. The trained VAs are then added to the image to extract clusters from images (Figure 9).

For accuracy assessment among different methods, the initial clusters were mapped into information classes. This is because, in unsupervised classification each, information class may contain more than one cluster. For example, in the University of Pavia dataset, the number of clusters and information classes were considered seven and five, respectively.
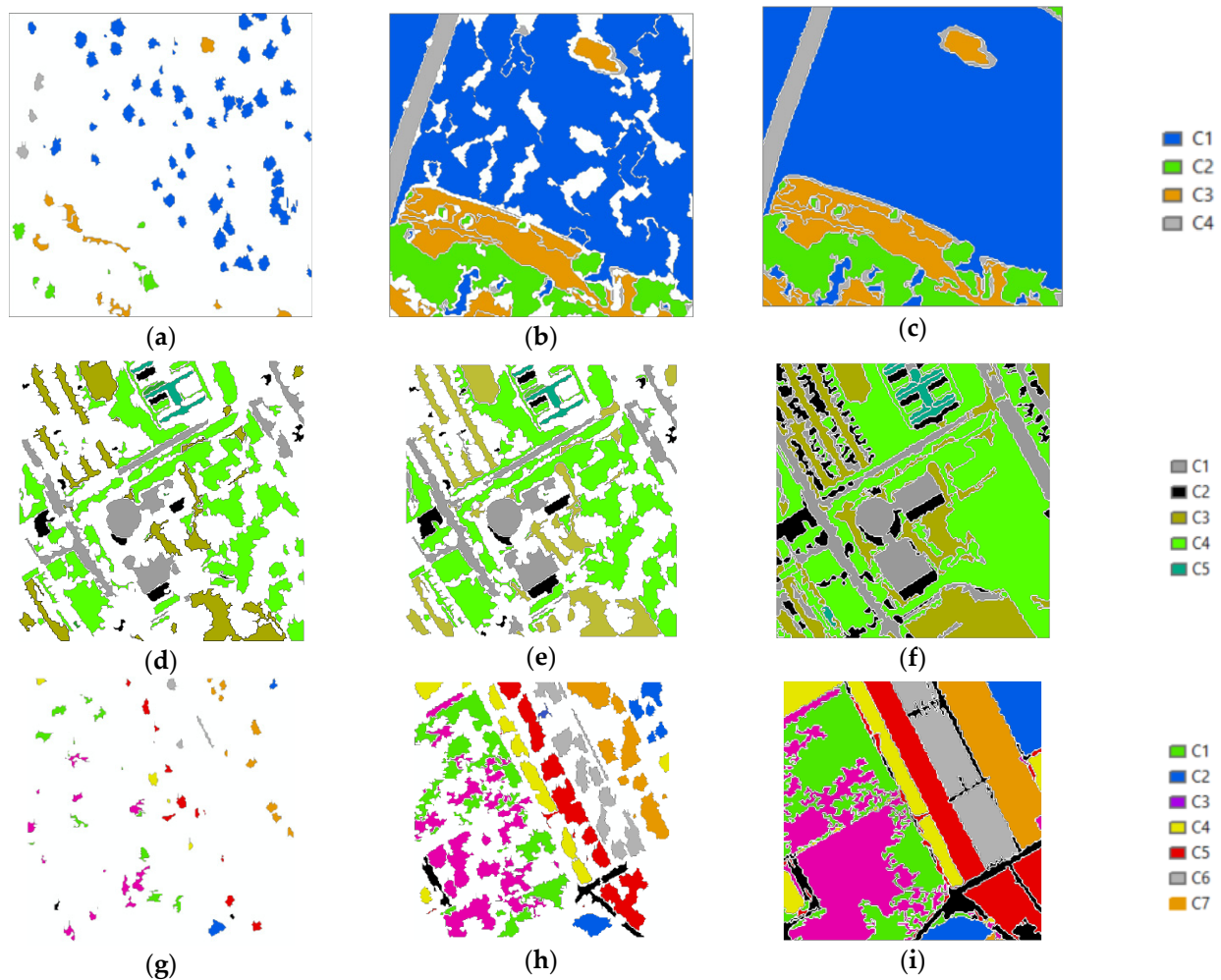
Figure 9. (**a**–**i**) display how the VAs geometrically evolve in the image at different successive time steps for three different datasets. (**a**–**c**) Five different types of VA mapped into four information classes according to the information classes in Figure 8b; (**d**–**f**) seven different types of the VA mapped into five information classes: shown in Figure 8d; (**g**–**i**) seven information classes shown in Figure 8f were considered to assess the results of the VA map.

4. Discussion

The main contribution of the proposed work is to introduce a geometry-led approach with minimum human intervention to reduce salt-and-pepper noise in classified maps. To assess the performance of the methods, different conventional spectral–spatial clustering methods were applied (described in Section 3.2). In the next section, the results of the proposed method will be discussed in more detail using different datasets.

4.1. Dataset 1

4.1.1. Speckled Noise Analysis

From Figure 10b,c,e–g, it can be seen how the use of spatial information can reduce the number of isolated pixels within the classified images. For example, the water object, C1, in Figure 10a,d, where only spectral information is applied for classification, contains some isolated bridge pixels shown by the black circle. The black circles in Figure 10a,e show that using the MF method effectively reduces noises within the C1 object. However, the MF method was unable to completely remove the isolated region (grey pixels) within the black circle due to its fixed geometric structure. An additional filtering step can be applied to reduce these noises. On the other hand, it can affect the quality of boundaries, especially if there are low local variations between objects with different classes.

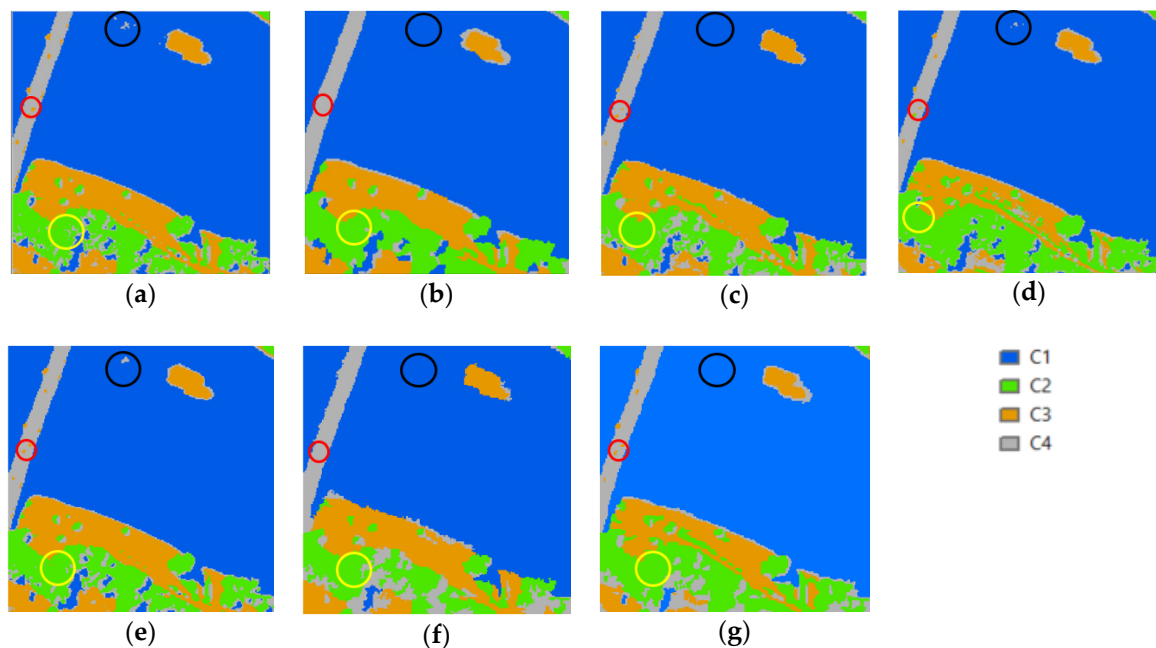**Figure 10.** (**a**) The k-means map, (**b**) the VA map, (**c**) the SSC map provided through using the MSS algorithm and k-means, (**d**) the SVM map that is created by manually selected training samples, (**e**) the MF map, (**f**) the MSS map generated using the MSS and the SVM, and (**g**) the MRS and the SVM are applied to label image objects.

The generated objects by the SSC, MSS and MRS methods are more patch-like than the created objects by the MF, SVM and k-means methods. The yellow circles illustrate the ability of these methods to deal with heterogeneous objects, i.e., C2 objects. This can be explained by the fact that these methods use a flexible neighbourhood system to classify the image. Whilst these methods provide a better appearance than the fixed neighbourhood systems, the results are subjective to the function and parameters specified by a human expert. For example, the MSS map provided a better appearance for the bridge objects (the red circles in Figure 10) compared to the MRS method. In contrast, the tree objects are more homogenous in the MRS map. The other issue for the geometry applied by these methods is that it lacks the ability to directly model the objects in the scene during the classification. Consequently, these methods can use spatial relationships once the image is segmented.

Conversely, the geometry of the VA model allows the algorithm to directly link an object in the simulation space to its corresponding object in the scene. It provides an actual geometry that allows the method to automatically formulate the topological relationship between and within objects without setting any geometric parameters. The output results in the VA map in Figure 10b show how the use of this geometry can improve the appearance of the classified maps and remove speckles within objects. For instance, there are no isolated pixels within C1 and C4 objects.

4.1.2. Accuracy Analysis

In Table 1, the highest values of each metric for each class are marked in bold and the lowest values are underlined. As can be seen from Table 1, the best results are achieved by the VA model. The OA value for the VA model is 99.97% in Table 1, which indicates that most pixels are correctly detected. However, this improvement is not significant, i.e., the OA of the k-means is 99.83. This is due to the imbalanced distribution of the ground data, almost 85% of ground data belong to the C1 class. Thus, the accuracy of the C1 class overshadows the inaccuracy of the C3 and C4 classes. Moreover, the ground data do not cover the whole C4 object, i.e., the bridge, even though it has a sharp and clear exterior boundary.

Table 1. Comparison between k-means, SSC, SVM, VA, MF, MSS, and MRS approach. The total number of C1–C4 ground pixels in Figure 8b is 22,469.

Class	Metrics	k-Means	SSC	SVM	VA	MF	MSS	MRS
C1	precision	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
	Jaccard-index	99.88	**100.00**	99.95	**100.00**	99.91	**100.00**	99.99
	F-score	99.94	**100.00**	99.98	**100.00**	99.96	**100.00**	**100.00**
	recall	99.88	**100.00**	99.95	**100.00**	99.91	**100.00**	99.99
C2	precision	**100.00**	99.86	94.29	99.46	99.87	**100.00**	98.28
	Jaccard-index	98.25	98.79	94.17	99.20	**99.60**	94.35	98.15
	F-score	99.12	99.39	97.00	99.60	**99.80**	97.10	99.07
	recall	98.25	98.92	**99.87**	99.73	99.73	94.35	99.87
C3	precision	99.94	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	99.88
	Jaccard-index	99.94	99.83	93.94	99.77	**99.94**	98.04	99.14
	F-score	99.97	99.91	96.88	99.88	**99.97**	99.01	99.57
	recall	100.00	99.83	93.94	99.77	99.94	98.04	99.25
C4	precision	95.22	98.66	91.34	**99.73**	97.49	93.89	91.34
	Jaccard-index	95.10	98.66	91.34	**99.73**	97.49	93.89	91.34
	F-score	97.49	99.33	95.47	**99.86**	98.73	96.85	95.47
	recall	99.86	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
C1–4	OA	99.83	99.95	99.49	**99.97**	99.91	99.66	99.93

The OA of the SVM model in this experiment is slightly lower than the OA of the k-means. This is because the C2 and C3 pixels are misclassified. However, the recall values show better performance of the SVM model to reduce the isolated pixels within C1, C2, and C4 compared to the k-means.

Table 1 shows the high performance of the VA model to identify pixels in C4 objects. The results of the VA model are better than the MRS method, with almost 7% improvement for the precision and Jaccard-index values. This is because the parameters specified by an analyst form the geometry of the segments, and the method models interior and exterior boundaries of objects separately. Thus, some isolated pixels or regions exist within segments that can change features' spectral behaviour, especially when they are close to borders in the feature space. The results in Table 1 also confirm the improved performance of the VA model compared to the MSS method, by more than 3%, to extract the C4 pixels.

Although the Recall values for the C4 class are 100%, the classified maps show some isolated pixels within the classified maps (see Figure 10d,e). Thus, the C4 recall values do not accurately reflect the actual geometric structure of the bridge object in the classified maps. This is because the ground data in Figure 8b do not cover the whole bridge structure. The F-score of the VA model is better than the F-score of the MRS and MMS algorithms for C2 and C3 objects. This indicates that the proposed geometry of the VA can accurately model heterogeneous objects such as C2 and C3 in the scene.

For better analysis of the performance of the VA model, we used the geometric parameters of two heterogeneous classes, C3 and C4, in this experiment. In Table 2, the P/A ratio is calculated using the geometric information of the objects in the C3 and C4 classes.

Table 2. Comparison between the P/A ratio of SSC, VA, MSS, and MRS approach.

Method	SSC		VA		MSS		MRS	
Class	C2	C3	C2	C3	C2	C3	C2	C3
Number	37	26	11	6	15	6	22	16
Area (m²)	8549.85	9621.59	9335.86	9820.32	7479.82	10,263.82	10,225.76	7947.24
Perimeter (m)	2513.01	2089.01	1533.01	1379.62	1855.19	1599.30	2282.84	1787.64
P/A ratio	0.29	0.21	0.16	0.14	0.25	0.16	0.22	0.22

Table 2 shows that the VA algorithm used 17 objects to extract C2 and C3 pixels. This number for the SSC map is 63, while the difference between calculated values for C2 and C3 class in Table 1 is less than 1% for these two methods. That means the C2 and C3 objects are less fragmented and more connected in the VA map. The number of regions for C2 and C3 classes in the ground truth map (Figure 8b) are 17 and 3, respectively. From Table 2, it can be seen that there is almost 3% difference between the total area of the C3 objects in the SSC and VA, while there is a 66% difference between the perimeters of the C3 objects. This means that the C3 objects in the SSC map are more fragmented. The reason for this is the isolated pixels and regions within the C3 objects, which increase fragmentation. These holes can be cut via an additional process. For example, the algorithm can use an additional filter to identify the holes within objects and cut these objects. However, the results would be subjective to the size and shape of filters and the shapes of holes. Additionally, this process can decrease the level of automation for unsupervised algorithms.

The geometric information of the C2 objects in Table 2 illustrates that the VA algorithm reduces the noises within C2 objects better than the SSC method. The MRS method applied 22 patches to cover the area of 10,225.76 m^2, while the VA model used 11 patches to extract C2 objects with an area of 9335.86 m^2. The MSS and VA algorithms generated six patches to address the C3 objects. However, the P/A ratio of the VA model is 0.02% lower than the MSS method for the Class 3 objects. This means the connectivity between C3 pixels in the VA map is better than the MMS map.

4.2. Dataset 2

4.2.1. Speckled Noise Analysis

It can be clearly observed from Figure 11b,e–g how the use of spatial information can significantly reduce the speckled noises and produce a better visual appearance for the classified maps.
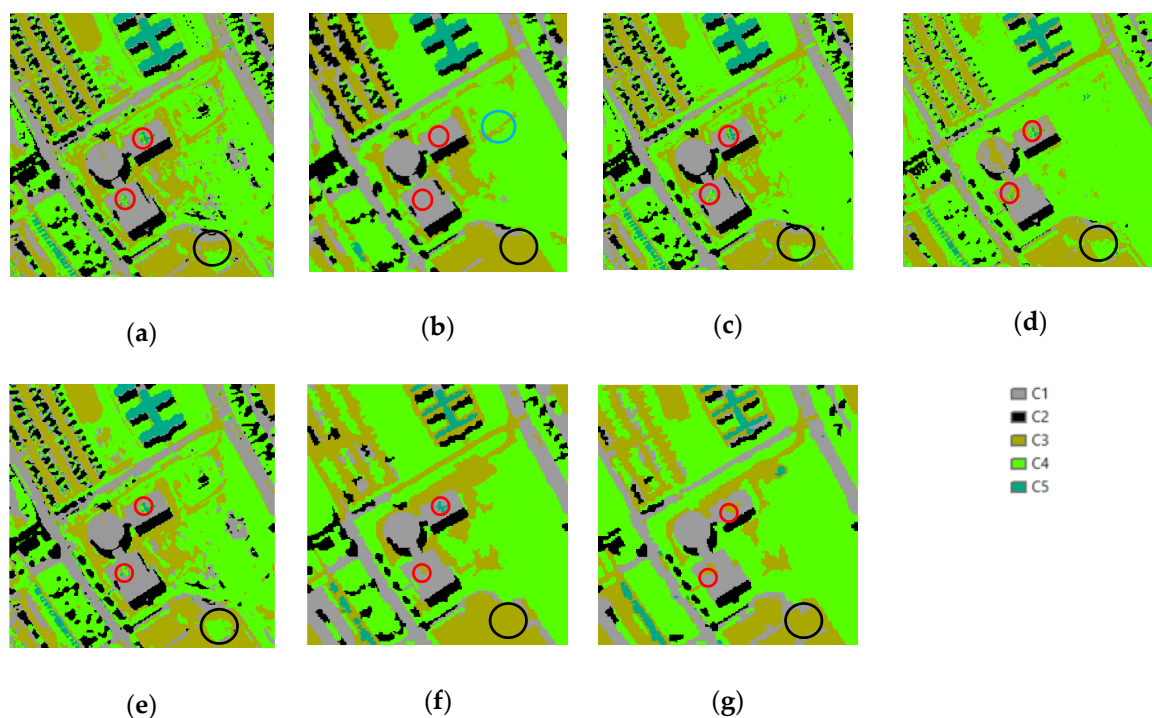

Figure 11. (**a**) The k-means map, (**b**) the VA map, (**c**) the SSC map provided through using the MSS algorithm and k-means, (**d**) the SVM map that is created by manually selected training samples, (**e**) the MF map, (**f**) the MSS map provided using the MSS and the SVM, and (**g**) the MRS and the SVM are applied to label image objects.

Figure 11 shows that the VA map generates a better appearance for the roof objects in the C1 class than the other methods. For example, the red circles in Figure 11 illustrate no isolated pixels or regions within the C3 area in the VA map. This is because the VAs use a unified geometry that enables them to simultaneously model objects' inner and outer boundaries in the image. If there is a geometric hole in the VA, it implies that the object inside the VA has a different class from the VA. The VA then removes this object by simply reconstructing its geometry. In contrast, the geometry applied by the MSS and MRS methods lacks this capability to determine the spatial connectivity within objects in an image. Therefore, an additional step is usually required to reformulate the neighbourhood system to address these holes after classification.

Figure 11 illustrates that the MF algorithm creates less noise within the C2 objects compared to the k-means algorithm. For example, the size of the isolated regions within the shadow object, located next to the building in the south of the classified map, is decreased in the MF map. The MSS and MRS algorithms perform better than the MF method to remove the isolated pixels within shadow objects. This is because the shadow objects have sharp and crisp exterior boundaries, and the size of isolated regions is relatively small. However, the distribution of the shadow objects within the MSS and MRS maps are different because these methods use different functions to segment the image. Figure 11 confirms no isolated pixels or regions within the classified maps for the C3 and C2 objects.

For the C4 objects, the SVM method shows better performance than the other methods (Figure 11d). For example, the VA map displays a few isolated regions within the C4 area, highlighted by the blue circle. There is no one-to-one relationship between information class and cluster in the scene (Figure 9f). Thus, the neighbouring VAs in the same class but with different clusters cannot join together to remove these regions.

The classified maps show no isolated pixels or regions within the C5 objects as they have distinctive spectral behaviours. However, a visual assessment between the classified maps in Figure 11 and the ground truth map in Figure 8d illustrates that the MSS and MRS methods significantly change the exterior boundaries of the C5 objects.

4.2.2. Accuracy Analysis

Table 3 includes the recall, precision, Jaccard index, and F-score values of the classified maps. In the table, the highest and lowest values are highlighted. It can be seen that the VA model generated better recall values for all clusters compared to the MSS and MRS methods. For example, there is an improvement of 10% for the VA recall value for the C2 objects compared to those in the MRS map. Although there are no isolated pixels and regions within C2 objects in the MRS map, the method could not accurately address the exterior boundaries of the shadow objects in the classified map. This is because the static geometry of the method does not allow the classifier to change the boundary of the objects during the classification step. The recall value for the C1 objects in the VA map is 99.44%, a 10% improvement compared to the MMS results. A visual assessment between the C1 objects in Figures 8d and 11f illustrates that the MSS algorithm could not accurately delineate the exterior boundaries of the C1 objects. This is due to the fact that the method uses a unique formula initialised by a human expert to model different shapes and neighbourhood distances. The red circles in the classified maps show no isolated regions or pixels within the roof objects in the VA map.

Table 3 shows that the results of the SSC method are better than the MSS and MRS methods. For instance, the F-score of the SSC map for the C5 objects is 20% more than the MRS method. The use of an advanced method that allows the classifier to modify the boundaries of segments during the classification step and use different parameters can improve the results for these methods. However, these changes are against the nature of unsupervised approaches that aim to increase the level of automation for image classification.

Table 3. Comparison between k-means, SSC, SVM, VA, MF, MSS, and MRS approach. The total number of C1–C5 ground pixels in Figure 8d is 9753.

Class	Metrics	k-Means	SSC	SVM	VA	MF	MSS	MRS
	precision	<u>76.82</u>	87.69	**99.38**	93.24	73.25	78.86	87.14
C1	Jaccard-index	75.41	87.19	82.55	**92.75**	<u>68.36</u>	72.43	81.85
	F-score	85.98	93.16	90.44	**96.24**	<u>81.21</u>	84.01	90.02
	recall	97.62	99.35	<u>82.98</u>	**99.44**	91.10	89.89	93.09
	precision	71.05	97.85	**100**	98.15	<u>61.07</u>	79.08	98.55
C2	Jaccard-index	71.05	97.85	**100**	98.15	<u>58.55</u>	73.48	84.21
	F-score	83.07	98.91	**100**	99.07	<u>73.85</u>	84.72	91.43
	recall	**100**	**100**	**100**	**100**	93.42	91.22	<u>85.27</u>
	precision	46.46	52.12	**68.14**	55.43	38.41	<u>27.94</u>	44.27
C3	Jaccard-index	42.47	43.98	**67.64**	52.01	32.48	<u>21.23</u>	37.35
	F-score	59.62	61.09	**80.7**	68.43	49.03	<u>35.02</u>	54.39
	recall	83.16	73.79	**98.93**	89.40	67.79	<u>46.92</u>	70.50
	precision	99.78	99.78	99.79	**99.87**	96.83	95.05	<u>94.95</u>
C4	Jaccard-index	67.63	84.2	**97.38**	82.96	65.72	74.05	78.84
	F-score	80.69	91.42	**98.67**	90.69	<u>79.32</u>	85.09	88.17
	recall	67.73	84.36	**97.58**	83.05	<u>67.17</u>	77.01	82.29
	precision	98.53	97.8	97.75	**100**	89.94	97.75	<u>87.77</u>
C5	Jaccard-index	97.99	97.09	87.07	**99.44**	78.81	63.74	<u>62.14</u>
	F-score	98.98	98.52	93.09	**99.72**	88.15	77.85	<u>76.65</u>
	recall	**99.44**	99.26	88.85	**99.44**	86.43	<u>64.68</u>	68.03
C1–5	OA	79.56	87.93	**93.88**	89.19	<u>74.85</u>	76.10	82.69

Table 3 shows the better performance of the SVM method compared to the other methods to identify the C3 pixels. For instance, the recall value of the VA model is 89.40%, while this number is 98.93% for the SVM method. This is because the C4 class in the VA map contains more than one cluster. As a result, the adjacent VAs could not merge with each other in the C4 zones (Figure 9f) in order to remove the isolated C3 areas (the blue circles in Figure 11b). This also decreases the recall value of the VA model for the C4 class compared to the SVM method. As the C4 class has the highest number of training samples, which is more than 60% of the total ground pixels, the accuracy of the VA model is reduced in comparison to the SVM method. From Table 3, it can be seen that the VA model does return better results for objects in the C5 class than the other methods.

Compared with k-means, the VA model improves the OA to a large degree by reducing the noises. The overall accuracy (OA) of the VA model and k-means is 89.19% and 79.56%, respectively. It confirms that using VAs can improve the accuracy of the results by more than 10%. However, the results of the SSC and VA approaches are almost 4% different. This is because the C4 class contains more than one cluster. Accordingly, the C4 VA could not join together and remove isolated C3 areas in the scene.

As most ground truth samples belong to the C4 class, this significantly affects the performance of the VA model compared to the SVM method in this experiment. Additionally, the ground pixels do not entirely contain the whole objects for man-made features. For instance, there are many isolated pixels and regions within the building objects generated by the SVM and SSC method, while there are no ground pixels for these areas.

Like the first experiment, we use the geometric information of the C1 and C4 classes to have a better assessment of the VA model (Table 4). The VA model used 40 patches to address the C1 objects. This number is 342 patches for the SSC method, while the area difference between the C1 objects in these two methods is less than 3%. That means that spatial connectivity between C1 objects in the VA map is higher than the SSC map. Although the MSS and MRS methods used fewer patches to address the C1 and C4 objects, they were unable to address the boundaries within these clusters accurately.

Table 4. Comparison between the P/A ratio of SSC, VA, MSS, and MRS approach.

Method	SSC		VA		MSS		MRS	
Class	C1	C4	C1	C4	C1	C4	C1	C4
Number	342	370	40	39	21	28	18	24
Area (m^2)	13,850.52	38,233.43	13,555.06	36,978.89	13,146.53	38,329.54	13,824.32	40,807.29
Perimeter (m)	7083.78	11,199.85	4140.10	6285.49	3123.36	5586.51	3762.62	5619.38
P/A ratio	0.51	0.29	0.30	0.17	0.23	0.22	0.30	0.17

4.3. Dataset 3

4.3.1. Speckled Noise Analysis

As with the previous experiments, we use a combination of unsupervised and semi-supervised methods to assess the performance of the proposed method. It can be clearly observed from Figure 12 that the VA, MSS, and MRS methods provide more homogenous areas than the other methods. For instance, the k-means map contains a lot of salt-and-pepper noises for Class 1 and Class 3, which gives the output a speckled appearance. Figure 12 also demonstrates that the VA map offers a better visual appearance than the SSC map. There are lots of isolated pixels within the SVM and SSC maps. The VA map shows no speckles within the C2 object. In contrast, a few speckles, especially near the boundaries, can be observed in the other classified maps. The MF map illustrates that the use of filters is effective in reducing noise. However, it can significantly affect the boundaries of objects in the scene. The classified maps show that there are a few isolated regions or pixels within the C4–C7 objects. However, the exterior boundaries of C4–C7 objects are different. This is due to the various geometric structures which are applied by these methods.
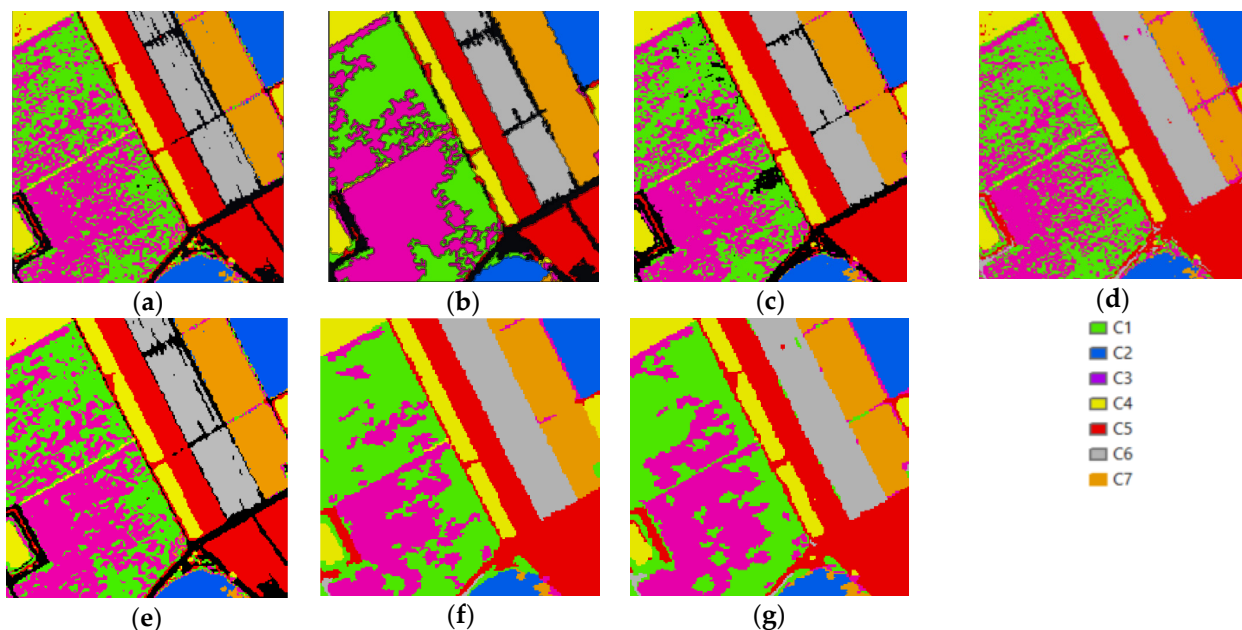

Figure 12. (**a**) The k-means map, (**b**) the VA map, (**c**) the SSC map provided through using the MSS algorithm and k-means, (**d**) the SVM map that is created by manually selected training samples, (**e**) the MF map, (**f**) the MSS map provided using the MSS and the SVM, and (**g**) the MRS and the SVM are applied to label image objects.

4.3.2. Accuracy Analysis

Table 5 lists precision, Jaccard index, F-score, and Recall values of the classification maps based on the different methods. In the table, the highest and lowest values are marked. The values in Table 5 confirm that the use of spatial information can significantly improve the performance of the k-means method. The highest recall value for Class 1 belongs to the MSS method, indicating that most C1 pixels are identified. Although the

Recall value of the VA model is 20% lower than the MSS method, the difference between the F-score values of these two methods is 7%. That indicates the FP value of the MSS is higher than the VA model for Class C1, which decreases the recall value of the MSS map for the C3 class by 13%. Although the recall value of the VA for the C1 class is lower than the MSS method, Figure 12 shows that there are no isolated regions within the C1 object. While the MSS map displays a few isolated areas within the C1 and C3 objects in the MSS map.

Table 5. Comparison between the results of k-means, SSC, SVM, and VA approach. The total number of C1–C7 pixels in Figure 8f is 27,659.

Class	Metrics	k-Means	SSC	SVM	VA	MF	MSS	MRS
C1	precision	58.34	61.83	54.83	**65.51**	61.6	63.58	62.23
	Jaccard-index	41.26	45.74	41.87	48.04	43.19	**56.47**	49.15
	F-score	58.41	62.77	59.02	64.9	60.33	**72.18**	65.9
	Recall	58.49	63.75	63.91	64.31	59.1	**83.46**	70.03
C2	Precision	98.92	**100**	**100**	**100**	96.5	**100**	**100**
	Jaccard-index	90.32	88.22	88.42	**94.81**	90.54	88.22	86.63
	F-score	94.91	93.74	93.86	**97.34**	95.04	93.74	92.83
	Recall	91.22	88.22	88.42	**94.81**	93.61	88.22	86.63
C3	Precision	65.27	68.58	64.3	70.84	66.47	**80.95**	71.78
	Jaccard-index	48.2	51.09	42.69	**55.33**	50.98	51.46	50.76
	F-score	65.05	67.63	59.84	**71.25**	67.53	67.95	67.34
	Recall	64.83	66.7	55.95	**71.66**	68.62	58.55	63.42
C4	Precision	97.93	97.54	**98.56**	97.39	86.45	86.33	86.69
	Jaccard-index	96.21	**96.99**	96.82	96.65	77.68	82.02	79.92
	F-score	98.07	**98.47**	98.38	98.29	87.44	90.12	88.84
	Recall	98.21	**99.43**	98.21	99.21	88.45	94.26	91.1
C5	Precision	97.6	98.36	96.48	**98.39**	93.74	91.21	90.65
	Jaccard-index	96.71	97.2	95.78	**97.23**	87.45	84.71	84.63
	F-score	98.33	98.58	97.85	**98.6**	93.30	91.72	91.67
	Recall	99.06	98.80	**99.25**	98.80	92.87	92.23	92.72
C6	Precision	**100.00**	**100.00**	99.75	99.97	**100.00**	99.70	99.44
	Jaccard-index	99.88	99.84	99.42	99.89	**100.00**	99.45	98.67
	F-score	99.94	99.92	99.71	99.94	**100.00**	99.72	99.33
	Recall	99.88	99.84	99.67	99.92	**100.00**	99.75	99.22
C7	Precision	98.16	98.35	98.50	**99.24**	98.82	98.41	98.19
	Jaccard-index	97.40	97.70	95.02	**99.13**	98.35	98.25	97.84
	F-score	98.68	98.83	97.45	**99.56**	99.17	99.12	98.91
	recall	99.21	99.32	96.42	**99.89**	99.52	99.83	99.63
C1–7	OA	75.91	78.04	76.34	80.01	76.79	**81.69**	79.52

Figure 12 shows a few isolated pixels within the classified maps for Class 2 and Class 4–7. However, the VA model exhibits better performance for the objects in Class 2 and Class 4–7 than the MSS, MRS, and SSC methods. The reason for this is that the misclassified pixels lie on the boundaries between objects, and the boundaries between objects are determined prior to classification. Additionally, the boundary changes during classification are limited to merging segments within the same class. In contrast, the geometry of the VA model allows the objects to affect the geometry of each other at both pixel and object levels, as demonstrated in Figure 6.

The OA of the SVM method and VA model is 76.34% and 80.01%, respectively. This can be due to the high spatial distribution of C1 pixels within the C3 class in the SVM map and vice versa. Moreover, the SVM method only applies spectral information to label pixels. As we expected, the VA model improves the OA accuracy of the k-means by almost 4.5%. The VA model also demonstrates better performance than the SSC method, with more than 2% improvement. The main reason for this improvement is the geometry of the VAs, which gives them the power to reduce the noises spatially in the image. Although the OA value of the MSS method is 2% more than the VA method, the recall values of the

VA method for C2–C7 objects are higher than the MSS method. This indicates that the proposed geometry by the VA can effectively remove the speckled noises.

We extracted the geometry information of the C1 and C3 classes in the classified maps to calculate the P/A ratio. In contrast to the previous experiments, the ground pixels are concentrated in one part of the GT map and cover the whole C1 and C3 objects in the image (Figure 8f).

From Table 6, it can be clearly observed that the VA map has a lower level of fragmentation compared to the SSC method. For example, the P/A value for the C1 class is 0.09 in the VA map, which is 0.21 in the SSC map. This is because the small and isolated patches in the classified maps, i.e., Figure 12c,d, increase the perimeter of features in the image. To address the C1 objects in the image, the VA method generated only 15 patches, the lowest number among all methods in Table 6. This can be explained by the fact that the VA model classifies the pixels in the scene instead of the feature space. Additionally, the unified geometric structure of the VA model allows them to remove the small holes within the objects in the image as shown in Figures 4 and 5.

Table 6. Comparison between the P/A ratio of SSC, VA, MSS, and MRS approach.

Method	SSC		VA		MSS		MRS	
Class	C1	C3	C1	C3	C1	C3	C1	C3
Number	316	268	15	19	15	23	34	29
Area (m^2)	104,838.89	119,822.50	102,764.97	126,549.36	147,059.32	93,143.21	128,887.22	114,776.89
Perimeter (m)	21,978.94	21,326.96	9577.48	9372.53	9001.97	7772.15	9153.52	8030.08
P/A ratio	0.21	0.18	0.09	0.07	0.06	0.08	0.07	0.07

For all the above experiments, the processing was carried out using Repast running on an Intel CPU at 3.40 GHz with 16 GB of memory. The average time for the VA method to classify the whole image was 450 s. Although the proposed algorithm is a little slow, the level of automation of the proposed method is higher than the MSS, MRS, MF, and SSC methods, which are formulated based on geometric parameters specified by a human expert. As the neighbourhood distances are automatically defined by the VAs, not a human expert, the results of the proposed method can be more consistent and robust.

5. Conclusions

Conventional k-means methods use pixels in isolation to cluster an image. Because of this, they lack the ability to deal with limitations, such as salt-and-pepper noise. This drawback is usually addressed through the use of a combination of spectral and spatial information. To extract spatial information around each pixel, classification algorithms generally apply a static geometry formulated based on a local fixed window or an irregular polygon. The properties of the geometry are generally defined by an expert user. The algorithm then applies this spatial information and the spectral information of pixels to classify the image or relabel pixels. The primary assumption in this structure is that there is a unique mathematical formula that can be applied to formulate spatial connectivity between all features in an image. If this assumption is violated, e.g., when there are complex or heterogeneous features in a scene, the clustering results may contain significant amounts of misclassifications.

In this paper, we presented a geometry-led approach, constructed based on the VA model, to remove the salt-and-pepper noise without setting any geometric parameters, e.g., scale, in unsupervised image classification. In the presented algorithm, we applied a unified and dynamic geometry, instead of using a predefined geometry or a hierarchical structure, to create an actual flexible neighbourhood system for extracting spatial information and removing speckle noise. The experimental results demonstrated the desirable performance of the VA model. For example, the P/A values in Tables 2, 4 and 6 highlighted that the VAs increase spatial connectivity between pixels and provide a better visual appearance

by simultaneously modelling the exterior and interior boundaries of objects in the images. The results in Tables 1, 3 and 5 also indicate better performance of the VA model to remove noise than the MSS and MRS methods.

For future research, we plan to improve the performance of the proposed method by reducing the processing time. This can be performed by adding the learning capabilities to VAs in order to find the shortest route to determine the boundary lines, which can help the algorithm to save memory and reduce simulation time. Another area for future research can be to adapt the proposed method for object extraction from remotely sensed imagery, e.g., road extraction.

Author Contributions: K.B.: conceptualisation, methodology, formal analysis, validation, software, writing—original draft preparation. A.B.M.: supervision, methodology, software, review and editing. A.N.H.: software, review and editing and P.S.: supervision, methodology. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tso, B.; Olsen, R.C. Combining spectral and spatial information into hidden Markov models for unsupervised image classification. *Int. J. Remote Sens.* **2005**, *26*, 2113–2133. [CrossRef]
2. Madhu, A.; Kumar, A.; Jia, P. Exploring Fuzzy Local Spatial Information Algorithms for Remote Sensing Image Classification. *Remote Sens.* **2021**, *13*, 4163. [CrossRef]
3. Tyagi, M.; Bovolo, F.; Mehra, A.K.; Chaudhuri, S.; Bruzzone, L. A context-sensitive clustering technique based on graph-cut initialisation and expectation-maximisation algorithm. *IEEE Geosci. Remote Sens. Lett.* **2008**, *5*, 21–25. [CrossRef]
4. Madubedube, A.; Coetzee, S.; Rautenbach, V. A Contributor-Focused Intrinsic Quality Assessment of OpenStreetMap in Mozambique Using Unsupervised Machine Learning. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 156. [CrossRef]
5. Chi, M.; Feng, R.; Bruzzone, L. Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. *Adv. Space Res.* **2008**, *41*, 1793–1799. [CrossRef]
6. Ragettli, S.; Herberz, T.; Siegfried, T. An Unsupervised Classification Algorithm for Multi-Temporal Irrigated Area Mapping in Central Asia. *Remote Sens.* **2018**, *10*, 1823. [CrossRef]
7. Ghaffarian, S.; Ghaffarian, S. Automatic histogram-based fuzzy C-means clustering for remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2014**, *97*, 46–57. [CrossRef]
8. Tarabalka, Y.; Chanussot, J.; Benediktsson, J.A.; Angulo, J.; Fauvel, M. Segmentation and classification of hyperspectral data using watershed. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Boston, MA, USA, 2008; pp. 652–655.
9. Zheng, N.; Zhang, H.; Fan, J.; Guan, H. A fuzzy local neighbourhood-attraction-based information c-means clustering algorithm for very high spatial resolution imagery classification. *Remote Sens. Lett.* **2014**, *5*, 1328–1337. [CrossRef]
10. Zhang, H.; Zhai, H.; Zhang, L.; Li, P. Spectral–Spatial Sparse Subspace Clustering for Hyperspectral Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3672–3684. [CrossRef]
11. Cui, W.; Zhang, D.; He, X.; Yao, M.; Wang, Z.; Hao, Y.; Li, J.; Wu, W.; Cui, W.; Huang, J. Multi-Scale Remote Sensing Semantic Analysis Based on a Global Perspective. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 417. [CrossRef]
12. Li, N.; Huo, H.; Zhao, Y.M.; Chen, X.; Fang, T. A spatial clustering method with edge weighting for image segmentation. *Geosci. Remote Sens. Lett.* **2013**, *10*, 1124–1128.
13. Miao, Z.; Shi, W. A New Methodology for Spectral-Spatial Classification of Hyperspectral Images. *J. Sens.* **2016**, 1–12. [CrossRef]
14. Dzung, L. Pham, Spatial Models for Fuzzy Clustering. *Comput. Vis. Image Underst.* **2001**, *84*, 285–297.
15. Hay, G.J.; Castilla, G.; Wulder, M.A.; Ruiz, J.R. An automated object-based approach for the multiscale image segmentation of forest scenes. *Int. J. Appl. Earth Obs. Geoinf.* **2005**, *7*, 339–359. [CrossRef]
16. Tian, J.; Chen, D.M. Optimisation in multi-scale segmentation of high-resolution satellite images for artificial feature recognition. *Int. J. Remote Sens.* **2007**, *28*, 4625–4644. [CrossRef]
17. Feizizadeh, B.; Garajeh, M.K.; Blaschke, T.; Lakes, T. An object based image analysis applied for volcanic and glacial landforms mapping in Sahand Mountain, Iran. *Catena* **2021**, *198*, 105073. [CrossRef]

18. Kurtz, C.; Passat, N.; Gançarski, P.; Puissant, A. Multi-resolution region-based clustering for urban analysis. *Int. J. Remote Sens.* **2010**, *31*, 59415973. [CrossRef]

19. Gençtav, A.; Aksoy, S.; Önder, S. Unsupervised segmentation and classification of cervical cell images. *Pattern Recognit.* **2012**, *45*, 4151–4168. [CrossRef]

20. Hesheng, W.; Baowei, F. A modified fuzzy C-means classification method using a multiscale diffusion filtering scheme. *Med. Image Anal.* **2009**, *13*, 193–202.

21. Fang, B.; Chen, G.; Chen, J.; Ouyang, G.; Kou, R.; Wang, L. CCT: Conditional Co-Training for Truly Unsupervised Remote Sensing Image Segmentation in Coastal Areas. *Remote Sens.* **2021**, *13*, 3521. [CrossRef]

22. Yin, S.; Qian, Y.; Gong, M. Unsupervised hierarchical image segmentation through fuzzy entropy maximization. *Pattern Recognit.* **2017**, *68*, 245–259. [CrossRef]

23. Hay, G.J.; Castilla, G. Geographic Object-Based Image Analysis (GEOBIA): A new name for a new discipline. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 75–89.

24. Baatz, M.; Hoffmann, C.; Willhauck, G. Progressing from object-based to object-oriented image analysis. In *Object-Based Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 29–42.

25. Blaschke, T.; Hay, G.J.; Kelly, M.; Lang, S.; Hofmann, P.; Addink, E.; Feitosa, R.Q.; van der Meer, F.; van der Werff, H.; van Coillie, F. Geographic object-based image analysis–towards a new paradigm. *ISPRS J. Photogramm. Remote* **2014**, *87*, 180–191. [CrossRef] [PubMed]

26. Troya-Galvis, A.; Gançarski, P.; Berti-Équille, L. Remote sensing image analysis by aggregation of segmentation-classification collaborative agents. *Pattern Recognit.* **2018**, *73*, 259–274. [CrossRef]

27. Hofmann, P.; Lettmayer, P.; Blaschke, T.; Belgiu, M.; Wegenkittl, S.; Graf, R.; Lampoltshammer, T.J.; Andrejchenko, V. Towards a framework for agent-based image analysis of remote-sensing data. *Int. J. Image Data Fusion* **2015**, *6*, 115–137. [CrossRef]

28. Torrens, P.M.; Benenson, I. Geographic automata systems. *Int. J. Geogr. Inf. Sci.* **2005**, *19*, 385–412. [CrossRef]

29. Hammam, Y.; Moore, A.; Whigham, P. The dynamic geometry of geographical vector agents. *Comput. Environ. Urban Syst.* **2007**, *31*, 502–519. [CrossRef]

30. Chang, C.C.; Lin, C.J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Systems Technol.* **2011**, *3*, 1–27. [CrossRef]

31. Moore, A. Geographical Vector Agent-Based Simulation for Agricultural Land Use Modelling. In *Advanced Geosimulation Models*; Marceau, D., Benenson, I., Eds.; 2011; pp. 30–48. Available online: http://www.casa.ucl.ac.uk/Advanced%20Geosimulation%20Models.pdf (accessed on 15 November 2021).

32. Howe, T. Containing agents: Contexts, projections, and agents. In *Proceedings of the Agent Conference on Social Agents: Results and Prospects*; Argonne National Laboratory: Argonne, IL, USA, 2006.

33. Borna, K.; Moore, A.B.; Sirguey, P. Towards a vector agent modelling approach for remote sensing image classification. *J. Spat. Sci.* **2014**, *59*, 283–296. [CrossRef]

34. Borna, K.; Moore, A.B.; Sirguey, P. An Intelligent Geospatial Processing Unit for Image Classification Based on Geographic Vector Agents (GVAs). *Trans. GIS* **2016**, *20*, 368–381. [CrossRef]

35. Mahmoudi, F.T.; Samadzadegan, F.; Reinartz, P. Object oriented image analysis based on multi-agent recognition system. *Comput. Geosci.* **2013**, *54*, 219–230. [CrossRef]