Unsupervised classification of very high remotely sensed images for grapevine rows detection

Nicola Puletti*1, Rita Perria2 and Paolo Storchi2

1Consiglio per la ricerca e la sperimentazione in agricoltura, Forestry Research Centre (CRA-SEL), vialle Santa Margherita, 80, 52100 Arezzo (AR), Italy
2Consiglio per la ricerca e la sperimentazione in agricoltura, Unità di ricerca per la viticoltura (CRA-VIC), Via Romea, 53, 52100 Arezzo (AR), Italy
*Corresponding author, e-mail address: nicola.puletti@entecra.it

Abstract
In viticulture, knowledge of vineyard vigour represents a useful tool for management. Over large areas, the grapevine vigour is mapped by remote sensing usually with vegetation indices like NDVI. To achieve good correlations between NDVI and other vine parameters the rows of a vineyard must be previously identified. This paper presents an unsupervised classification method for the identification of grapevine rows. Only the red channel of an RGB aerial image is considered as input data. The image is first masked preserving only the considered vineyard and then pre-processed with a high pass filter. The pixel populations are split in “row” and “inter-row” subset through a Ward’s modified technique. The proposed methodology is compared with standard object oriented procedure tested on six vineyards located in Tuscany using as reference manually digitalized vine rows.

Keywords: precision viticulture, unsupervised classification, NDVI, Airborne Remote Sensing.

Introduction
Remotely sensed data represents a support for decision-making in agriculture, because they provide precise information over large areas. In winegrowing regions - particularly ones covered by Protected Denomination of Origin (D.O.P.) - accurate digital vineyards maps could be useful to help monitoring of vigour, erosion, flood risks, etc. [Rabatel et al., 2008]. For their capacity in providing direct information on the management of vineyard and practical application of precision viticulture, vigour maps are the ones mostly required from vine growers [Hall et al., 2003]. Since healthy and vigorous vine will show strong near-infrared reflectance and low red reflectance, vigour maps were usually produced from Normalized Difference Vegetation Index (NDVI) [Proffit et al., 2006].

The extraction of vine rows from multispectral image scenes could be important because it (i) allows to concentrate the analysis on vine canopy and (ii) ensures a more reliable
correlation of NDVI (or other indices) with specific vine parameters (e.g. biomass or production of grapes).

The steps for vine rows detection consist in the discrimination of vineyards from other land cover types [Richardson and Wiegand, 1977] and, subsequently, in the discrimination of single rows in the vineyard. These operations can be done both manually or automatically. Manual delineation of rows is more precise but it is really time-consuming and this is one of the reasons why automatic procedures are preferred.

The automatic approaches include region growing techniques [Chang et al., 1994], thresholding techniques [Tellaeche et al., 2008], edge-based detection techniques [Bobillet et al., 2003] or texture analysis techniques. To segment vineyards blocks, one of the most robust texture analysis method uses the Fast Fourier Transform (FFT) or the Gabor filters [Ranchin et al., 2001; Wassenaar et al., 2002; Rabatel et al., 2008], although they are sensitive to row spacing and works only on linear rows.

Image segmentation methods could be considered a good alternative to the conventional per-pixel methods, since they consider the spatial characteristics, in the identification of an “object”; this allowing object-oriented analysis [Flanders et al., 2003; Benz et al., 2004; Lamonaca et al., 2008].

For some authors [e.g. Smit et al., 2010], thresholding techniques are the most efficient. An application of a thresholding segmentation technique was used to distinguish between weeds and crops in a field [Tellaeche et al., 2008]. In this case, the threshold has been subjectively chosen analysing the histogram of the grey-scale image.

In this work, a procedure based on Ward’s technique (referable to the thresholding techniques) is proposed and tested, to automatically and objectively obtain a threshold. The application of this method is compared with an object-oriented procedure based on eCognition software [Definiens, 2003]. Results on six experimental plots are evaluated using as reference manually digitized vine rows.

Test area and data
The study area is located in Tuscany within the Barone Ricasoli estate (Gaiole in Chianti, Siena, Italy), in the heart of Chianti Classico area (Fig. 1). It displays the pedological and climatological characteristics best suited for the production of quality wines [Bucelli et al., 2010]. The average dimension of the vineyards is about 4 hectares, a surface representative of many other Tuscan estates located in hilly, often steeply sloping areas. The vineyards are planted on hillsides, at altitudes ranging from 180 to 490 meters, on slopes facing south and southwest. The rows are traditionally oriented with the “rittochino” system, perpendicular to contour lines. In our test areas, the inter-row spaces are not grass-covered.

Our dataset consists of a real colour image, used for segmentation, and a reference dataset. The real colour image, acquired in 2010 (7th of July) from aircraft through a Canon camera with ground cover of 1051 x 700.8 m, had a spatial resolution of 0.3 m and a windows extent that ranges from 697200 to 697830 East and from 4806090 to 4807250 North in WGS 84, Zone 32-Nord.

The reference dataset is the rasterization of the vector layers produced by manual delimitation of the vine rows border. Each row or row segment is represented as a polygon by the GIS operator [Quantum GIS, 2013].
Methods
The methodology here proposed uses a plot-based analysis approach. In our case, an “experimental plot” corresponds to a vineyard. The main task is to find - at plot level - the threshold that splits the pixel population into two groups: rows and inter-rows. The entire procedure can be divided in two steps: “pre-processing” and “unsupervised classification” (Fig. 2).
Pre-processing
The first task of pre-processing is the delineation of experimental plots and this can be done both manually or automatically [Da Costa et al., 2007; Delenne et al., 2010].

As in the experience of other authors [Delenne et al., 2010], preliminary tests done on the blue, green, red, near infrared channels have shown that best results are obtained with the red channel. This is mainly due to the higher contrast between vine rows (vegetation) and interrows, generally occurring in the red channel, especially when the interrows are covered by grass. This difference produces high contrast between vine rows and soil, local minima of mean DN make rows identification easier (Fig. 3).

Then, a 3x3 high pass filter (HPF) – i.e. a filter that passes high frequencies [Schowengerdt, 1980] - is used. The HPF emphasizes the boundaries between rows and inter-rows and reduces the mixed reflectance effects. Typically [Schowengerdt, 1980], HPF uses a kernel with a high central value surrounded by negative weights [eq. 1]:

![Figure 3 - (above) Window of 40x25 pixels cropped from the red channel; (below) vertical projection profile of DN values at specified (red dots) sample points. Lower values correspond to rows, higher ones to inter-rows (soil).](image)
\[ P_{HPF} = \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \] [1]

Note that the values in the kernel sum to zero, and so they are normalized.

**Object-oriented approach with eCognition**

In eCognition’s [Definiens, 2003] the segmentation procedure divides the original image into small objects (i.e. polygons). The user can specify several options (scale, colour, shape ratio, compactness, smoothness and individual weightings of input layers). Different configurations were tested, the final software parameters here adopted are: scale = 5; colour = 0.7; shape ratio = 0.3; compactness = 0.5; smoothness = 0.5: for the classification a fuzzy-sinusoidal membership was used. The weight of input layer is irrelevant because only the red channel was considered. In each experimental plot a subset of objects derived from segmentation where manually classified into two classes (rows and inter-rows) and were used as training for the automatic image classification.

**Ward’s technique**

Ward’s method minimizes the variance within groups, so it can only be used for quantitative variables. Starting with elementary sets, at each step, this algorithm optimises the partition obtained aggregating two elements in a class. A partition improves if classes are more homogeneous inside and more different among each other. In other words, at high values of the variance between classes correspond low values of the variance inside the classes.

In our work we use the version of Ward’s technique modified by Longobardi and Villani [2009], where a maximum number of two clusters is allowed and the maximization of the Euclidean distance to the cluster mean of each case (second term of [eq. 2]), is considered as the target. Maximization of the deviance between clustering groups corresponds to the maximization of differences between groups and thus to the minimization of within cluster variability [Longobardi and Villani, 2009]. The deviance \( (dev(X)) \) of the process can be written as:

\[
dev(X) = \sum_{i=1}^{n_1+n_2} (X_i - \mu_X)^2 = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (X_{i,j} - \mu_i)^2 + \sum_{j=1}^{2} (\mu_{X_j} - \mu_X)^2 \] [2]

where \( X \) is the variable (here the values of digital number of the red channel), \( n_1 \) and \( n_2 \) are the sizes of sub-sets (rows and inter-rows, respectively).

**Procedure assessment**

The performances of classifications derived from Object Oriented Procedure (OOP) and Ward’s Modified Method (WMM) were synthetically evaluated through the computation of four indices based on the use of an error matrix (or contingency table). In this work, where only two categories (row and inter-row) are considered, the error matrix has two columns and two rows: columns refer to reference dataset while rows contain the classified dataset (Tab. 1).
Table 1 - Error matrix schema when only two categories are considered. \( n_{TP} \) is the number of pixels correctly classified as rows; \( n_{FP} \) is the number of pixels incorrectly classified as rows; \( n_{FN} \) is the number of pixels incorrectly classified as inter-rows; \( n_{TN} \) is the number of pixels correctly classified as inter-rows; \( T_{R1} \) and \( T_{R2} \) are the per row sums; \( T_{C1} \) and \( T_{C2} \) are the per column sums; \( N \) is the total amount of pixels within experimental plot.

<table>
<thead>
<tr>
<th>Classified Data</th>
<th>Reference Data</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>( n_{TP} )</td>
<td>( n_{FN} )</td>
</tr>
<tr>
<td>Inter-rows</td>
<td>( n_{FP} )</td>
<td>( n_{TN} )</td>
</tr>
<tr>
<td>Total</td>
<td>( T_{C1} )</td>
<td>( T_{C2} )</td>
</tr>
</tbody>
</table>

Following Congalton [1991], the accuracy measures adopted are (1) Overall Accuracy \( \hat{p} \), (2) KHAT statistic (KHAT), (3) User’s Accuracy \( A_u \) and (4) Producer’s Accuracy \( A_p \) (Tab. 2). Overall accuracy is computed by dividing the total correct (i.e., the sum of the major diagonal) by the total number of pixels in the error matrix. KHAT statistic is a measure of the difference between the actual agreement between reference data and the results of classification, and the chance agreement between the reference data and a random classifier. Producer’s accuracy (so called because the producer of the classification is interested in how well a certain area can be classified) indicates the probability of a reference pixel being correctly classified and is really a measure of omission error. User’s accuracy estimates the probability that a pixel classified as belonging to a specific category actually represents that category on the reference.

Table 2 - Calculation algorithms of indices used to assess quality of image classification. See table 1 for nomenclature.

<table>
<thead>
<tr>
<th>Index</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall Accuracy</td>
<td>( \hat{p} = \frac{(n_{TP} + n_{FN})}{N} )</td>
</tr>
<tr>
<td>KHAT statistic</td>
<td>( KHAT = \frac{\hat{p} - \hat{\theta}}{1 - \hat{\theta}} ) where ( \hat{\theta} = \frac{1}{n^2} \sum_{j=1}^{2} T_{Rj} \cdot T_{Cj} )</td>
</tr>
<tr>
<td>Producer’s Accuracy for rows detection</td>
<td>( A_p = \frac{n_{TP}}{T_{C1}} )</td>
</tr>
<tr>
<td>User’s Accuracy for rows detection</td>
<td>( A_u = \frac{n_{TP}}{T_{R1}} )</td>
</tr>
<tr>
<td>Producer’s Accuracy for inter-rows detection</td>
<td>( A_p = \frac{n_{TN}}{T_{C2}} )</td>
</tr>
<tr>
<td>User’s Accuracy for inter-rows detection</td>
<td>( A_u = \frac{n_{TN}}{T_{R2}} )</td>
</tr>
</tbody>
</table>
Results

Table 3 shows the evaluation statistics for each experimental plot. The values of Overall Accuracy ($\hat{p}$) are always high (mean $\hat{p}_{\text{OOP}} = 0.84$; mean $\hat{p}_{\text{WMM}} = 0.82$). The minimum $\hat{p}$ value is recorded in the experimental plot 1 ($\hat{p}_{\text{OOP}} = 0.78$; $\hat{p}_{\text{WMM}} = 0.77$) while the best performances are in experimental plot 3 ($\hat{p}_{\text{OOP}} = 0.87$; $\hat{p}_{\text{WMM}} = 0.87$).

Table 3 - Values of Overall Accuracy ($\hat{p}$), KHAT, Producer’s Accuracy ($A_p$) and User’s Accuracy ($A_u$) calculated for Object-oriented procedure (OOP) and Ward’s Modified Method (WMM).

<table>
<thead>
<tr>
<th>Experimental Plot (ID)</th>
<th>Classification method</th>
<th>($\hat{p}$)</th>
<th>KHAT</th>
<th>Rows</th>
<th>Inter-rows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$A_p$</td>
<td>$A_u$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$A_p$</td>
<td>$A_u$</td>
</tr>
<tr>
<td>1</td>
<td>OOP</td>
<td>0.78</td>
<td>0.54</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.77</td>
<td>0.54</td>
<td>0.75</td>
<td>0.84</td>
</tr>
<tr>
<td>2</td>
<td>OOP</td>
<td>0.84</td>
<td>0.66</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.83</td>
<td>0.63</td>
<td>0.63</td>
<td>0.92</td>
</tr>
<tr>
<td>3</td>
<td>OOP</td>
<td>0.87</td>
<td>0.72</td>
<td>0.93</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.87</td>
<td>0.74</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td>4</td>
<td>OOP</td>
<td>0.83</td>
<td>0.65</td>
<td>0.84</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.79</td>
<td>0.58</td>
<td>0.79</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>OOP</td>
<td>0.86</td>
<td>0.69</td>
<td>0.88</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.79</td>
<td>0.59</td>
<td>0.73</td>
<td>0.94</td>
</tr>
<tr>
<td>6</td>
<td>OOP</td>
<td>0.84</td>
<td>0.69</td>
<td>0.75</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.83</td>
<td>0.65</td>
<td>0.69</td>
<td>0.94</td>
</tr>
<tr>
<td>Mean</td>
<td>OOP</td>
<td>0.66</td>
<td>0.84</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>WMM</td>
<td>0.62</td>
<td>0.82</td>
<td>0.74</td>
<td>0.90</td>
</tr>
</tbody>
</table>

All other indices confirm this trend. As an example, within experimental plot 1 KHAT assumes values slightly higher than 0.5 for both procedures, while in experimental plot 3 it has values greater than 0.70 ($KHA\text{T}_{\text{OOP}} = 0.72$; $KHA\text{T}_{\text{WMM}} = 0.74$). This can be connected to different spectral characterization of row/inter-row patterns that in experimental plot 3 is more homogeneous with respect to experimental plot 1 (Fig. 4) and, consequently, more easily distinguishable also by the algorithm.
Mean values of $A_p$ and $A_u$ are higher than 0.75, with maximum value higher than 0.90, demonstrating a good ability of WMM in detection of both row and inter-row.

**Conclusion**
A method to extract row pixels within delimited vineyards is presented, tested over six experimental plots belonging to a private property in the Chianti area and, lastly, assessed and compared with an object-oriented classification. The procedure appears easy to use, needs just the red channel of very high-resolution images and requires only two steps, after a preliminary effort that produces the polygons delimiting the areas to be analysed. Choice of the red channel is due to the higher contrast between grapevine rows and interrows and to the simple identification of grapevine DN that can be obtained.

Before classification with Ward’s modified technique, red channel image is pre-processed using a high pass filter, which removes the slowly varying components and enhances the high-frequency local variations and consequently emphasise the transition between areas with low DN values (rows) and those with high ones (soil, inter-rows).

The results obtained with this method are compared both with manually drawn polygons, considered as reference, and with a standard image object-oriented procedure developed within eCognition’s environment. As expected, the method works better when the contrast between plant and soil DNs is high.
Areas with low or medium vigour, in which this distinction is less pronounced, make the job more complex, tough still statistically and operationally useful.

Acknowledgements
We acknowledge two anonymous reviewers, whose comments helped us to improve the quality of the original manuscript. This work was developed under the IMViTo project founded by Regione Toscana.

References