

Article

Assessment of the Potential of UAV Video Image Analysis for Planning Irrigation Needs of Golf Courses

Alberto-Jesús Perea-Moreno ^{1,*}, María-Jesús Aguilera-Ureña ¹, José-Emilio Meroño-De Larriva ² and Francisco Manzano-Agugliaro ^{3,4}

¹ Department of Applied Physics, University of Cordoba, CEIA3, Campus de Rabanales, 14071 Córdoba, Spain; fa1agurm@uco.es

² Department of Graphic Engineering and Geomatics, University of Cordoba, CEIA3, Campus de Rabanales, 14071 Córdoba, Spain; jemerono@uco.es

³ CIAIMBITAL (Research Center on Agricultural and Food Biotechnology), University of Almeria, 04120 Almeria, Spain; fmanzano@ual.es

⁴ Department of Engineering, University of Almeria, CEIA3, 04120 Almeria, Spain

* Correspondence: aperea@uco.es; Tel.: +34-957-212-633; Fax: +34-957-212-068

Academic Editors: Hongjie Xie and Xianwei Wang

Received: 9 October 2016; Accepted: 2 December 2016; Published: 8 December 2016

Abstract: Golf courses can be considered as precision agriculture, as being a playing surface, their appearance is of vital importance. Areas with good weather tend to have low rainfall. Therefore, the water management of golf courses in these climates is a crucial issue due to the high water demand of turfgrass. Golf courses are rapidly transitioning to reuse water, e.g., the municipalities in the USA are providing price incentives or mandate the use of reuse water for irrigation purposes; in Europe this is mandatory. So, knowing the turfgrass surfaces of a large area can help plan the treated sewage effluent needs. Recycled water is usually of poor quality, thus it is crucial to check the real turfgrass surface in order to be able to plan the global irrigation needs using this type of water. In this way, the irrigation of golf courses does not detract from the natural water resources of the area. The aim of this paper is to propose a new methodology for analysing geometric patterns of video data acquired from UAVs (Unmanned Aerial Vehicle) using a new Hierarchical Temporal Memory (HTM) algorithm. A case study concerning maintained turfgrass, especially for golf courses, has been developed. It shows very good results, better than 98% in the confusion matrix. The results obtained in this study represent a first step toward video imagery classification. In summary, technical progress in computing power and software has shown that video imagery is one of the most promising environmental data acquisition techniques available today. This rapid classification of turfgrass can play an important role for planning water management.

Keywords: water management; golf course; memory-prediction theory; object-based classification; unmanned aerial vehicle

1. Introduction

As a case of precision agriculture, golf courses can be considered; this is called precision turfgrass in the literature [1]. The huge dimensions of maintained turfgrass can be highlighted by the fact that it is estimated to cover 20 million ha in the USA [2]. Spatio-temporal variation of soil, climate, plants and irrigation requirements are new challenges for precision agriculture and, above all, complex turfgrass sites [3]. The irrigation of golf courses is a major concern in this crop maintenance, especially in a Mediterranean climate, both in the USA and in Europe [4]. Golf courses in the southwestern United States are rapidly transitioning to reuse water (treated sewage effluent), as municipalities

provide price incentives or mandate the use of reuse water for irrigation purposes [5]. So, when reuse water of poor quality is used, as on golf courses in the arid southwestern United States, proper irrigation management is critical [6], so greenkeepers should pay attention to irrigation strategies employed on reuse water irrigated golf courses to properly manage for higher nitrogen and salt loads. In Spain, it is estimated that water consumption for a golf course is 6.727 m³/ha per year (this is due to the use of poor water, 2.5 dS/m) [7].

Recently, unmanned aerial vehicles (UAVs) have provided a technological breakthrough with potential application in PA [8,9]. UAVs enable the quick production of cartographic material because they rely on different technologies, including cameras, video and GPS (Global Positioning System) [10]. Even though an UAV has very restricted, heavy limitations, the minimization of the sensors during the last year is allowing the use of lighter vehicles, or the use of more features and sensors to a given platform [11]. The opportunity offered by UAVs to observe the world from the sky provides the opportunity to study crops or turfgrass from an unusual viewpoint, allowing the visualization of details that cannot be easily seen from the ground [12,13].

Regarding agricultural purposes, aerial photography and colour video from UAVs presents an alternative to imagery from satellite and aerial platforms [14], which are often difficult to obtain or expensive [15]. Hassan et al. [15] highlight the problem of conventional methods in the classification process, using high resolution images to overcome or minimize the difficulty in classification of the mixed pixel areas. A huge number of applications are achieved using UAVs to monitoring the health of crops through spectral information, e.g., stressed or damaged crops change their internal leaf structure which could be rapidly detected by a thermal infrared sensor [16], therefore, this information is very important to detect stress such as water and nutrient deficiency in growing crops [17].

On the other hand, texture measurements from images obtained by UAVs have been integrated in object-oriented classification, specifically in the classification and management of agricultural land cover [18]. Likewise, there are studies that demonstrate the feasibility of using UAVs with thermal multispectral cameras for estimating crop water requirements, determining the ideal time for watering and saving water consumption without affecting productivity [19–21]. Therefore, the technology is versatile and capable of producing very useful cartographic material for working with PA; the technique also facilitates working with aerial photography in addition to LIDAR (Laser Imaging Detection and Ranging) or video cameras [22,23]. UAVs on golf courses have been used for some time to monitor certain agronomic variables, such as nitrogen [24], and should be considered as a valuable tool to monitor plant nutrition. In this case study, a rapid classification of turfgrass, among others, can play an important role to determine the water requirements of the different areas in order to plan water use.

For this purpose, information of important use can be analyzed and extracted from the images through the employment of powerful and automatic software. The object-based image classification techniques are applied not only for a high level of adaptability but automation as well. These techniques overcome some limitations of pixel-based classification by creating objects on the image through segmentation, using adjacent pixels with a spectral similarity [25]. Subsequently, object-based classification combines spectral contextual information for these objects to perform more complex classifications. These techniques have been successfully applied to images obtained by UAVs in agricultural [26,27], aquatic ecosystems [28] and urban [29] areas.

Therefore, for golf courses, irrigation need planning, especially if it is employed in large areas, and has to be monitored more frequently than other crops. UAVs, due to low cost and fast response time, are the technology that allows this monitoring. A monitoring system based on the video image analysis and classification, will allow a real-time control of crops. Thus, this research is the first step to show the technical viability of real-time control of crops.

Thus, given the positive results previously obtained in the classification of images and given that the applications developed using the HTM algorithm are capable of analyzing video images, the objective of the current study is to develop a recognition methodology for golf courses in real-time

using video images taken by an UAV based in a HTM for possible application in planning irrigation needs in order to maximize the water use efficiency and help to plan water requirements of reuse water.

2. Material and Methods

2.1. UAV and Sensor Description

The material used in this work included images taken by an UAV DJI Phantom 2 Vision+ (Figure 1a) with a flight control system Naza-M V2, that has a range of 700 m and an altitude of 300 m; a HD integrated camera; and a 3-axis gimbal correcting movements in any axis and direction [30]. The Phantom 2 Vision+ is a rigid quadcopter with a maximum ascent rate of 6 m/s, a maximum descent rate of 2 m/s and a maximum flight speed of 15 m/s.

The camera of the Vision+ (Figure 1b) has a 140° FOV (Field of View), F2.8 connected to a 2.3'' sensor with 14 megapixels that can capture images in Adobe DNG and JPEG format as well as recording 1080-pixel and 30-fps videos or recording in a slow camera mode at 60-fps 720-pixel resolution [30]. The equipment also features a streaming video and telemetry data with a range of up to 700 m to a phone, tablet or computer and has a 5200-mAh lithium battery that can hold the quadcopter in the air for up to 25 min. The operator can control the camera using Wi-Fi to manage pan, tilt and camera light sensitivity, video or image modes. The Wi-Fi computer camera system is a very important element that allows for real-time viewing of everything being seen by the camera and the obtainment of video images in real-time.

The equipment also includes an inertial sensor and a barometric altimeter to measure altitude and latitude.



Figure 1. Details of the DJI Phantom 2 Vision+. (a) General image of the quadcopter; and (b) the details of the camera.

2.2. Study Site

The first stage of this study is to propose a new methodology for analysing geometric patterns of video data acquired from UAVs using a new Hierarchical Temporal Memory (HTM) algorithm.

The information used in this research during the training phase includes simple and short videos, as these videos represent a first step in the integration of UAV video cameras into this technology and the search for the most suitable videos for the proposed purposes. The analyzed patterns to check the accuracy of the method were grapes (*Vitis vinifera*; see Figure 2) and other non-agricultural uses, namely urban and wood areas.

For each of these categories, 300 training videos and 150 testing videos with a total duration of 60 min were used. The videos were obtained in different areas of Redwood City, San Mateo County, California, United States (37°30.128' N 122°12.758' W; Figure 3).



Figure 2. Sample image of the video sequences studied.



Figure 3. Location map.

2.3. HTM Methodology

In recent years, the technology involved in remote sensing and object recognition has considerably advanced [31,32], with diverse applications ranging from recognition and vehicle classification [33] to the facial recognition of individuals [34]. Studies on detection and object recognition can be classified into two categories: keypoint-based object detection [35] and hierarchical and cascaded classifications [36]. Parallel to this development, a new technology applicable to the classification of digital pictures emerged: the Hierarchical Temporal Memory (HTM) learning algorithm. This classification technology is based on both neural networks and Bayesian networks but involves a particular algorithm based on a revolutionary model of human intelligence—the memory-prediction theory developed by Jeff Hawkins [37]. This theory is based on the workings of the human cerebral cortex, which has a structure in the form of “layers” in which information flows bidirectionally from the senses to the brain. From this operating hierarchy, a hypothesis of how the human mind works is created. The key point of this algorithm is found in the duality of the information received. All information we perceive has a spatial component and a temporal one; information is received

by the human brain not as an isolated pattern but as a succession of patterns. The cerebral cortex stores the patterns that we perceive and how they are ordered in time. In light of that concept, the memory-prediction theory states that the cerebral cortex stores the new patterns and their evolution over time so that once these sequences stabilize, the brain can make predictions (or inferences) enabling it, without observing a full sequence, to know what pattern it is observing because it knows the sequence in which the patterns occur over time [37].

Thus, this new technology developed by Jeff Hawkins not only presents a new model of how human intelligence functions but also models a neural network system capable of emulating this theory. This classification technology is not specific to image analysis but is versatile for any type of information (from medical information to economic data), with a dual role: learning and pattern recognition in data flows and classifying unknown data according to the training received. Currently, we can find this technology integrated into the free software application NuPIC developed by NUMENTA® (Redwood City, CA, USA), which is used to classify data streams [38]. These data can be of many types, ranging from sign language [39] to eye retinal images for biomedical purposes [40]. There are open areas of research using HTM as a classifier for land planning, which is where our work focuses. In a previous study, Perea et al. [41] conducted an analysis of high-resolution images for classification and land planning in agricultural environments; starting from images from a UltracamD® (Graz, Austria) photo sensor of a region of southern Spain, classification results were obtained that recognized the ground cover up to 90.4%. In a similar fashion, using HTM in the recognition and object-oriented classification, the technology was successfully applied in the recognition of urban areas and green areas; the classification results obtained were approximately 93.8% [42].

Objects with a hierarchical structure, in both space and time, compose the world; this same concept is used by HTM to generate a series of interconnected nodes organized in a tree hierarchy [43]. Thus, the HTM presents a hierarchical structure either in space or time and represents the structure of the world [44].

The HTM learning algorithm implemented in the HTM Camera Toolkit free Application Programming Interface (API) was used in this experiment. This API allows easy implementations of HTM learning algorithms using real world images. Although this API can be used in a variety of contexts, in this paper we focus only on visual recognition applications (i.e., inputs are UAV videos). This API is built and configured by writing Python scripts, allowing researchers to design and configure the hierarchy of nodes based on their input data. To improve the accuracy based on node parameters configuration it is necessary to work with iterations.

As commented before, the principal objective of this first stage in this investigation is to propose a new methodology for analysing geometric patterns of video data acquired from UAVs using a new Hierarchical Temporal Memory (HTM) algorithm. For this purpose, the parameterization and structure of the HTM algorithm for learning and inference have been analyzed and constructed.

Figure 4 shows the overall methodology for HTM design and implementation. There are five phases in this methodology, from the definition and configuration of the data and HTM network to the training and its evaluation.

Once the data to be used have been defined, two steps were necessary to create this network: the creation of the architecture using the Python programming language and the formation of a set of training patterns.

Based on the experience of the research group in previous work [41,42,45], the HTM network was defined in three levels: the first two levels are composed by two sub-levels (a sub-level which analyses the spatial component and another sub-level that analyses the temporal component), and finally there is a classifier which sorts the images into common categories. The level 1 or input level is composed of 8×8 pixel input nodes, each associated to a single pixel. Nodes from the first level go through the raw image and receive a characteristic of the training pattern image, creating an entry vector formed by digital levels of 8×8 pixels. Level 2 is composed of 16 nodes that receive the information from the previous level; therefore, each level 2 node is formed by four primary child nodes (arranged in

a 2×2 region). Finally, level 3 or higher comprises a single node, and it has 16 child nodes (arranged in a 4×4 region) and a receptive field of 64 pixels. In Figure 5, the downward connection of one node per level is shown. This system operates in two phases: the training phase and the inference phase. During the training phase, the network is exposed to training patterns and builds a model that categorizes patterns. During the inference phase, new patterns will be distributed in these categories. All nodes (except the initial node) process information in the same manner and consist of two modules: temporal and spatial [44]. Understanding an HTM node involves understanding the operation of these modules during the learning and training phases.

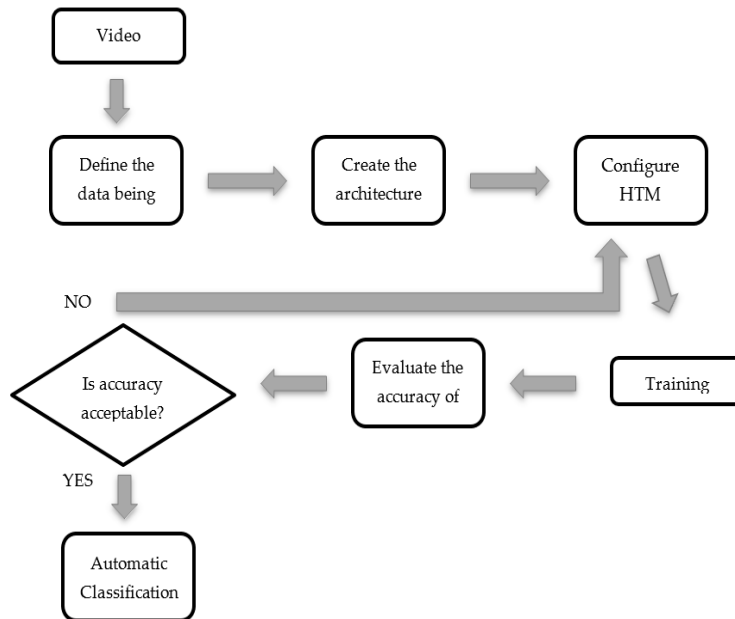


Figure 4. Overall methodology diagram.

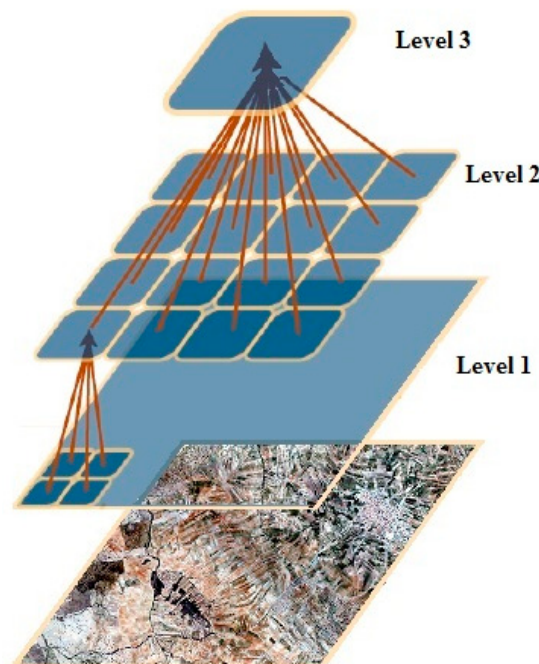


Figure 5. Details of the HTM structure. Level 1 is composed of 64 nodes; Level 2 is composed of 16 nodes and Level 3 comprises a single node.

2.3.1. Training Phase

During the training phase, the spatial module learns to classify input data based on the spatial coincidence of the elements that compose them. The input vector is compared with other vectors already stored. The exit of the spatial module (temporal module entrance) occurs in terms of their matches and can be seen as a pre-processing stage for the temporal module, simplifying entry. The temporal module learns temporary groups, which are groups of coincidences that frequently occur [44].

Spatial Module

The spatial modules of the input nodes receive raw data from the sensor; spatial modules of the upper nodes receive the output data from their lower nodes. The input of the spatial module in the upper layers is the concatenation of the order set by the output of the nodes below. Its input is represented by a series of vectors, and the function of the spatial module is to build a matrix (match matrix) of input vectors that have recently occurred. There are several algorithms for the spatial modules, such as the Gaussian and Product algorithms. The Gaussian algorithm is used for nodes in the input level, and the top nodes of the hierarchy use the spatial module Product.

The Gaussian algorithm compares the input vector without dealing with the existing matches in the match matrix. If the Euclidean distance between the input vector and the existing match is sufficiently small, then the entry is considered as the same match, and the match count is incremented and stored in the memory. The distance between an input vector and previously stored vectors is:

$$d^2(x, w_j) = \sum_{i=1}^D (x_i - w_j)^2 \quad (1)$$

where D is the dimension of the vector (64 in the first level), x_i is the i th element of the input vector and w_j is the position i of the vector j in the match matrix W . The match threshold of an input vector to an existing match is the Maximum distance parameter.

The product algorithm calculates the probability of similarity ($belief_i$), Equation (2), between an input in the inference and a vector that had been previously memorized by the spatial module:

$$belief_i = \prod_{j=1}^{nchildren} y_i(child_j) * x(child_j) \quad (2)$$

where $nchildren$ is the number of secondary nodes (previous level) that the parent node has, x is the input vector, y_i are the vectors previously stored by the spatial module and $(child_j)$ is the part of a vector obtained from $nchildren$ secondary nodes.

Temporal Module

The temporal module forms groups of matches in time, called temporal groups. Subsequently, a temporal match matrix is built. After the training phase, the temporal module uses this matrix to create the temporal groups. This module uses the sum algorithm, which takes the best representations of all groups to classify new input patterns during inference. When a new input vector is presented during the training phase, the spatial module represents the input vector as one of the learned matches. This process increases the elements (j, i) of the temporal match matrix and is controlled by the *transitionMemory* parameter. This increment (It) is calculated as follows, Equation (3):

$$It = transitionMemory - t + 1 \quad (3)$$

where t is the training; the HTM time is in seconds between the current match and the past match.

2.3.2. Inference Phase

After training a node, the network transitions to the inference mode. When the complete network is trained, all of the nodes are in the inference state, and the network is capable of performing inference with new input patterns. Initially, a probability distribution is generated for the categories that were used during training.

Spatial Module

When an input pattern arrives to the spatial module, the network will generate a distribution of beliefs about the categories that have been created in the training phase. Both the Gaussian spatial module and the Product spatial module work differently during the inference stage, but both turn an input vector into a belief vector around the matches.

In the Gaussian spatial module, the distance between an input vector x and each of the trained matches w_j is calculated using Equation (1).

This distance becomes a probability vector considering x as a random sample drawn from a set of multi-dimensional Gaussian probability distributions, all of them based in one of the trained matches. All of these distribution probabilities have the same constant variance in all dimensions, controlled by the *Standard Deviation (SD)* parameter, which is the square root of the variance. Each element i of the probability vector b , which represents the probability of the input vector x having the same cause as the match i , is calculated using the following equation:

$$y_i = \exp \left\{ -\frac{d^2(x, w_j)}{2SD^2} \right\} \quad (4)$$

where d^2 is defined in Equation (1) and w_j is the match of the position j in the match matrix W .

The algorithm of the Product spatial module divides the input vector at the outputs of each one of its subgroups. The algorithm uses the dot product with the same parts of the match and then calculates the products of these numbers, resulting in a probability vector element in matches in the match matrix.

Temporal Module

During the inference phase, the temporal module receives a probability vector concerning the matches in the spatial module. Subsequently, the module calculates the probability distribution of the groups. A choice is made between two different algorithms in the temporal module during the inference: *maxProp* and *sumProp*, controlled by the *PoolerAlgorithm* time parameter. These algorithms are defined in detail in [46].

2.4. HTM Design and Implementation

As commented before, we used the HTM Camera Toolkylt API, developed by Numenta® (Redwood City, CA, USA), in order to design the HTM network used in this investigation.

Once the network has been built, the second step is to configure the information handling and training process. Here, the key parameter is the number of iterations performed with the training images. In this case, 2000 iterations were performed at three levels. Experiments have demonstrated that increasing up to double the number of iterations (4000) does not result in a significant increase in the accuracy of the analysis [41,42,45].

In Table 1, the most relevant parameters of the network-training phase are presented, as are the starting values of the core network as recommended by [43].

Figure 6 presents images of each level of the network structure. Each image that is contained in a video is analyzed by the network. As a pre-treatment, all frames are rescaled to a specific resolution as many times as the parameter *ScaleCount* indicates (Figure 6a—original image, Figure 6b—rescaled

image), after which the information goes through the first level of nodes (level 1). This first level is called the S1 layer, and it uses a filter (Gabor filter) to help in recognizing input patterns and making a selection among a series of categories based on geometric and temporal similarities. To extract features and analyze texture, Gabor filters are used [47].

Table 1. Parameters used during training.

Parameter	Description	Values
maxDistance (<i>maxDist</i>)	Minimum Euclidean distance for storing a pattern as a new category, in the lower level of the training phase.	1
Scale factor (<i>ScaleCount</i>)	Number of scales of the same image that the sensor introduces into the network.	1
Spatial reference (<i>spatialRF</i>)	Size of the information reception field with respect to the total.	0.2
Temporal groups (<i>requestGroupCount</i>)	Sets the maximum number of temporal groups that will be created.	24
Spatial overlap (<i>spatialOverlap</i>)	Overlap between nodes of the same level according to the information received from child nodes.	0.5
Scale reference (<i>ScaleRF</i>)	Number of scales of which the node receives information.	2
Categories (<i>outputElementCount</i>)	Number of categories.	3

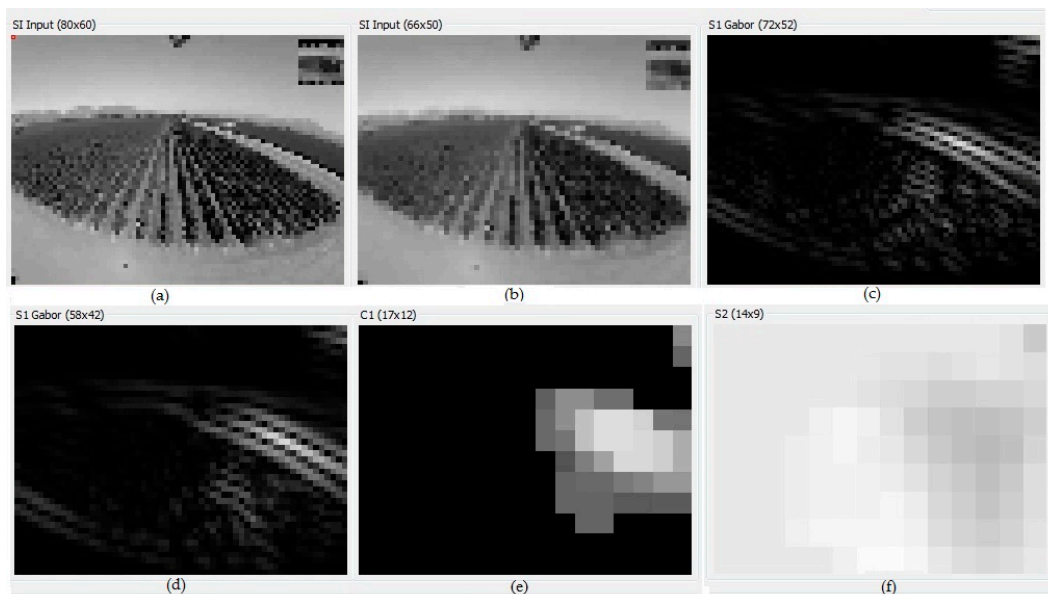


Figure 6. Operation of network levels during the training phase. (a) Input with 80×60 pixels resolution; (b) input with 66×50 pixels resolution; (c) S1 layer with 72×52 pixels resolution (Gabor filter); (d) S1 layer with 58×42 pixels resolution (Gabor filter); (e) C1 layer with 17×12 pixels resolution; and (f) S2 layer with 14×9 pixels resolution.

Due to this initial screening, we generate a database of the most common patterns and reduce the infinite number of patterns that we could receive in each image to a limited number. This level produces a set of patterns that are common or that are strongly present in the image as an output.

The input for Level 2 (Figure 6c—obtained from the image, Figure 6d—obtained from the rescaled image), designated C1 Layer (Figure 6e), is the output of the previous level (S1). Level 2 is where the clustering of time sequences occurs. In this level, grouping is performed based on the information of the previous layer, with the base patterns (equivalent to the invariant representations of the HTM theory) creating pattern sequences or pattern clusters using geometric criteria. These sequences are stored, generating more complex patterns.

The information travels up the network to level 3, called the S2 layer (Figure 6f), where information from the preceding level 2 (C1) arrives. Level 3 is where an initial classification is performed. During the training phase, a set of prototypical patterns are memorized through the sequences received from the classification made by the lower layer (C1). When the network is trained, the new data stream in this sub-layer will be compared to the memorized sequences performing an initial classification.

2.5. Inference Phase

Once the network has been trained with the data set that was provided, indicating the categories that we want it to recognize, we move to the inference phase. In the inference phase, we supply the network with a set of unknown images for it to classify according to what the network learned and memorized in the previous phase.

Figure 7 presents the system working in the inference phase. The status of any of the nodes of the different levels (Figure 7a–e) can be visualized while the network is processing the information. Finally, we have the C2 Layer, in which the process of grouping already classified patterns is repeated. This process is performed to convert the information into a probability vector, which collects the sequences with the maximum response to the classification process. Behind this last layer, we have a support vector machine (SVM) classifier. SVM, as a kernel learning method, is used for classification problems, performing a non-linear classification [48]. This classifier memorizes the categories with which we are working; these categories were defined during the training phase. This classifier is responsible for assigning the class to which each classified image belongs (Figure 7f). Once the inference stage is completed, a confusion matrix is obtained.

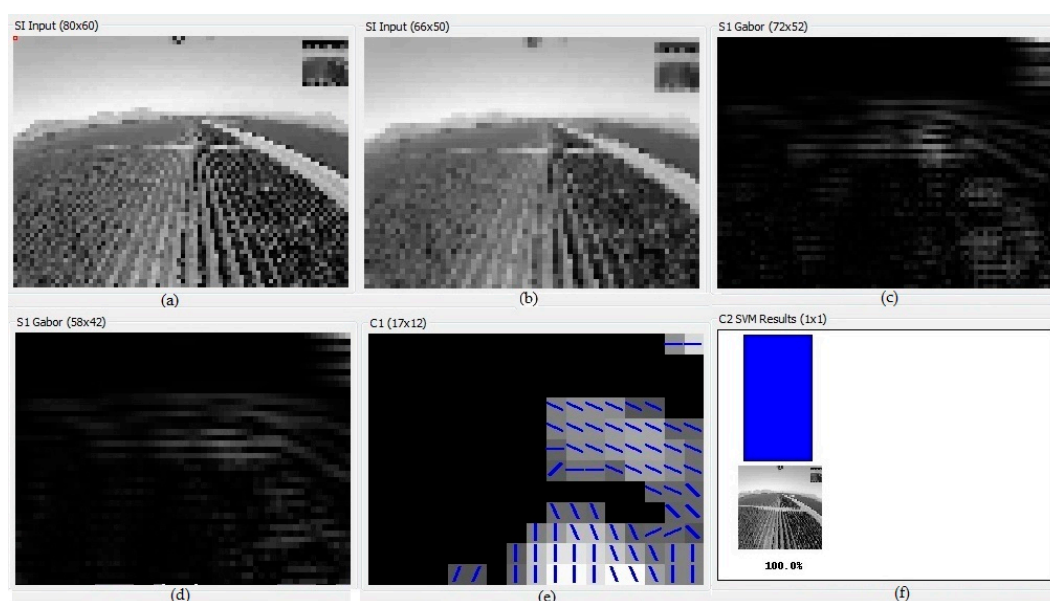


Figure 7. Operation of network level in the training phase. (a) Input with 80×60 pixels resolution; (b) input with 66×50 pixels resolution; (c) S1 Layer with 72×52 pixels resolution (Gabor filter); (d) S1 layer with 58×42 pixels resolution (Gabor filter); (e) C1 Layer with 17×12 pixels resolution; (f) SVM classifier.

3. Results and Discussion

During the experiments, internal network parameters that affect the learning process were modified, with the main goal of obtaining an optimal methodology for the recognition of video image patterns.

As mentioned above, the *maxDist* parameter defined the Euclidean distance between a known pattern and a new one, which is critical in the recognition and classification of patterns. An optimal

value is essential for the successful creation of temporal groups during the training phase. A high value of the *maxDist* parameter contributes to the formation of fewer temporal groups, which could seriously impact the total recognition accuracy. On the other hand, a low value of the *maxDist* parameter generates a high number of temporal groups, which on top of the large memory demand, also results in poor recognition performance. To avoid these undesirable effects, it is very important to evaluate the optimal value for *maxDist* to achieve the best accuracy in the classifications.

In the original configuration, the *maxDist* parameter has a starting value of 1, and the influence of this parameter on the overall accuracy values in the different classifications was studied. The *maxDist* values (Table 2) used in this experiment were defined based on the results of the initial studies performed [41,42,45].

Table 2 presents the *maxDist* parameter values with respect to the overall accuracy obtained for each of the test classifications. The maximum accuracy value was 96% and was obtained at an intermediate value for a *maxDist* of 3. After this value, there is nearly a linear drop in the overall accuracy of the classifications. This drop is due to the number of coincidences detected during the training phase and the temporal groups formed.

Table 2. Overall accuracy and average number of coincidences and temporal groups learned in the 64 bottom nodes for different values of *maxDist*.

<i>maxDist</i>	Overall Accuracy (%)	Number of Coincidences	Number of Temporal Groups
1	86.77	55.00	25.00
3	96.00	44.79	20.00
6	83.13	17.94	13.65
9	76.37	12.20	8.21
12	64.50	10.12	5.67

For the previously mentioned optimal value of *maxDist*, the Urban class was the class that obtained the largest number of misclassified frames, as seen in Table 3, whereas the Grape class reached the highest accuracy of all the classes during classification.

Table 3. Confusion matrix for the optimum value of *maxDist*.

Classes	Grapes	Urban	Woods
Grapes	985	5	10
Urban	6	913	81
Woods	8	10	982

Looking at the second and third columns of Table 2, a large number of matches was not related to a greater overall accuracy of classification, as the number of matches in input patterns might be unrealistic, classifying new similar patterns in different categories. For example, if we set a low value for the parameter *maxDist*, it is forcing the creation of many different, but similar, groups. So, several categories may correspond to the same pattern.

For the case with *maxDist* of 3, which can be considered optimal, the number of matches obtained was 44.79. On the other hand, the effect of the value of the *maxDist* parameter on the creation of temporal groups during the training phase of the network can be seen in Table 2; the smaller the *maxDist* parameter, the greater the number of temporal groups was obtained, leading similar patterns to be classified in different classes. Conversely, increasing the value of the *maxDist* parameter reduces the formation of temporal groups, an effect that is not conducive in any way to obtaining an optimal accuracy in the classification, as the images of wineries and images of forest areas are classified in the same category (Table 4). For the case with the optimal *maxDist* value of 3, the number of temporal groups obtained was 20.

The effect of the *SD* parameter on the accuracy of the classification was verified. This parameter is calculated as the square root of the *maxDist*. This value is a reasonable starting value for *SD* because

the distances between the matches are calculated as the square of the Euclidean distance instead of the normalized Euclidean distance.

Table 4. Confusion matrix for a *maxDist* value of 12.

Classes	Grapes	Urban	Woods
Grapes	600	5	395
Urban	157	767	76
Woods	407	25	568

Figure 8 presents the overall accuracy values obtained for different *SD* values. Similar to the *maxDist* parameter, there is growth in the overall accuracy value until it reaches a maximum of 96% for an *SD* value of 1.73. Smaller *SD* parameter values cause high beliefs to be assigned only to matches that are very close to the inferred pattern. Conversely, when using lower *SD* values, between 1 and 1.73, all of the matches receive high belief values independent of their distance to the inferred pattern.

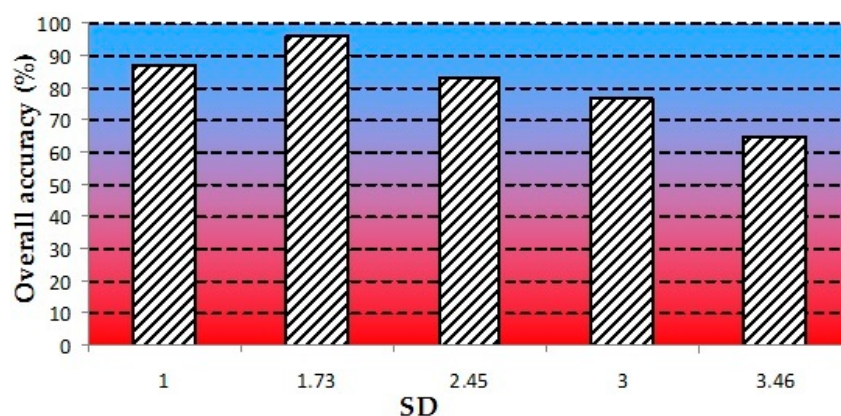


Figure 8. Overall accuracy for five setups of the *SD* parameter.

Based on the optimal *maxDist* and *SD* values previously discussed, we studied the effect of the *ScaleRF* and *ScaleOverlap* parameters on the network training and overall accuracy obtained in the classification of the images.

As mentioned above, the *ScaleRF* and *ScaleOverlap* parameters are related to the scale or the resolution of the images that are presented to the network; thus, by changing these parameters, we can vary the number of different scales of the image that are presented to the nodes and the overlap among them. This change is critical because changes in the image resolutions allow the network to extract patterns of the same image in different levels to create invariant representations (or models of stored patterns) used to classify new images.

The basic network starts from intermediate values of *ScaleOverlap* and *ScaleRF* (1 and 1, respectively). Figure 9 presents a bar chart in which the *ScaleOverlap* and the *ScaleRF* parameters are related to the overall accuracy for each case. The highest overall accuracy (97.1%) was obtained for a value of 4 for the *ScaleRF* parameter and 1 for the *ScaleOverlap*. The worst results were obtained for a *ScaleOverlap* parameter value of 0; this value creates no spatial overlap among the input patterns, worsening the training stage in the temporal module and thereby reducing the number of temporal groups formed and their time sequence.

In general, it is observed in this study that a value of 4 for the *ScaleRF* parameter optimizes the capacity of the network to extract patterns from images at different resolutions. From a value of 5, the overall classification accuracy starts to fall again.

After the analysis of the videos, the abilities of the model to learn the invariant representation of the visual pattern, to store these patterns in the hierarchy and to automatically retrieve them associatively, was verified.

For this experience, the maximum overall accuracy obtained among the different classifications made was 97.1% (Figure 10), avoiding problems related to the use of images with high spatial resolution, as in the salt-and-pepper noise effect. The salt-and-pepper effect makes it difficult to obtain and cleanly classify images, resulting in different cases for a plot where there should only be a single case.

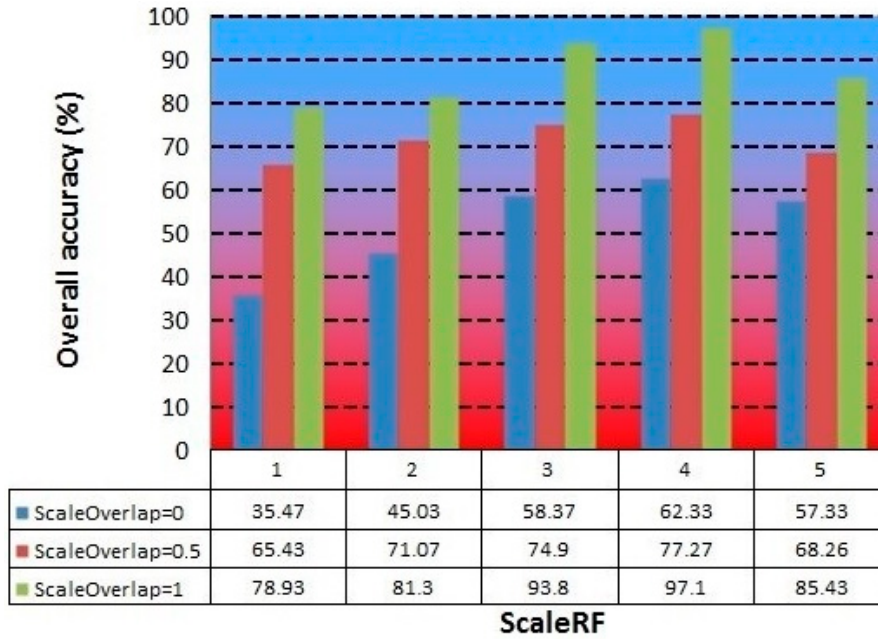


Figure 9. Overall accuracy for different values of *ScaleRF* and *ScaleOverlap*.

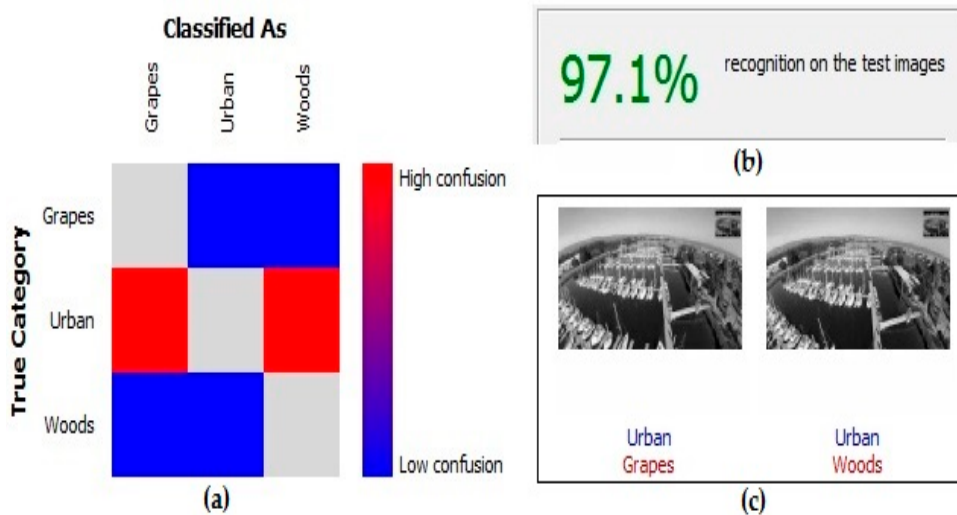


Figure 10. Classification results for the best HTM configuration presented by the HTM Camera Toolkit API (Application Programming Interface). (a) Confusion matrix; (b) overall accuracy; (c) clicking on confusion matrix the user can display the misclassified frames (for example, Urban class classified as Grapes or Urban class classified as Woods).

Comparing the results of the Confusion matrix (Table 5), lower accuracy in the Urban class is observed; there were a few misclassified frames because in the same image, two different classes could coexist, such as buildings and parks (Table 5). In 59 frames, the Urban class was classified as the Woods class, and in 11 frames, it was classified as the Grapes class. The higher accuracy obtained was for the Grapes class, where one frame was classified as Urban class and five frames as Woods class.

Table 5. Confusion matrix of the best performing system.

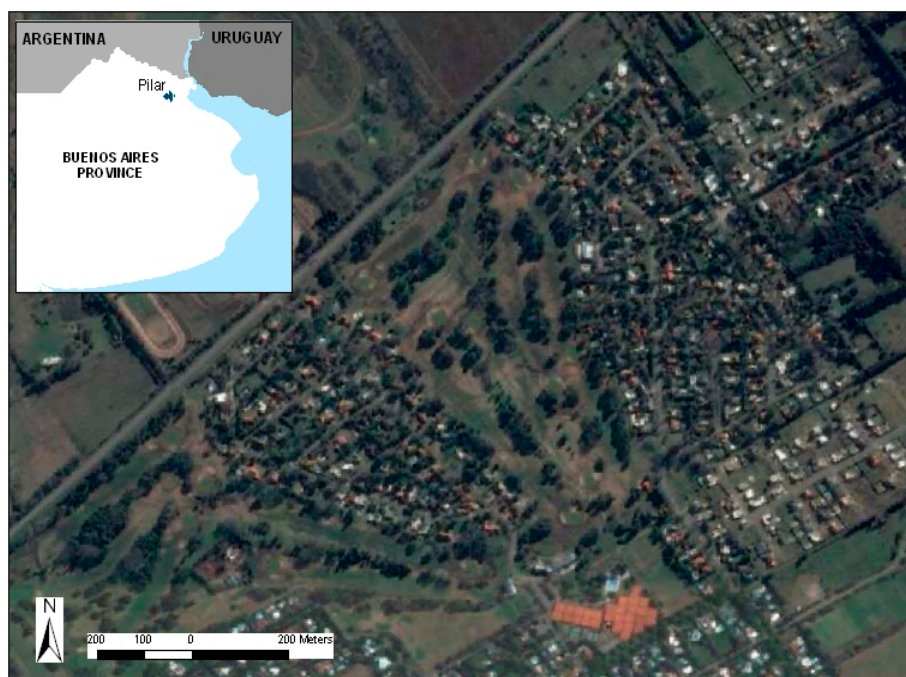
Classes	Grapes	Urban	Woods
Grapes	994	1	5
Urban	11	930	59
Woods	4	7	989
Overall accuracy	97.10%		

Case Study: Golf Course

The analyzed patterns to check the accuracy of this case study were turfgrass (see Figure 11) and other uses, namely urban, water, bunker and wood areas.

**Figure 11.** Sample image of the video sequences studied.

For each of these categories, 300 training videos and 150 testing videos with a total duration of 60 min were used. The videos were obtained in different areas of a golf course in Pilar, Buenos Aires ($34^{\circ}29'52.62''$ S; $58^{\circ}56'11.68''$ O; Figure 12).

**Figure 12.** Golf course location: case study.

Based on the optimal parameter values previously discussed, we studied the effect of and the overall accuracy obtained in the classification of the images.

For this case study, the overall accuracy obtained, using the optimal values parameters studied above, was 98.28% (Table 6).

Table 6. Confusion matrix of the best performing system.

Classes	Turfgrass	Urban	Water	Bunker	Woods
Turfgrass	986	1	3	4	6
Urban	0	980	7	3	10
Water	2	7	974	15	2
Bunker	3	3	7	986	1
Woods	3	4	0	5	988
Overall accuracy	98.28%				

We compared our results to those of other works. For example, Revollo et al. [49] develop an autonomous application for geographic feature extraction and recognition in coastal videos and obtained an overall accuracy of 95%; Duro et al. [50] used object-oriented classification and decision trees in Spot images to identify vegetal coverings and obtained an overall accuracy of 95%; Karakizi et al. [51] developed and evaluated an object-based classification framework towards the detection of vineyards reaching an overall accuracy rate of 96%.

Therefore, the accuracy value obtained from the classification using the algorithm based on HTM is similar or superior to values obtained by other authors using object-oriented classification and neural networks, which demonstrates that the methodology is appropriate for discriminating agricultural covers in real-time.

Furthermore, as an added benefit, HTM and the methodology developed in this study enable the classification and decision making to be performed in real-time. As we commented before, the operator can control the camera using Wi-Fi. The Wi-Fi computer camera system allows for real-time viewing of everything being seen by the camera, even without taking an image or video. Once the network has been trained and tested, the algorithm classifies the videos, which are received in real-time from the Wi-Fi computer camera system of the DJI Phantom 2 Vision+ (Figure 13).

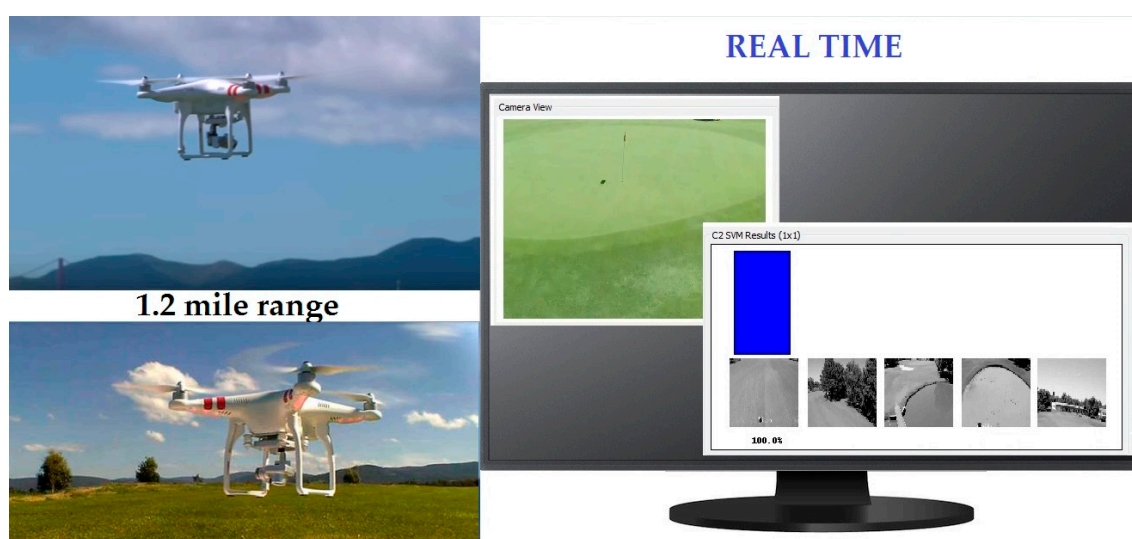


Figure 13. Display of the classification in real-time.

In contrast, in the works [1,2,51] of classical classification, post-processing work was required.

4. Conclusions

Pattern recognition is an important step in remote sensing applications for precision agriculture. Unmanned aerial vehicles (UAVs) are currently a valuable source of aerial photographs and video images for inspection, surveillance and mapping in precision agriculture purposes. This is because UAVs can be considered in many applications as a low-cost alternative to classical remote sensing. New applications in the real-time domain are expected. The problem of video image analysis taken from an UAV is approached in this paper. A new recognition methodology based on the Hierarchical Temporal Memory (HTM) algorithm for classifying video imagery was proposed and tested for agricultural areas.

As a case study of precision agriculture, golf courses have been considered, namely precision turfgrass. The analyzed patterns to check the accuracy of this case study were turfgrass (see Figure 11) and other uses, namely urban, water, bunker and wood areas.

In the classification process, based on the optimal parameter values obtained during the first stage, a maximum overall accuracy of 98.28% was obtained with a minimum number of misclassified frames. In this case study, a rapid classification of turfgrass, among others, can play an important role to determine water requirements of the different areas in order to plan water use.

Additionally, these results provide evidence that the analysis of UAV-based video images through HTM technology represents a first step for video imagery classification. As a final conclusion, the use of HTM has shown that it is possible to perform, in real-time, pattern recognition of video data images taken from an UAV. This opens new perspectives for precision irrigation methods in order to save water, increase yields and improve water, as well as indicating many possible future research topics.

Acknowledgments: The authors would like to thank all of the researchers who kindly shared the codes used in our studies. This research was supported by the Universities of Cordoba and Almeria.

Author Contributions: Alberto-Jesús Perea-Moreno and José-Emilio Meroño-De Larriva conceived and designed the experiments; Alberto-Jesús Perea-Moreno, Francisco Manzano-Agugliar, José-Emilio Meroño-De Larriva, and María-Jesús Aguilera-Ureña performed the experiments; Alberto-Jesús Perea-Moreno, José Emilio Meroño-De Larriva analyzed the data; Alberto-Jesús Perea-Moreno, Francisco Manzano-Agugliar and María-Jesús Aguilera-Ureña contributed equipment and analysis tools; Alberto-Jesús Perea-Moreno and Francisco Manzano-agugliar wrote the paper.

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

Abbreviations

The following abbreviations are used in this manuscript:

DTM	Digital terrain model
GCP	Ground control point
HTM	Hierarchical temporal memory
LIDAR	Light detection and ranging
NDVI	Normalized difference vegetation index
OBIA	Object-based image analysis
PA	Precision agriculture
SD	Standard deviation
SVM	Support vector machine
UAV	Unmanned aerial vehicle

References

1. Carrow, R.N.; Krum, J.M.; Flitcroft, I.; Cline, V. Precision turfgrass management: Challenges and field applications for mapping turfgrass soil and stress. *Precis. Agric.* **2010**, *11*, 115–134. [[CrossRef](#)]
2. Beard, J.B.; Kenna, M.P. *Water Quality and Quantity Issues for Turfgrasses in Urban Landscapes*; CAST Special Publication 27; Council for Agricultural Science and Technology: Ames, IA, USA, 2008.
3. Baghzouz, M.; Devitt, D.A.; Morris, R.L. Evaluating temporal variability in the spatial reflectance response of annual ryegrass to changes in nitrogen applications and leaching fractions. *Int. J. Remote Sens.* **2006**, *27*, 4137–4158. [[CrossRef](#)]

4. López-Bellido, R.J.; López-Bellido, L.; Fernández-García, P.; López-Bellido, J.M.; Muñoz-Romero, V.; López-Bellido, P.J.; Calvache, S. Nitrogen remote diagnosis in a creeping bentgrass golf green. *Eur. J. Agron.* **2012**, *37*, 23–30. [[CrossRef](#)]
5. Devitt, D.A.; Morris, R.L.; Kopec, D.; Henry, M. Golf course superintendents' attitudes and perceptions toward using reuse water for irrigation in the southwestern United States. *Horttechnology* **2004**, *14*, 1–7.
6. Devitt, D.A. Irrigation management considerations when using reuse water on golfcourses in the arid southwest. In Proceedings of the World Environmental and Water Resources Congress 2008, Ahupua'a, HI, USA, 12–16 May 2008; p. 316.
7. Manzano-Agugliaro, F. *Study for Economic and Environmental Sustainability in the Planning, Construction and Maintenance of Golf Courses in Mediterranean Climate*; National Report for Ministry of the Presidency of the Government of Spain 2012; Grant Project Number 190/UPB10/12; Ministry of Education, Culture and Sport: Madrid, Spain, 2012.
8. Watts, A.C.; Ambrosia, V.G.; Hinkley, E.A. Unmanned Aircraft Systems in Remote Sensing and Scientific Research: Classification and Considerations of Use. *Remote Sens.* **2012**, *4*, 1671–1692. [[CrossRef](#)]
9. Borra-Serrano, I.; Peña, J.M.; Torres-Sánchez, J.; Mesas-Carrascosa, F.J.; López-Granados, F. Spatial Quality Evaluation of Resampled Unmanned Aerial Vehicle-Imagery for Weed Mapping. *Sensors* **2015**, *15*, 19688–19708. [[CrossRef](#)] [[PubMed](#)]
10. Feng, Q.; Liu, J.; Gong, J. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China. *Water* **2015**, *7*, 1437–1455. [[CrossRef](#)]
11. Vallet, J.; Panissod, F.; Strecha, C.; Tracol, M. Photogrammetric performance of an ultra-light weight singlet UAV. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Archives, Zurich, Switzerland, 14–16 September 2011; Volume 38 (1C22), pp. 253–258.
12. Candiago, S.; Remondino, F.; De Giglio, M.; Dubbini, M.; Gattelli, M. Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sens.* **2015**, *7*, 4026–4047. [[CrossRef](#)]
13. Torres-Sánchez, J.; López-Granados, F.; Serrano, N.; Arquero, O.; Pena, J.M. High-Throughput 3-D Monitoring of Agricultural-Tree Plantations with Unmanned Aerial Vehicle (UAV) Technology. *PLoS ONE* **2015**, *10*, e0130479. [[CrossRef](#)] [[PubMed](#)]
14. Hassan, F.M.; Lim, H.S.; MatJafri, M.Z. Cropcam UAV images for land use/land cover over Penang Island, Malaysia using neural network approach. In Proceedings of the SPIE: Earth Observing Missions and Sensors: Development, Implementation, and Characterization, Incheon, Korea, 13–14 October 2010; Volume 7862, p. 78620.
15. Wang, J.; Wang, L.; Yue, X.; Liu, Y.; Quan, D.; Qu, X.; Gan, H.; Wang, J. Design and test of unmanned aerial vehicle video transfer system based on WiFi. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 47–51.
16. Yang, G.; Li, C.; Yu, H.; Xu, B.; Feng, H.; Gao, L.; Zhu, D. UAV based multi-load remote sensing technologies for wheat breeding information acquirement. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 184–190.
17. Kusnierek, K.; Korsaeht, A. Challenges in using an analog uncooled microbolometer thermal camera to measure crop temperature. *Int. J. Agric. Biol. Eng.* **2014**, *7*, 60–74.
18. Peña-Barragán, J.M.; Ngugi, M.K.; Plant, R.E.; Six, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sens. Environ.* **2011**, *115*, 1301–1316. [[CrossRef](#)]
19. Berni, J.A.J.; Zarco-Tejada, P.J.; Suarez, L.; Fereres, E. Thermal and Narrow-band Multispectral Remote Sensing for Vegetation Monitoring from an Unmanned Aerial Vehicle. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 722–738. [[CrossRef](#)]
20. Zarco-Tejada, P.J.; Berni, J.A.J.; Suárez, L.; Sepulcre-Cantó, G.; Morales, F.; Miller, J.R. Imaging Chlorophyll Fluorescence from an Airborne Narrow-Band Multispectral Camera for Vegetation Stress Detection. *Remote Sens. Environ.* **2009**, *113*, 1262–1275. [[CrossRef](#)]
21. Suárez, L.; Zarco-Tejada, P.J.; González-Dugo, V.; Berni, J.A.J.; Sagardoy, R.; Morales, F.; Fereres, E. Detecting water stress effects on fruit quality in orchards with time-series PRI airborne imagery. *Remote Sens. Environ.* **2010**, *114*, 286–298. [[CrossRef](#)]
22. Yang, B.; Chen, C. Automatic registration of UAV-borne sequent images and LiDAR data. *ISPRS J. Photogramm.* **2015**, *101*, 262–274. [[CrossRef](#)]

23. Zhang, L.Y.; Peng, Z.R.; Li, L.; Wang, H. Road boundary estimation to improve vehicle detection and tracking in UAV video. *J. Cent. South Univ. Technol.* **2014**, *12*, 4732–4741. [[CrossRef](#)]
24. Caturegli, L.; Corniglia, M.; Gaetani, M.; Grossi, N.; Magni, S.; Migliazzi, M.; Angelini, L.; Mazzoncini, M.; Silvestri, N.; Fontanelli, M.; et al. Unmanned aerial vehicle to estimate nitrogen status of turfgrasses. *PLoS ONE* **2016**, *11*, e0158268. [[CrossRef](#)] [[PubMed](#)]
25. Perea-Moreno, A.J.; Meroño-De Larriva, J.E.; Aguilera-Ureña, M.J. Comparison between pixel based and object based methods for analysing historical building façades. *Dyna* **2016**, *91*, 681–687.
26. Ma, L.; Cheng, L.; Han, W.Q.; Zhong, L.S.; Li, M.C. Cultivated land information extraction from high-resolution unmanned aerial vehicle image data. *J. Appl. Remote Sens.* **2014**, *8*, 083673. [[CrossRef](#)]
27. Díaz-Varela, R.A.; Zarco-Tejada, P.J.; Angileri, V.; Loudjani, P. Automatic identification of agricultural terraces through object-oriented analysis of very high resolution DSMs and multispectral imagery obtained from an unmanned aerial vehicle. *J. Environ. Manag.* **2014**, *134*, 117–126. [[CrossRef](#)] [[PubMed](#)]
28. Chung, M.; Detweiler, C.; Hamilton, M.; Higgins, J.; Ore, J.-P.; Thompson, S. Obtaining the Thermal Structure of Lakes from the Air. *Water* **2015**, *7*, 6467–6482. [[CrossRef](#)]
29. Feng, Q.; Liu, J.; Gong, J. UAV Remote Sensing for Urban Vegetation Mapping Using Random Forest and Texture Analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [[CrossRef](#)]
30. DJI Company. Phantom 2 Vision+ User Manual, V1.6. Available online: http://download.dji-innovations.com/downloads/phantom_2_vision_plus/en/Phantom_2_Vision_Plus_User_Manual_v1.6_en.pdf (accessed on 19 October 2015).
31. Andreopoulos, A.; Tsotsos, J.K. 50 Years of object recognition: Directions forward. *Comput. Vis. Image Underst.* **2013**, *117*, 827–891. [[CrossRef](#)]
32. Li, Y.; Wang, S.; Tian, Q.; Ding, X. Feature representation for statistical-learning-based object detection: A review. *Pattern Recogn.* **2015**, *48*, 3542–3559. [[CrossRef](#)]
33. Battiato, S.; Farinella, G.M.; Furnari, A.; Puglisi, G.; Sniijders, A.; Spiekstra, J. An Integrated System for Vehicle Tracking and Classification. *Expert Syst. Appl.* **2015**, *42*, 7263–7275. [[CrossRef](#)]
34. Siddiqi, M.H.; Ali, R.; Khan, A.M.; Kim, E.S.; Kim, G.J.; Lee, S. Facial expression recognition using active contour-based face detection, facial movement-based feature extraction, and non-linear feature selection. *Multimed. Syst.* **2015**, *21*, 541–555. [[CrossRef](#)]
35. Hare, S.; Saffari, A.; Torr, P.H. Efficient online structured output learning for keypoint-based object tracking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; IEEE: San Diego, CA, USA, 2012; pp. 1894–1901.
36. Li, Y.; Gong, J.; Wang, D.; An, L.; Li, R. Sloping farmland identification using hierarchical classification in the Xi-He region of China. *Int. J. Remote Sens.* **2013**, *34*, 545–562. [[CrossRef](#)]
37. Hawkins, J.; Blakeslee, S. *On Intelligence*, 1st ed.; Henry Holt: New York, NY, USA, 2004; p. 296.
38. Hawkins, J.; Ahmad, S.; Dubinsky, D. Hierarchical Temporal Memory including HTM Cortical Learning Algorithms. Available online: http://numenta.org/resources/HTM_CorticalLearningAlgorithms.pdf (accessed on 19 October 2015).
39. Rozado, D.; Rodriguez, F.B.; Varona, P. Extending the bioinspired hierarchical temporal memory paradigm for sign language recognition. *Neurocomputing* **2012**, *79*, 75–86. [[CrossRef](#)]
40. Boone, A.; Karnowski, T.P.; Chaum, E.; Giancardo, L.; Li, Y.; Tobin, K.W. Image Processing and Hierarchical Temporal Memories for Automated Retina Analysis. In Proceedings of the Biomedical Sciences and Engineering Conference (BSEC), Oak Ridge, TN, USA, 25–26 May 2010; IEEE: San Diego, CA, USA, 2010.
41. Perea, A.J.; Meroño, J.E.; Aguilera, M.J. Application of Numenta[®] Hierarchical Temporal Memory for land-use classification. *S. Afr. J. Sci.* **2009**, *105*, 370–375.
42. Perea, A.J.; Meroño, J.E.; Crespo, R.; Aguilera, M.J. Automatic detection of urban areas using the Hierarchical Temporal Memory of Numenta[®]. *Sci. Res. Essays* **2012**, *7*, 1662–1673.
43. Numenta Inc. *HTM Cortical Learning Algorithms*; Numenta, Inc.: Redwood City, CA, USA, 2010.
44. Hawkins, J.; George, D. *Hierarchical Temporal Memory, Concepts, Theory, and Terminology*; Numenta, Inc.: Redwood City, CA, USA, 2007.
45. Perea, A.J.; Meroño, J.E. Comparison between New Digital Image Classification Methods and Traditional Methods for Land-Cover Mapping. In *Remote Sensing of Land Cover: Principles and Applications*, 1st ed.; Giri, C., Ed.; CRC Press Taylor and Francis Group: Boca Raton, FL, USA, 2012; pp. 1662–1673.
46. George, D.; Jaros, B. *The HTM Learning Algorithms*; Numenta, Inc.: Redwood City, CA, USA, 2007.

47. Weldon, T.P.; Higgins, W.E.; Dunn, D.F. Efficient Gabor filter design for texture segmentation. *Pattern Recogn.* **1996**, *29*, 2005–2015. [[CrossRef](#)]
48. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
49. Revollo, N.V.; Delrieux, C.A.; Perillo, G.M.E. Automatic methodology for mapping of coastal zones in video sequences. *Mar. Geol.* **2016**, *381*, 87–101. [[CrossRef](#)]
50. Duro, D.C.; Franklin, S.E.; Dubé, M.G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* **2012**, *118*, 259–272. [[CrossRef](#)]
51. Karakizi, C.; Oikonomou, M.; Karantzas, K. Vineyards Detection and Vine Variety Discrimination from very high Resolution Satellite Data. *Remote Sens.* **2016**, *8*, 235. [[CrossRef](#)]



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).