A Bayesian framework for joint structure and colour based pixel-wise classification of grapevine proximal images

Florent Abdelghafour, Roxana Rosu, Barna Keresztes, Christian Germain, Jean-Pierre da Costa

A Bayesian framework for joint structure and colour based pixel-wise classification of grapevine proximal images.

Abdelghafour F.[1,2], Rosu R.[1,2], Keresztes B.[1,2], Germain C.[1,2] and Da Costa J.P.[1,2]

[1]Univ. Bordeaux, IMS UMR 5218, F-33405 Talence, France

[2]CNRS, IMS UMR 5218, F-33405 Talence, France.

florent.abdelghafour@ims-bordeaux.fr, roxana.rosu@ims-bordeaux.fr,
barna.keresztes@ims-bordeaux.fr, christian.germain@ims-bordeaux.fr,
jean-pierre.dacosta@ims-bordeaux.fr

Keywords: Proximal sensing ; grapevine ; texture ; parametric classification ; Local Structure Tensor.

Abstract

Estimating the spatial variability of basic agronomic parameters at the scale of the plant is of prime importance for the development and monitoring of Precision Agriculture applications. It is all the more crucial in viticulture where intra-plot variabilities are exacerbated. This paper focuses on the description of the structure of the canopy at the plant scale by proximal imaging. A new framework is proposed for the pixel-wise classification of the grapevine canopy into organs at different phenological stages. The proposed processing chain proceeds in four steps: (i) foreground extraction, (ii) pixel-wise feature extraction, (iii) pixel-wise classification and (iv) spatial regularization. Step (i) is based on colour information only. For step (ii), colour is represented using an RGB triplet while texture is captured using the local structure tensor (LST). Two variants are proposed to associate colour and LST information into a single Euclidean vector. Step (iii) is a Bayesian decision process based on the joint modelling of colour and texture using multivariate Gaussian distributions. Finally, step (iv) combines stochastic relaxation and morphological filtering, allowing for the spatial regularisation of the classification output. This processing chain is applied to the pixel-wise classification of proximal images into grapevine organs. Images were taken from two $0.2ha$ plots planted with the red variety "*Merlot Noir*" in Bordeaux area. Images were taken from an embedded acquisition system at three key phenological stages: *flowerhood falling*,

*pea-sized berries*and *berries touching* (BBCH 68, 75 and 79). Results are produced with *leave-one-out cross validations* where models are estimated from only 15 images per stage containing about 3.2×10^6 labellised pixels. The resulting classification performances are measured in terms of recall and precision that reached overall between 85% and 95% depending on the stage while overall accuracies range between 88% and 93%.

1 Introduction

The development of Precision Viticulture (PV) applications has considerably improved the efficiency of vineyard management strategies in terms of productivity, quality and environmental impact (Bramley, 2010). PV is the ability to control and adapt the spatial distribution of inputs or mechanical operations within a parcel, according to site-specific characteristics of the vines. In opposition to the conventional uniform practices, PV is a promising solution for more viable and sustainable grape-growing (Tona *et al.*, 2017). However, the efficiency of PV applications relies mostly on the abundance, reliability and resolution of *in-vivo*-measured agronomic parameters (Taylor *et al.*, 2007). In order to acquire and analyse agronomic data at the scale of the plant, proximal sensing, notably using optical sensors, is a very promising automated and non-intrusive technique. Indeed, with relatively low costs in terms of instrumentation, labour and time duty, it enables the local assessment of various agronomic parameters and of their intra-parcelvariability across large acreage.

The PV literature includes several examples of research intended to describe grapevine organs (leaves, berries, stems) in the trellising plane using optical sensors, with various applicative interests. For instance, detecting, counting and measuring grape bunches or berries allow for early estimations of yield. At the scale of PV, such information about plant productivity enables to monitor locally fertilisation or irrigation. For this purpose, Nuske *et al.* (2012) proposed to detect grape berries on proximal colour images thanks to geometric criteria. Later, Liu *et al* (2015) proposed the use of 3-D stereo vision system to estimate the volume of grape bunches. More recently, Abdelghafour *et al.* (2017) presented a colour and texture based machine learning application for the detection and counting of inflorescences and grape clusters at early fruiting stages on proximal images. Similarly, Keresztes *et al.* (2018) proposed a combination of geometrical criteria and Deep Learning to

detect grape berries and then to reconstruct and count grape bunches on proximal images.

Besides the fruit bearing part of the trellising plane, PV research also addresses the description of the foliage. Indeed, local estimations of its volume enable to monitor vigour-control and aeration operations such as defoliation, trimming, thinning or precision spraying (Tisseyre *et al.*, 2007). Pfeiffer *et al.* (2018), proposed to estimate the canopy crown surface and its porosity thanks to the analysis of topographic images obtained with a Lidar sensor. Vieri et al. (2013) presented an autonomous spraying robot able to adapt the flow-rate of pesticides according to foliage density which was estimated with ultrasonic sensors.

Concerning the remaining organs of the grapevine, stems and shoots have also been the subject of proximal sensing studies. Counting shoots and measuring stems provide an insight on the future yields and on the energetic reserves contained in the vine wood. Liu *et al.* (2017) presented a computer vision system for early stage grape yield estimation based on shoot detection. At a different stage Demestihas *et al.* (2018), presented a laser based sensor used to count and measure the diameters of stems. This measure is strongly correlated to the pruning weight, an indicator of vigour during the cold hardening (*i.e.* lignification) phase taking place in August (in the northern hemisphere). For the same purpose, Keresztes *et al.* (2018) proposed a proximal imaging application for the estimation of vine shoots volume and morphology.

However, to this date, no proposal of methods able to describe and locate altogether the different objects present in the trellising plane has been made. Yet, describing the plant architecture *i.e.* the spatial distribution and proportions of leaves, stems and fruits in the trellising plane is equally important (Mathews and Jensen, 2013). Indeed, many decisions regarding cultivation operations are not only based on the physiological expression of a single type of organ. It is often rather based on equilibriums and relative expressions within the global architecture of the plant. For instance, the requirements in defoliation or trimming do not only depend on the volume of foliage. Essentially, these operations are adjusted according to the balance between canopy vigour and fruit load, the exposition of fruits to natural light, the amount of secondary shoots or the balance between young-upper leaves and aged leaves. All the information necessary to describe the grapevine trellising plane architecture are nonetheless present on proximal images of the plant. It is a reasonable assumption that statistical indices extracted on these images

could be transcripted into useful agronomic parameters in the PV context. Even though this particular agricultural problem has not been addressed yet, there are nevertheless well-established methods in the field of image processing that could prove to be efficient solutions.

Describing the trellising plane architecture can be seen as a conventional object-detection / classification problem in image processing. This problem faces the same issues encountered in plant phenotyping where the purpose is to describe some parts of the plant in order to determine agronomic properties. The strategy commonly involved relies on three major steps : (i) choosing appropriate features based whether on textural, colour or geometric properties, (ii) modelling these features for the desired classes, (iii) applying a classifier (usually supervised) based on field data used to learn the parameters and validate the decision model. Several plant phenotyping applications have been developped according to this strategy. For instance Yalcin (2015) proposed to estimate growth stages thanks to proximal textural analysis. More recently Zhang *et al.* (2018) proposed to characterise the intensity of the flowering stage for canola flowers.

Following this general strategy, this paper proposes a solution for the mapping of the trellising plane into classes of organs. It is an essential step for the characterisation of architectural and agronomic properties of the grapevine vegetation. The proposed framework relies on a pixel-wise classification based on the parametric modelling of both textural and colorimetric local properties. Texture is described using Local Structure Tensors (LST) as in Rosu *et al.*, (2016, 2017) while colour is considered as Gaussian filtered triplets in the RGB space. Two novel representations joining texture and colour information into log-Euclidean vectors are introduced. These representations are adapted for the use of probabilistic modelling tools, multivariate Gaussian models or Gaussian mixtures in this case. For both representations, a model is learnt on a collection of pixel samples in images containing the different organs of the grapevine. The eventual pixel-wise classification is performed by Bayesian MAP estimation (Maximum a posteriori probability) based on the previously learnt model's parameters. This decision process results in a pixel-wise classification map that is further regularised both by probabilistic relaxation and by morphological filtering.

The designed processing chain is an original combination of methods that have not

been implemented in the context of agricultural applications. It constitutes the first methodological contribution of this work. The second main contribution concerns the feature extraction step (section 3.3). In this step, two solutions are presented to capture colour and structure information into joint vector representations that are mathematically tractable for the subsequent stochastic modelling and Bayesian decision processes. The first one, called LEEST, already presented in (Abdelghafour *et al.*, 2018) consists in mapping the extended structure tensor, a feature proposed in (De Luis-Garcia *et al.* (2008) into the Log-Euclidean space. The second solution, called CELEST, has never been presented before. This is an alternative to join colour and structure into a more compact representation, easier to manage in the Bayesian machine learning process. Finally, authors also propose a spatial criterion in the use of a priori probabilities in the decision process detailed in section 3.4.2. This proposition accounts for the very unbalanced and varying spatial distribution of organs on the trellising plane.

2 Plant material and instrumentation

2.1 Vine plots

The plant material used for experimental study is composed of two 0.2 ha plots with 120 cm row-spacing and planted with the red wine grape variety Merlot Noir in "Le Domaine de la Grande Ferrade", a public experimental facility in the area of Bordeaux (INRA, French National Institute of Agricultural Research). Between May and September 2017, the two plots were extensively photographed weekly. The resulting image database contains more than 30,000 images covering phenological stages ranging from "*inflorescence swelling*" to "*half-ripening*" according to the conventional BBCH (Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie) phenological scale established by Lorenz et al. (1995) *i.e.* BBCH 53 to 83.

2.2 Acquisition system

An image database had to be constructed for the learning and validation steps of the proposed approach. Images had to present homogeneous properties in terms of

resolution, distance and angle of capture and also in terms of illumination. The achievement of such prerequisites is not so trivial in uncontrolled and highly variable outdoor environments. Therefore, instrumentation and acquisition methods have been designed to minimize the impact of natural light on the levels and variations of illumination in images in order to preserve their intrinsic textural properties.

The device used for image acquisition is composed of a 5 Mpx industrial Basler Ace (acA2500-14gc GigE) RGB camera with a 55° horizontal field of view lens, a high-power 58GN xenon flash (Neewer speedlite 750ii) with short exposure time ($250 - 300\ \mu sec$), a $12V$ battery and an industrial computer (built around a low consumption 4-core ARM chip), all placed in a compact and watertight case (Fig.1b). The device is completed with a GNSS receiver (G-star IV) for georeferencing and an ultrasonic sensor which provides the distance between the camera and the trellising plane, allowing to compute pixel size. The device was embedded on a vineyard tractor at 70 cm above ground and at 50 cm from the target (Fig. 1a). Each image covers an area including a full vine stock and its canopy with a resolution of 2592×2048 pixels and around 3 $px.mm^{-1}$. Acquisitions are adapted for the work-rate in vineyards ($3-8\ km.h^{-1}$). Images were taken between 7 am and 5 pm. Despite the various insolations and cloud coverages encountered during the season, the intensity of the light emitted by the 58GN xenon flash during the short exposure time provides images with consistent illumination of the foliage regardless of the natural lighting conditions. Images taken with different conditions are shown in figure 1c and 1d which depict two examples of images acquired by the device (Fig. 1b) for a same phenological stage (BBCH 68 "*flowerhoods falling*"). These two images were captured with different natural lights.The sky and background is a lot darker on image figure 1(d) However, foliages present similar illumination on both images. In these conditions natural light and insolation have little impact.

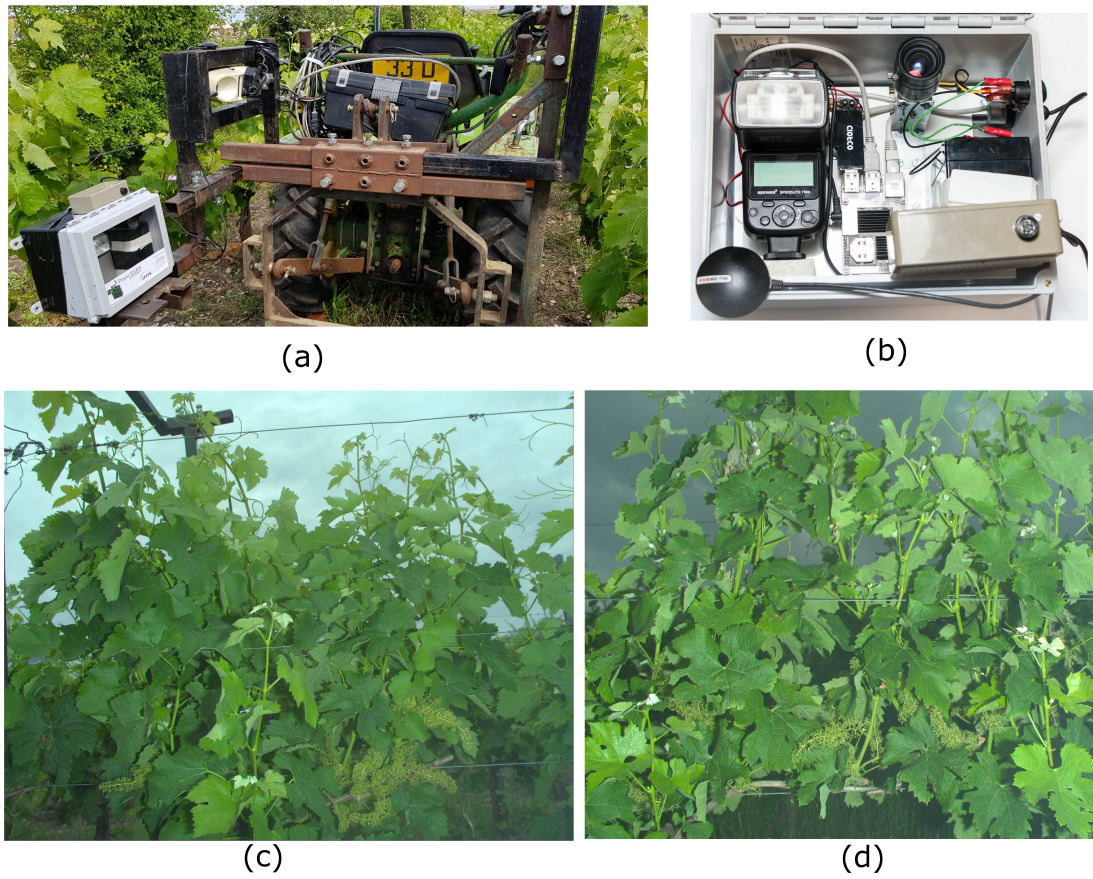


FIGURE 1 – Instrumentation : embedding on a vineyard tractor (a), device compounds (b), example of a resulting image in moderate sunlight (c) and example of image in a low-sunlight (d)

2.3 Groundtruthing

In order to perform the learning and validation phases of the classification process, a labelling procedure was conducted on 16 images (*i.e.* 16 vinestocks) for each of the three phenological stages. Images were selected randomly among the plots under study and present the various morphologies encountered for the variety and the cultivation system. On each image, around 2.0×10^5 pixels w labelled in one of the four classes: leaf cores, leaf edges, berries/inflorescences or stems. Classes are not evenly represented, classes which are naturally more represented on the images are also more represented in the database of labels. In total the database contains 3.2×10^6 pixels per phenological stage. In practice labelling consists in delineating homogeneous area then sampling pixels inside it.

3 Image Processing chain

3.1 Framework overview

The purpose of the proposed framework is to provide a pixel-wise classification of grapevine colour images into the different classes of organs which are visible in the trellising plane. The process consists in classifying pixels into one of the following classes: foliage, stems or reproductive organs (*i.e.* berries, flowers or buds depending on phenological stages).

The classification process is based essentially on the estimation of the likelihood of the local properties of the pixel and its close neighbourhood with parametric models describing classes. The maximum likelihood obtained for a class determines the eventual affiliation of the pixel to this class. This process is based on a parametric modelling of local pixel properties such as anisotropy and colour. These properties are captured by an extended form of the Local Structure Tensor (Bigun *et al.*, 1991). The following sub-parts aim at describing the different steps of the proposed processing chain (Fig. 2).

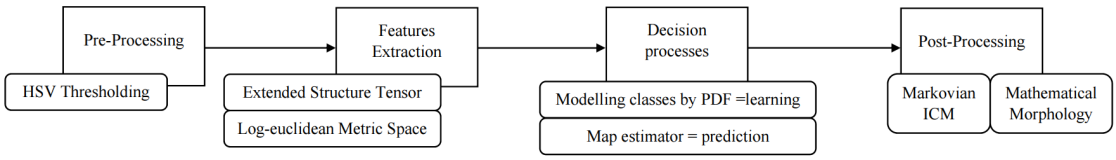


FIGURE 2 – Processing Chain.

3.2 Foreground extraction

The pre-processing aims at narrowing down areas where the main process operates by removing irrelevant parts of images, *i.e.* objects not belonging to the canopy (sky, ground, trellising wires, poles or grass). Authors propose a simple thresholding in the HSV colour space followed by simple morphological operations. Hue and Saturation channels enable to easily discard objects with colours which are implausible for foliage. While yellow, green and brown colours are retained, grey, blue and red colours are discarded. The Value channel is used to discard shadows and background but also overexposed or underexposed parts of the foliage which present extreme or marginal values.

3.3 Features extraction

Organs are not only characterised by their variations of colours but mainly by geometric properties like the anisotropy of their contours or textural properties. These particular properties can be extracted thanks to the local structure tensor (LST) that can be extended so that it also includes colour information.

3.3.1 Local Structure Tensor

The LST is a reference tool developed by Knutsson (1989) that extracts geometric information and orientation trends in local patterns within grayscale images. It is commonly defined as the local covariance of gradients (Bigün *et al.*, 1991 ; Rosu *et al.*, 2016). The computation of a LST field is a two step process, starting with estimating local gradients in the neighbourhood of every pixel in an image. Given an image I of size $[M \times N]$, the gradient image $\vec{\nabla} I$ is estimated as :

$$\vec{\nabla} I = [I_x, I_y]^t = [I * G_x, I * G_y]^t, \tag{1}$$

where t denotes the matrix transpose operator, $*$ denotes convolution, I_x and I_y represent respectively estimates of the horizontal and vertical derivatives of image I obtained by applying Gaussian derivative kernels G_x and G_y.

The LST field is then computed by smoothing the product $\vec{\nabla} I \vec{\nabla} I^t$ with a Gaussian filter W_T with a standard deviation σ_T:

$$Y = W_T * \vec{\nabla} I \vec{\nabla} I^t = W_T * \begin{bmatrix} I_x.I_x & I_x.I_y \\ I_x.I_y & I_y.I_y \end{bmatrix}. \tag{2}$$

Thus, for every pixel $(i, j) \in [1, N] \times [1, M]$ there is a corresponding local structure tensor, in the form of a 2×2 symmetric matrix : $Y(i,j) = \left[\begin{smallmatrix} y_{xx}(i,j) & y_{xy}(i,j) \\ y_{xy}(i,j) & y_{yy}(i,j) \end{smallmatrix}\right]$.

3.3.2 Log-Euclidean mapping of structure tensors

Structure tensors being covariance matrices, they belong to the Riemannian manifold of Symmetric Positive-Definite (SPD) matrices. The use of standard tools of

Euclidean geometry and Gaussian statistics on such variables is not straightforward (Arsigny *et al.*, 2007). For instance, computing a centre of mass or fitting a probabilistic distribution such as a multivariate Gaussian are not trivial tasks and should be carried out by considering the properties of the Riemannian manifold. Saïd et al. (2017) proposed several methods and parametric models adapted to the geometry of LST and notably Riemannian Gaussian distributions for strictly positive definite matrices. A more simple and convenient way to handle LST's is to map them into the Log-Euclidean space before applying geometric tools (Arsigny *et al.*, 2006) or standard probabilistic models (Rosu *et al.*, 2017). Rosu *et al.* (2017) successfully applied these methods to the classification of patches of remote sensing images of forests and oyster fields. Experimental comparisons between different models proved that the LE metric lead to equivalent or better results with a significant decrease in computation time. These contributions motivate the choice of authors to focus on LE approaches for tensor field modelling.

The mapping of a tensor Y onto the LE space is achieved by computing its matrix logarithm:

$$Y_{LE} = log_m(Y) = \begin{bmatrix} y_{LExx} & y_{LExy} \\ y_{LEyx} & y_{LEyy} \end{bmatrix}. \tag{3}$$

Let consider the factorization $Y = RDR^{-1}$ where $D = \left[\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right]$ is the diagonal matrix of Y, λ_1 and λ_2 are the eigenvalues of Y and R is the rotation matrix composed of unitary eigenvectors of Y. Then, $log_m(Y) = R\left[\begin{smallmatrix} log(\lambda_1) & 0 \\ 0 & log(\lambda_2) \end{smallmatrix}\right]R^{-1}$, is the matrix logarithm of Y.

As mentioned in (Arsigny, 2006), a more convenient way to handle the matrix $log_m(Y)$ is to express it in the vector form $\vec{Y}_{LE}$.

$$\vec{Y}_{LE} = (y_{LExx}, y_{LEyy}, \sqrt{2}y_{LExy}), \tag{4}$$

The mapping to the LE space allows the use of classical Euclidean geometry and probabilistic tools for tensor modelling while preserving the main properties of the tensor

space.

3.3.3 Extending structure tensors with colour information

De Luis-Garcia *et al.* (2008) proposed a method to join structural and colorimetric information into a single descriptor. The method consists in computing the structure tensor from an extended gradient where RGB intensity values are concatenated to the two directional derivatives to obtain a colour Extended Structure tensor $Y_{ce} = W_T * [\vec{\nabla} I_{ce} \vec{\nabla} I_{ce}^t]$, where $\vec{\nabla} I_{ce} = \left[I_x, I_y, R, G, B \right]$

The resulting structure tensor Y_{ce} is then a 5×5 SPD matrix representing the covariance of colour extended gradients. Alike the common LST, it is proposed here to map the colour extended structure tensor into the LE metric-space thanks to the matrix logarithm transform. The resulting vectorised form $\vec{Y}_{LEEST}$ is then a 15 dimensions descriptor. This representation will be referred further as the $LEEST$ representation for Log-Euclidean Extended Structure Tensor .

An alternative representation is proposed here which consists in extending the structure tensor to colour after the mapping into the LE space. Since [R,G,B] values are already euclidean variables it is also possible to concatenate directly colorimetric features within the vectorized form of LST in the LE space $\vec{Y}_{LE}$ to obtain a Colour Extended Log-Euclidean Structure Tensor ($CELEST$): $\vec{Y}_{CELEST}$.

$$\vec{Y}_{CELEST} = [y_{LExx}, y_{LEyy}, \sqrt{2} y_{LExy}, R, G, B]. \quad (5)$$

This is a smaller dimensionality representation where the correlations between oriented gradients and colour are not primarily established in the descriptors like in De Luis-Garcia *et al.* (2008) but established further in the modelling process.

3.3.4 Rotation Invariance

When comparing the textures present in images of grapevine, it is not their global orientation nor position in images that describe their intrinsic properties, but rather the combination of orientations in various proportions that can be represented by degrees of anisotropy. Some structural patterns, stems for instance, can be anisotropic, *i.e.* they present a predominant orientation. However, a stem is still a stem whether it is horizontal

or vertical, and a leaf is a characteristic pattern regardless the orientation of its veins. It is rather the level of anisotropy than the global orientation that constitutes a coherent dissimilarity criterion. It is then meaningful to ensure rotation invariance for the designed tensor-based representations. For $LEEST$ representation, the extended gradient $\vec{\nabla} I_{ce}$ does not contain only structural information but also colorimetric information, applying rotations to this peculiar form of gradient has no interpretable geometrical meaning. Concerning $CELEST$ representation, since the diagonal matrix of a given tensor provides a unique set of eigenvalues for different possible rotation matrices, it is possible to ensure rotation invariance by retaining only the eigenvalues. When applied to $CELEST$, the rotation invariant descriptor is expressed into its vectorised form as :

$$\vec{Y}_{RI-CELEST} = [log(\lambda_1), log(\lambda_2), R, G, B], \tag{6}$$

where λ_1 and λ_2 are the eigenvalues of D computed from the diagonal decomposition of the matrix RDR^{-1}. In practice $CELEST$ representation will always be implemented with its rotation invariant form $\vec{Y}_{RI-CELEST}$.

3.4 Decision process : pixel-wise classification

3.4.1 Maximum a posteriori probability estimation (MAP)

The purpose of this step is to determine from the observation of a structure tensor Y that describes a pixel, to which class c this pixel belongs to. This decision is based on a MAP estimator, a Bayesian method based on the determination of $\underset{c \in L}{argmax} \quad p(c|Y)$. According to Bayes theorem, $p(c|Y) = \dfrac{f(Y|c)p(c)}{f(Y)}$ then:

$$\underset{c \in L}{argmax} \quad p(c|Y) = argmax \quad f(Y|c)p(c), \tag{7}$$

where, $f(Y|c)$ Probability Density Function (PDF) f describing the distribution of a subset of structure tensors Y in class c. $p(c)$ is an *a priori probability* represents the relative proportion of class c. Both $f(Y|c)$ and $p(c)$ can be learnt from representative samples of structure tensors in each class.

3.4.2 A priori probabilities

Three different assumptions can determine the possible values of $p(c)$:

— Even distribution of classes : $p(c) = \frac{1}{C}$, where C is the number of classes .

— Uneven distribution classes : $p(c) = \pi_c$, the statistical frequency of the class c

— Heterogeneous distribution classes : $p(c) = \pi_c(h)$, a function of space *i.e.* the a priori probability depends on the location of the pixel in the image.

The most realistic assumption is the latter. Indeed, images of grapevine plants are spatially structured, within such images, the different types of objects and organs are not homogeneously distributed. Indeed it is more likely to observe grapes and inflorescences in the lower part of the canopy with fewer leaves, when its core is more abundant with dense foliage partially occulting stems and its upper part contains only thin foliage showing stem's apexes and no fruits. It is then conceivable to consider a priori probabilities as functions of the height at which pixels are located.

The decision criterion can then be based on a likelihood $f(Y|c)$ modulated as a product of $\pi_c(h)$ representing the relative proportion between organs labelled c at a level h of the canopy.
Authors propose to vertically divide images into 3 parts of equal heights. Where for each part a different value of c per class is estimated thanks to the average proportions observed on labelled images.

3.4.3 Probability Density Functions, parametric models

The distributions of structure tensors are represented by multivariate Gaussian distributions and multivariate Gaussian mixtures. A multivariate Gaussian distribution can be expressed with only 2 parameters, a covariance matrix Σ and a centre of mass vector $\vec{\mu}$. Every class c of organs can be then described by a multivariate Gaussian model (Σ_c, μ_c), established from a group of labelled images. Such models are essentially probability density functions. The estimated class $\hat{c}$ is determined by the maximum value

obtained with the MAP estimator among all c-classes.

Given an image containing the same classes of organs at a similar phenological stage and a given pixel within this image, it is possible to determine to which class this pixel most probably belongs to by computing the maximum likelihood that is obtained for the different possible models.

For a given class $c \in C$, which is described by a dataset of N-dimensional random variables, the likelihood of a structure tensor in the LE-space with multivariate Gaussian distribution is given by the following equation:

$$f_c(\vec{Y}_{LE}|\vec{\mu_c}, \Sigma_c) = \frac{1}{(2\pi)^{N/2}|\Sigma_c|^{1/2}} exp(-\frac{1}{2}(\vec{Y}_{LE} - \vec{\mu}_{LE})^t \Sigma_c^{-1}(\vec{Y}_{LE} - \vec{\mu}_{LE})). \quad (8)$$

The most probable class $\hat{c}_{opt}$ for $\vec{Y}_{LE}$ is given by :

$$\hat{c}_{opt} = \underset{c \in L}{argmax} \quad (p(c|Y)). \quad (9)$$

Gaussian mixtures The classes of interest are not necessarily uniform in terms of texture, for instance leaves sometimes present different properties depending if it is the upper or lower side that is visible, a better representation of the distributions of structure tensors within diverse classes can be Gaussian mixtures. Gaussian mixtures are composed of independent Gaussian density functions each representing a sub-part of the whole distribution. A mixture of K Gaussian probability density functions is given by :

$$f(\vec{Y}_{LE}|(\omega_k, \vec{\mu}_k, \Sigma_k)_{k=1:K}) = \sum_{k=1}^{K} \omega_k p_k(\vec{Y}_{LE}|\vec{\mu}_k, \Sigma_k), \quad (10)$$

the parameters $\omega_k > 0$ are the weights of sum equal to 1. The mixture model parameters ω_k , $\vec{\mu}_k$, and Σ_k are estimated by employing the expectation-maximization algorithm (Titterington *et al.*, 1985).

3.5 Post-processing : spatial regularisation

The classification process employed is a probabilistic decision made independently for each pixel, without considering the decisions reached for its neighbours. Nevertheless,

flora images naturally present spatial organisations into arrangement of organs having locally homogeneous structural properties . Therefore it is very unlikely to observe sparse distributions of labels within continuous regions. However the proposed classification can produce such erratic results. In order to enhance the efficiency/veracity of this classification, authors propose two methods to perform spatial regularisation.The first one is based on Markovian fields and ICM algorithm, the second one is based on mathematical morphology.

Stochastic relaxation: ICM algorithm (iterated conditional modes) In this process, classification results (*i.e.* field of labels) are considered as Markov Random Fields (MRF) where each label depends only on the labels of its direct neighbours (8-connectivity cliques). ICM algorithm is an optimisation algorithm designed to reach the most stable field of labels regarding the underlying parametric model and local dependencies of labels (Besag, 1986). It is essentially a trade-off between statistical classification and spatial coherence. It usually results in smoother classifications. The ICM algorithm runs with an autologistic potential and 8-connectivity cliques of two pixels.

Mathematical morphology This regularization process aims also at discarding sparse distributions, but for larger objects (connected components), *i.e.* misclassified group of labels whose neighbours are also misclassified. In practice this consists in filling gaps and holes in continuous regions and removing small connected components that differ from the main region thanks to opening and closing operations (Serra, 1986).

3.6 Analytical protocol

Different variants are proposed for each step of the processing chain. The purpose of this analysis is to compare different combinations of these variants in terms of classification performances. The $LEEST$ and $CELEST$ representations are compared for decisions based either on Gaussian MultiVariate (mvG) PDF or on Gaussian mixtures (mvGM) with various managements of a priori probabilities and regularisations.

The analysis is conducted for three phenological stages : flowerhoods falling (BBCH 68), pea-sized berries (BBCH 75) and majority of berries touching (BBCH 79). For each stage four classes are considered : leaf core, leaf edges, grape bunches / inflorescence and stems (leaf edges are differentiated in the modelling process from leaf core so that some external

parts of the foliage are not confused with stems because both objects are a sort a frontier between foreground and background). For each stage, 16 manually labelled images with around 2.10^5 sample sites manually selected per image per class are used for the estimation of models. Performances are evaluated with a *leave-one-out cross-validation* process, where each image to be tested is iteratively removed from the learning database on which models are computed. Performances on each image are produced by comparing the groundtruth pixels with the classes determined by the processing chain in confusion matrices. Resulting from the confusion matrix, Overall accuracy (OA) summarise the global performances reached for all the classes in the confusion matrix. OA represent the proportion of correctly classified instances over the total number of instances. In addition to OA, two different metrics are used to describe performances for each class, Precision and Recall. Precision represents for a class, the fraction of relevant instances among the retrieved instances. Recall represents for a class, the fraction of relevant instances that have been retrieved over the total amount of relevant instances.

A confusion matrix, is a specific table layout that allows visualisation of the performance of a classification algorithm. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class. From a confusion matrix, performances metrics are defined from four values, True Positive rate (TP , True Negative rate (TN), False Negative rate (FN) and False Positive rate (FP) as :

$$\text{For a class } k: \qquad Precision_{(k)} = \frac{TP_k}{TP_k + FP_k}; \qquad Recall_{(k)} = \frac{TP_k}{TP_k + FN_k}; \tag{11}$$

4 Results

Figure 3 presents and summarises the outputs of each step of the processing chain presented in figure 2. The presented results are produced for stage BBCH 79 (majority of berries touching) with the best performing variant of the processing chain which are described in table 2 ($CELEST$ + mvGM+sprob+ICM+Morph). The following parts of the result section aim at describing and discussing the results obtained at each step of the

processing chain, depending on the parameters and different proposed variants.

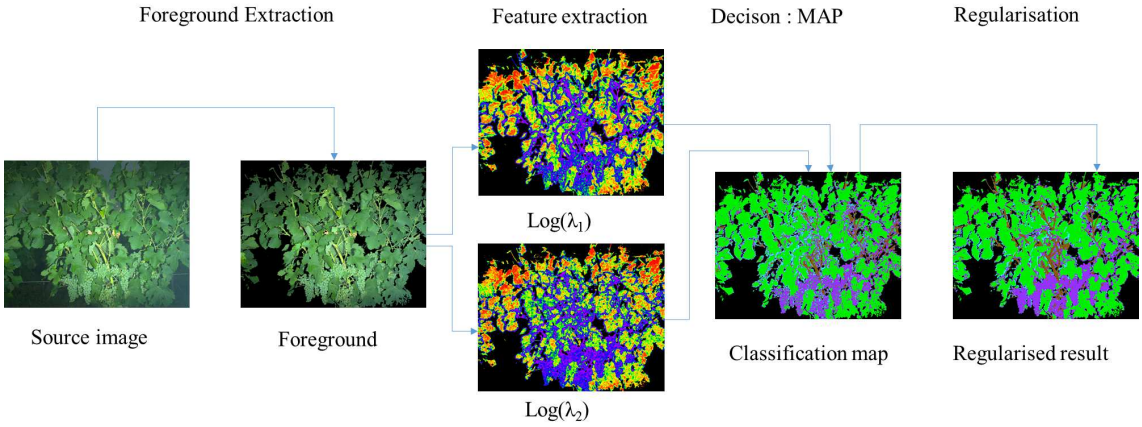


FIGURE 3 – Outputs of the processing chain for stage BBCH 79 (berries touching) obtained with the best performing variants for $CELEST$ representation.

4.1 Pre-processing : discarding background components

The use of a powerful flash during short exposure time brightens the foliage and softens shadows while the background is out of range and remains then in the dark. Therefore it possible to discard the entire background, even when it contains green objects such as grass and foliage of further rows because their V and S components are much smaller. Results of pre-segmentations are visible on figure 4 where all pre-segmented parts of the image are replaced with black pixels. The thresholds for both H, S and V channels are selected thanks to the histograms of each channel. Figure 5 presents the histogram of Hue values from 1000 images at stage BBCH 79 (berries touching). The histogram shows two Gaussian-like modes, the first distribution with the bigger amplitude is mainly composed of foliage pixels and the second one is composed of background pixels. The H threshold is then determined by the limits of the main Gaussian mode *i.e.* $H \in [45, 80]$. The same phenomenon is observed for S and V channel for all phenological stages, S and V threshold are then determined for $S \in [65, 230]$ and $V \in [80, 230]$

(a) (b)

FIGURE 4 – Example of pre-processing : HSV thresholding of the background components. (a) original image at stage BBCH 79 (berries touching) and (b) foreground extracted.

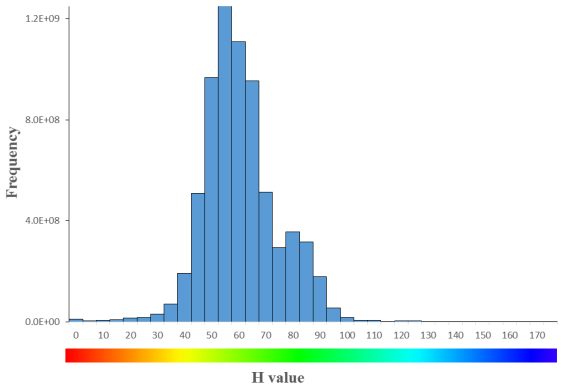
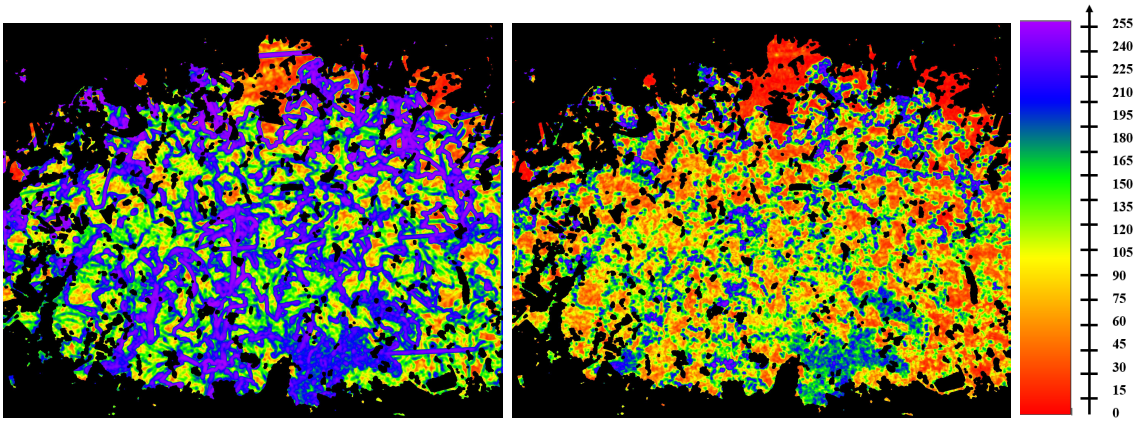


FIGURE 5 – Hue histogram of the dataset

4.2 Features extraction

The features used to extract and describe the different textures of organs are composed of the colour components visible on RGB images and of structural components resulting from Local Structure Tensors. In the $CELEST$ representation, it is possible to visualise the differences in structural properties of organs with two of its components $log(\lambda_1)$ and $log(\lambda_2)$ normalised on a $0-255$ scale. Figure 6 presents an example of structural maps produced from the original image on figure 4. Figure 6 shows that stems are characterised by their great anisotropy, presenting high values for λ_1 and very low values for λ_2. Leaves which are smooth with isotropic textures present low values for both eigenvalues *i.e.* low gradients in all directions except for their edges. As for grape bunches which are characterised by the radial symmetry of berries and chaotic regions in between berries, they present high

445 values for both components.



(a) (b) (c)

FIGURE 6 – structure components of *CELEST* representation (computed with $[\sigma_1 = 3.5, \sigma_2 = 5.5]$). (a) is a map of normalised $log(\lambda_1)$, (b) represents $log(\lambda_2)$ and (c) the normalised $[0 - 255]$ scale.

446 4.2.1 Optimal scale for the extraction of structural properties

447 The computation of structure tensors as defined in equation (1) depends on two scale
448 parameters. σ_g determines the scale at which image gradients are computed. σ_t determines
449 the scale at which structure information within a set of gradients is pooled into a structure
450 tensor. While the former should be chosen according to the size of the elementary observable
451 patterns, the choice of the latter should be related to the scale at which texture (*i.e.* local
452 organisation of patterns) is observable. The choice of these two scale parameters may thus
453 affect the descriptive capabilities of the structure tensor and may differ according to the
454 class of interest or the vegetative stage.

455 For example, at small scales it is the *granular* appearance of leaves with sparse veinlets
456 that is described, whereas at larger scales the structure tensors describe mainly larger
457 features. Similarly, small scales describe textural properties within a berry or a flower
458 when larger scales describe more entropic patterns containing several berries, stalks or
459 peduncles.

460 It is then not so obvious to determine the scale at which textural properties best
461 describe and discriminate classes while being robust enough to local noises. Figure 7
462 illustrates the multi-scale behaviour of structure tensors as energy maps (sum of
463 eigenvalues of LST's) for 3 couples of scale parameters (b,c,d) for original image (a) at

stage BBCH 68 (flowerhoods falling). Energy values have been normalized between 0 and 255 and represented using a coloured palette (Fig. 7e). On the three different maps (b,c,d) it can be clearly identified that the inflorescence visible in figure 7(a) is highlighted for energies between 150 and 255 when the underlying leaf could be segmented for values ranging between 15 and 90. However it is not possible to determine which scale provides the most reliable classification. It is then necessary to determine analytically for each phenological stage, the optimal couple (σ_g, σ_t) that offers the best trade-off in terms of performance for all classes. Scale parameters should be chosen to maximize the performances for the classes of primary interest (leaf cores and fruits) while ensuring reasonable performances for the classes of secondary interest (leaf edges and stems). Figure 8 presents the relative performances obtained for different couples of scale parameters, results are presented only for the stage *berries touching* (BBCH 79), they are obtained with the $CELEST$ representation for mvG without any regularisation.

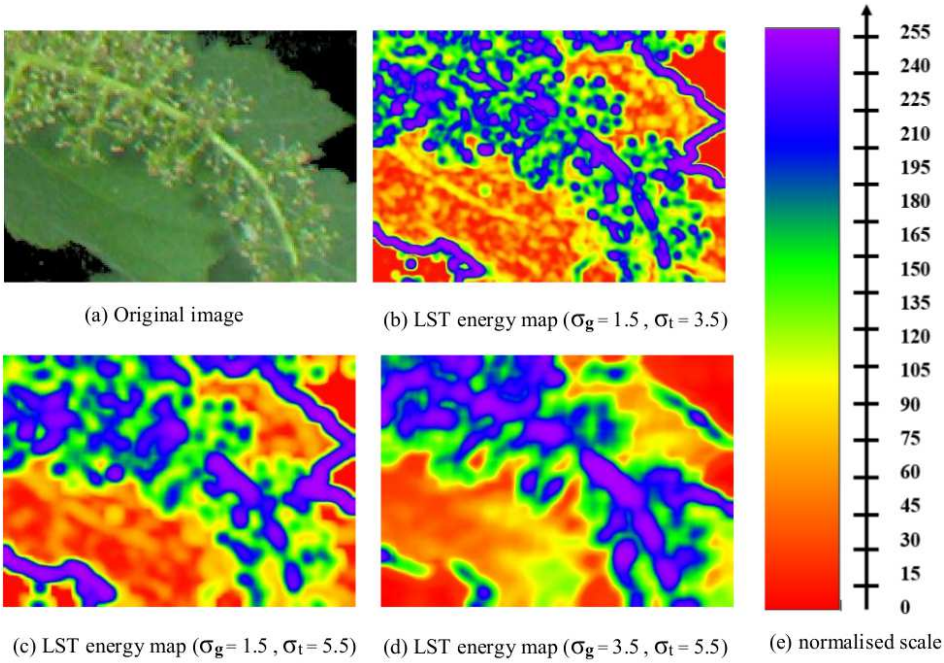


FIGURE 7 – An example of the multi-scale properties of LST applied to stage BBCH 68 *i.e.* flowerhood falling, (b,c,d) are normalised energy maps resulting from original image (a) according to scale parameters, (e) is the normalised displaying scale.

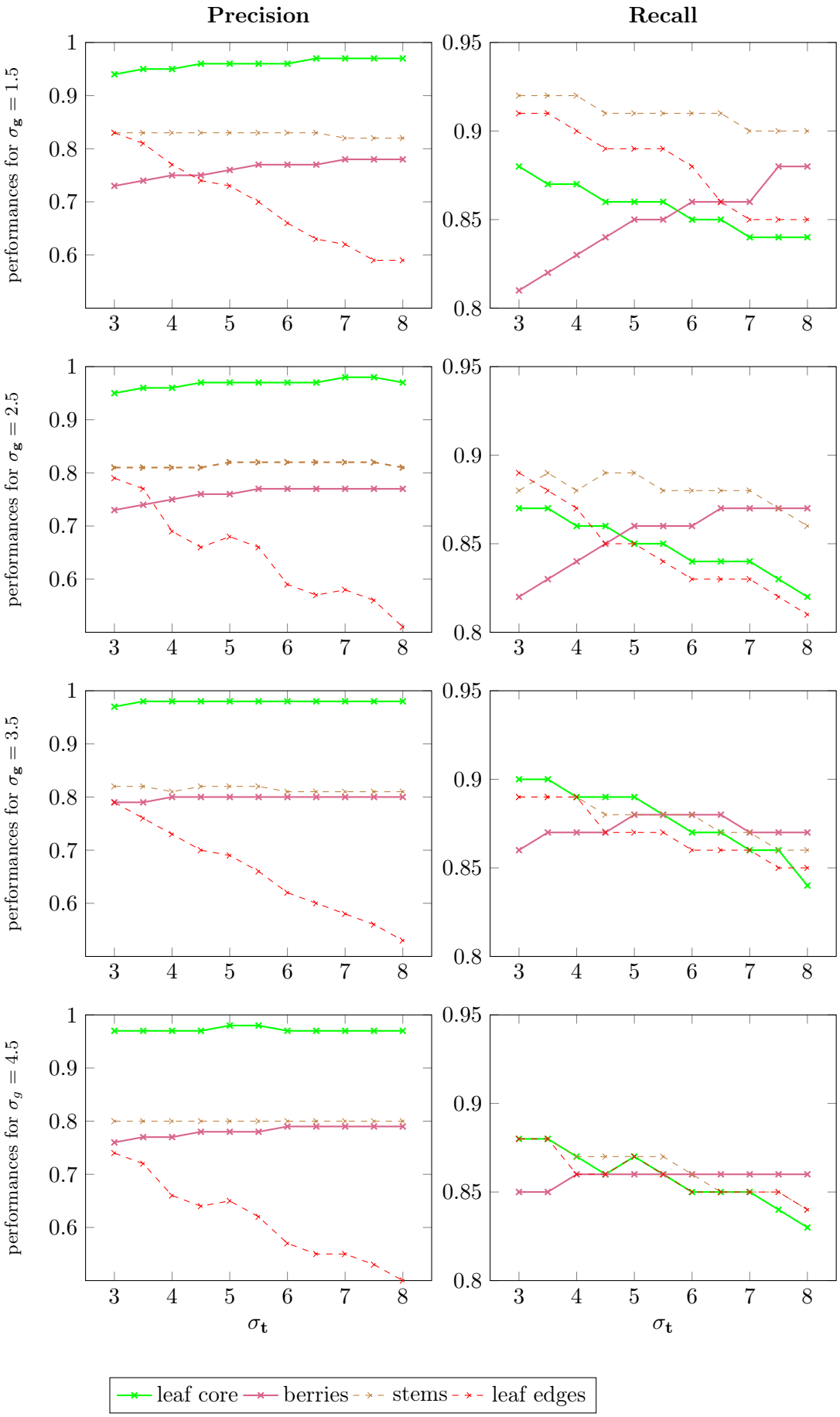


FIGURE 8 – Influence of σ_t scale for different values of σ_g on precision and recall metrics for stage BBCH 79 (berries touching)

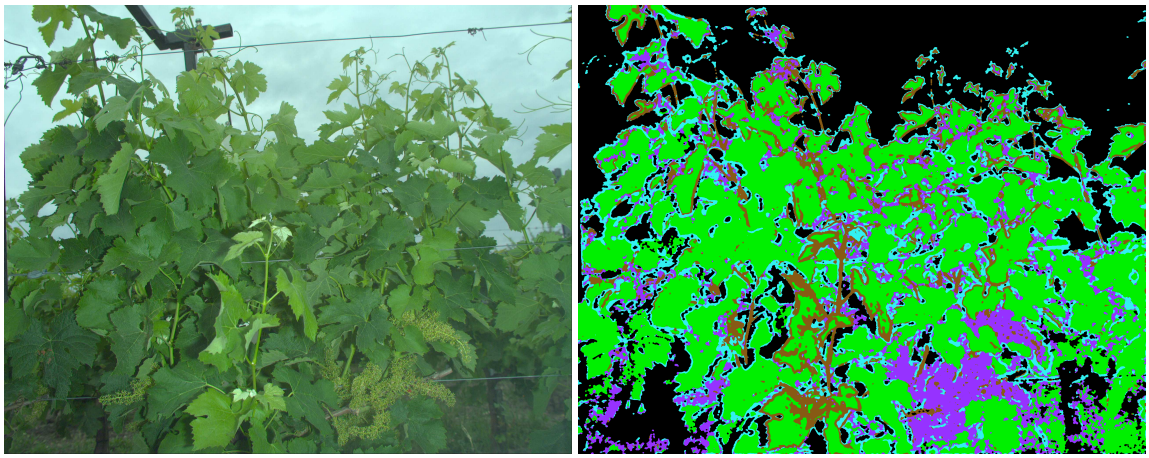
For all classes except *leaf edges*, the precision metric is very stable to variations of tensor scales σ_t given a fixed scale of gradient σ_g. Leaf edges being the smallest structures (<4px), it is then very difficult to extract its properties for scales larger than their size. In practice the proportion of leaf edges classified as leaf core is higher with growing values of σ_g and σ_t. The only case where an 80% precision is reached for berries is for a gradient scale $\sigma_g = 3.5px$, which is also the scale that maximises the precision for leaf cores that reached 98% while ensuring a minimum of 80% for stems.

Recall rates are less stables, they tend to decrease by around 5% for stems and leave cores for increasing values of σ_t while increasing about the same amount for berries. A compromise has then to be found to ensure altogether the maximisation of recall rates for both primary classes. Such a compromise can be found at the intersections of the berry and the leaf core curves for σ_t ranging between 5.0 and 5.5. Eventually the optimal couple is ($\sigma_g = 3.5px$ and $\sigma_t = 5.5px$) that both maximises recall for berries (88%) while ensuring recall above 88% too for leaf cores.

The optimal parameters are then similar for both precision and recall metrics. In the following, performance tests are conducted with the couple ($\sigma_g = 3.5$ px ; $\sigma_t = 5.5$ px).

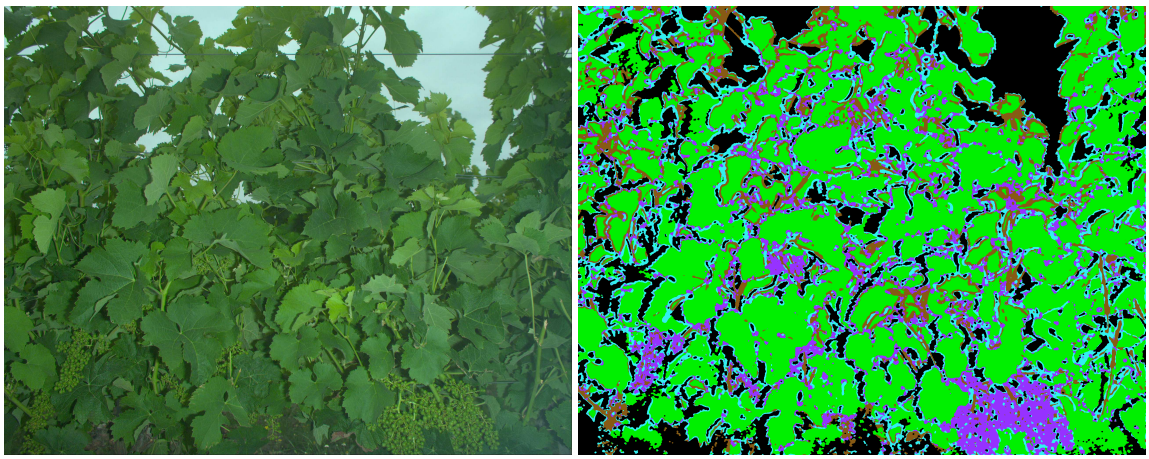
4.3 Decision : pixel-wise classification for Bayesian MAP estimator

Figure 9 shows examples of pixel-wise classification maps obtained from Bayesian MAP estimations for the three phenological stages of interest. On these maps, each pixel is assigned a colour corresponding to the class having the maximum likelihood. Results are produced with $CELEST$ representation. The raw classification results are overall satisfying. It is possible at this stage of the processing chain to recognize the shapes of organs. However the results present a certain amount of noise and inconsistencies inside regions that are supposed to be continuous. These errors will be corrected by the post-processing.



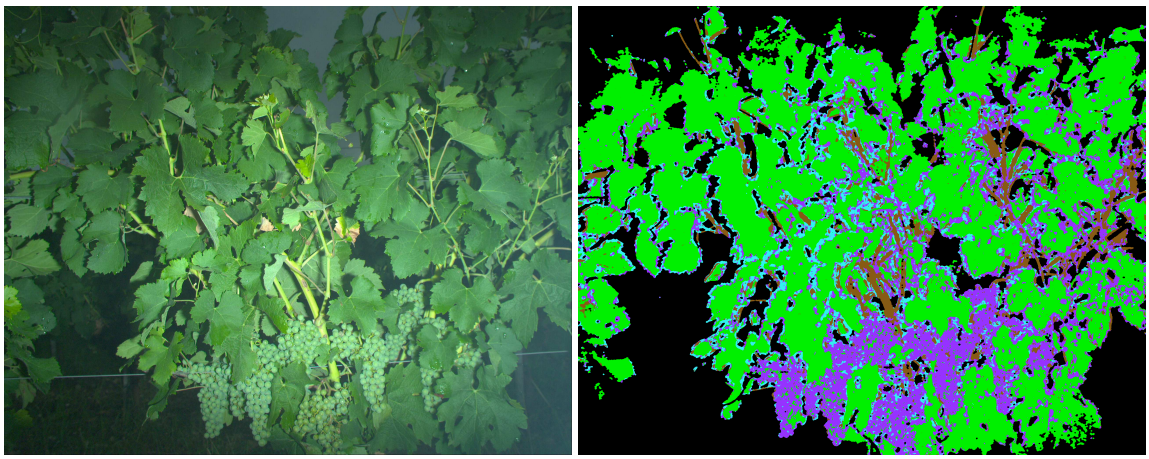
(a) Original image BBCH 68

(d) *CELEST* classification map BBCH 68



(b) Original image BBCH 75

(e) *CELEST* classification map BBCH 75



(c) Original image BBCH 79

(f) *CELEST* classification map BBCH 79

Leaf Cores Berries Stems Leaf edges

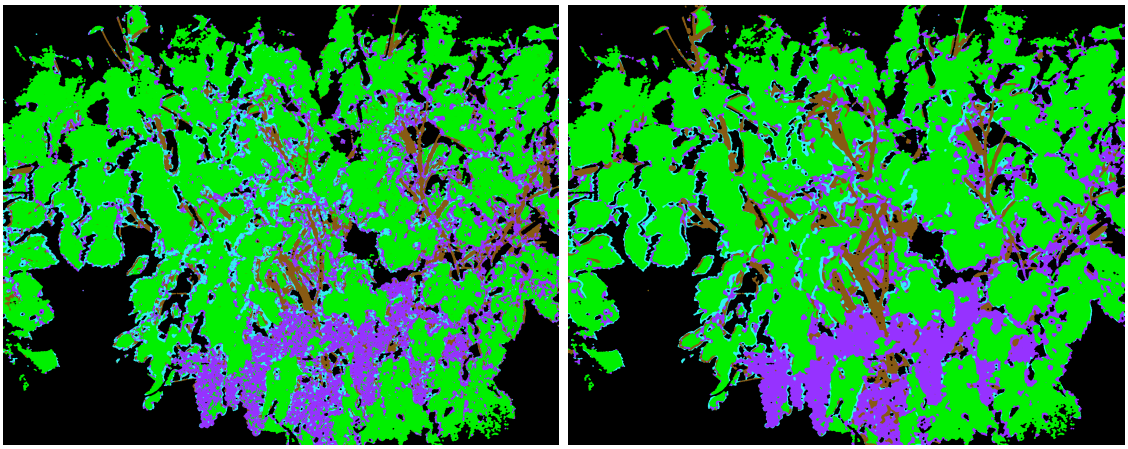
FIGURE 9 – Examples of images and pixel-classification maps obtained with the CELEST and mvGM variant before regularisation, implemented for stages *flowerhoods falling* BBCH 68 (a-d), *pea-sized* BBCH 75 (b-e) and majority of *berries touching* BBCH 79 (c-f).

4.4 Post-processing : Probabilistic relaxation and mathematical morphology

Figure 10 presents the corrections operated by the post-processing steps on a an example of classification map produced with the Map estimator. The example on figure 10 (a) is the result with *CELEST* representation presented for stage BBCH 79 in figure 9 (c). Figure 10 (b) is the output of probabilistic relaxation conducted with ICM Algorithm and morphological filtering applied to the image in figure 10 (a).

This post-processing produces more spatially coherent results. Indeed the post-processed images present more homogeneous regions without the sporadic variabilities of classes produces by the raw decisions. The redundant errors, such as small clusters of pixels classified as leaf or leaf-edges contained within grape bunches or occurrences of stems and grapes at the edges of leaves are corrected. However, in some cases the post-processing can oversimplify the contours of irregular objects and then produce additional errors. Nevertheless, in such images with rather large objects, the proportion of contours is substantially lower compared to the cores of objects.

Eventually this post-processing improves the performances even if it not its primary purpose. The improvement in performances can be quantified in regards to the validation dataset and is shown in table 1. Table 1 compares the confusion matrices and metrics resulting from the 16 validation images for the stage BBCH 79 before and after post-processing. The results are computed from about 3.2×10^6 labelled pixels which represent roughly 1/12 pixels from the foreground of the 16 validation images. The overall accuracy is increased from 92% to 94% after post-processing. It is mainly due to a significant improvement of both precision and recall for leaves, berries and leaf edges. For instance about 50% of the instances of leaves misclassified as berries and 68% of the instances of berries misclassified as leaves are corrected by the post-processing. Similarly 70% of the instances of leaf edges misclassified as berries and 50% misclassified as stems are also corrected. However the errors in instances of stems misclassified as leaves increased by 25% which affected the precision and recall rates for stems that drops both of 2%.



(a) CELEST classification map BBCH79 (b) post-processed map BBCH 79

FIGURE 10 – Example of regularisation result (b) obtained from pixel-wise classification map at stage majority of berries touching BBCH 79 (a) with *CELEST* representation

TABLE 1 – Impact of post-processing on classification maps , example for an image at stage BBCH 79 (before post-processing (a) / Fig. 10 (a), after post-processing (b) / Fig. 10 (b)

(a) without post-processing

		Predicted classes						
		Leaf Core	Berries	Stems	Edges	Total	Recall	OA
True class	Leaf Core	1702436	50162	9269	5131	1766998	0.96	0.92
	Berries	49196	690458	2020	10292	751966	0.92	
	Stems	29106	18887	379623	11733	439349	0.86	
	Edges	37791	15951	9248	172535	235525	0.73	
	Total	1818529	775458	400160	199691			
	Precision	0.94	0.89	0.95	0.86			

(b) after post-processing

		Predicted classes						
		Leaf Core	Berries	Stems	Edges	Total	Recall	OA
True class	Leaf Core	1709775	23367	23543	10313	1766998	0.97	0.94
	Berries	16769	722641	1678	10878	751966	0.96	
	Stems	39259	15579	371062	13449	439349	0.84	
	Edges	15074	6612	4726	209113	235525	0.89	
	Total	1780877	768199	401009	243753			
	Precision	0.96	0.94	0.93	0.86			

4.5 Comparison of feature representations, decision and regularisation methods

LEEST and *CELEST* representations are two variants of descriptors joining colorimetric and structural information. They differ from their dimensions and their

robustness to rotation. *LEEST* is an extensive representation with 15-dimensions, while *CELEST* is a compact 5-dimensions and rotation invariant representation. *CELEST* representation could prove more convenient to estimate model parameters with limited number of samples for the learning phase and more robust to numerical instability, especially since it involves the computation and inversion of covariance matrices. Table 1 presents a comparison between the performances achieved with simple colorimetric descriptors (RGB) or simple structural descriptors (LST) and the performances achieved with *LEEST* and *CELEST* representations.

TABLE 2 – Comparison of the different representations of structure tensor. Results are obtained with mvG for stage BBCH 79 (berries touching) without any regularisation

Representation	Precision				Recall				OA
	leaf core	berries	stems	edges	leaf core	berries	stems	edges	
RGB	0.48	0.37	0.29	0.17	0.69	0.51	0.43	0.22	0.41
LST	0.96	0.61	0.68	0.27	0.79	0.74	0.53	0.55	0.72
LEEST	0.95	0.79	0.80	0.45	0.86	0.79	0.81	0.75	0.83
CELEST	**0.98**	**0.80**	**0.82**	**0.67**	**0.88**	**0.88**	**0.89**	**0.87**	**0.88**

Alone, colorimetric information is not sufficient to describe and discriminate the grapevine organs, it results in random classifications with an overall accuracy of 40%. The pure structural information provided by LST's better describes the textural properties encountered. Indeed the overall accuracy exceeds 70%. However it is also insufficient to achieve a satisfying classification (recall rates are all below 80%, precision is below 65% for berries). The colour extension of structure tensor is essential to capture the distinctive properties of the textures appearing on grapevine images. Both *LEEST* and *CELEST* representations improve the classification performances with overall accuracy of 83% and 88% .

CELEST representation, while being a compact representation, proves to perform always better than the extensive *LEEST* representation, especially in terms of recall rates for the berry class. Table 2 presents the performances achieved with different variants in the processing steps applied for *CELEST* representation. In the following table, *mvG* stands for multivariate Gaussian, *mvGM* for multivariate Gaussian mixture, *sprob* refers to the spatial management of a priori probabilities and *Morph* to morphological filtering.

TABLE 3 – Comparison of the different variants for *CELEST* representation at stage BBCH 79 (berries touching)

Methods			Precision				Recall				OA
Representaiton	Decision	PP	leaf core	berries	stems	edges	leaf core	berries	stems	edges	
	mvG	∅	0.98	0.80	0.82	0.67	0.88	0.88	0.89	0.87	0.88
	mvG + sprob	∅	0.94	0.88	0.89	0.82	0.95	0.89	0.86	0.76	0.90
	mvG + sprob	ICM	0.97	0.85	0.86	0.77	0.94	0.90	0.89	0.90	0.91
CELEST	mvGM	∅	0.98	0.83	0.87	0.75	0.90	0.91	0.92	0.86	0.90
	mvGM + sprob	∅	0.94	0.89	0.95	0.86	0.96	0.92	0.86	0.73	0.92
	mvGM+ sprob	ICM	0.91	0.94	0.95	0.81	0.97	0.93	0.84	0.93	0.92
	mvGM + sprob	ICM + Morph	0.96	0.94	0.93	0.86	0.97	0.96	0.84	0.89	0.93

The proposed method for regularisation and for the management of a priori probabilities tends to improve selectively the performances for some classes while slightly decreasing performances for other classes. However the combination of these propositions improve results in all classes. The use of multivariate Gaussian mixtures results in better performances for all classes. The mixtures are estimated with $K = 3$, mixtures between 2 and 5 Gaussians were tested, ultimately $K = 3$ provided best results. When combining all propositions with mvGM, performances metrics are above 90% for the primary classes. Overall, each step leads to a slight increase in the global performances for all classes as it is shown by the evolution of overall accuracies from 88% with the ra map to 93% with the full variant.

4.6 Robustness of the representations for different phenological stages

Metrics	Precision				Recall				OA
Phenological stage	Leaf core	Berries	stems	Leaf edges	Leaf core	Berries	stems	Leaf edges	
Flowerhoods falling (BBCH 68)	0.90	0.88	0.93	0.84	0.94	0.90	0.83	0.75	0.91
Pea-sized berries (BBCH 75)	0.86	0.82	0.89	0.69	0.84	0.92	0.73	0.77	0.88
Berries touching (BBCH 79)	0.96	0.94	0.93	0.86	0.97	0.96	0.84	0.89	0.93

TABLE 4 – Comparison of best performances achieved with CELEST mvGM and full regularisation (ICM relaxation + morphological post-processing) for three phenological stages.

$CELEST$ representation produces recall and precision performances over 80% and up to 97% for the three phenological stages. However classification performances tend to be lower for the earlier phenological stages. It could possibly be explained by the greater variability of morphologies of the leaves and berries encountered during early stages which can present textural properties in between two transitioning morphologies. It has to be noted that the best performances obtained for stage BBCH 75 (pea-sized) are achieved without any spatial considerations for *a priori probabilities*. In this case, it tends to lower performances, it is mainly due to a higher variability of the spatial repartition of grape bunches and stems that produces less coherent estimations of $sprob : \pi_c(h)$.

5 Conclusion

In order to solve common problems regarding the classification of objects within natural outdoor images of grapevines in proximal sensing, a new framework has been proposed. The proposition is based on the joint parametric modelling of structure and colour.

The proposed framework includes two vector representations $LEEST$ and $CELEST$, both based on colour extended structure tensors which were modelled with multivariate Gaussians in the Log Euclidean space. In addition, a spatial management of a priori probabilities in the MAP estimator and two methods of spatial regularisation (ICM algorithm and mathematical morphology) were tested to improve the results.

$CELEST$ representation, in comparison with other tensor-based approaches, produces compact and rotation invariant descriptors respecting the riemannian geometric properties of structure tensors. These properties enable the estimation of coherent and numerically stable parameters for the models that produce reliable classifications with reasonable learning samples. The MAP estimator based decision process includes management of a priori probabilities accordingly to spatial considerations that improves performances. In addition the decision system requires few manual settings from users. The only parameters to be tuned by user are rather intuitive scale parameters highly correlated to texture sizes. Moreover the $CELEST$ representation is quite robust to these scale parameters. The proposed framework is easily applied to different

phenological stages with satisfying results in each case.

The proposed framework was tested on a considerable number of pixel samples but in a limited number of images. However, despite the low number of images used for the learning and test phases, models and performances are relatively stable for considering a single variety with uniform stages and cultivation methods. To a certain extent, in a uniform case, images present strong similarities in terms of textural properties. It can then be considered that the sampling could represent the larger dataset and is robust to variations of morphologies. While the developed process can be applied to any number of images and provide complete classification maps, it is not possible to assess its performances for a more representative number and variety of images without more abundant ground-truthing. Such a database could enable to compare the proposed method with reference classification methods such as SVM and Neural networks fed with similar textural / structural and colorimetric features. The robustness of the method and its models to varying varieties of grapevine or different cultivation systems are not tested yet. A complementary study is necessary to determine the versatility and the amount of data indexing required to apply this method at the scale of a vineyard.

One of the major upcoming challenge is to transcript the statistical parameters that can be estimated from the classification results (Leaf area, number and size of grape bunches, gaps in the canopy ect.) into formal agronomic parameters. This essential step for the development of innovative PV applications requires the acquisition of field data and direct measurements to establish correlations between what is estimated with image processing and what can be measured on the plots with well-acknowledged methods.

For now the framework concerns merely computer vision and was only applied to healthy vinestocks. As a perspective it could be considered to apply the same framework with plants presenting symptoms of fungal diseases such as powdery and downy mildew or Blackrot. Finally the proposed framework could be easily transposed to crops with similar structures such as fruit trees.

6 Acknowledgement

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 731884. Authors would like to thank the French National Institute for Agricultural Research (INRA, UMR SAVE 1065 and UMR EGFV 1287) and the French National Institute of Vines and Wine (IFV) who made available the vineyards and the farm equipments and who also provided data and expertise regarding the monitoring of agronomic parameters of the plots under study.

7 References

Abdelghafour, F., Keresztes, B., Germain, C. and Da Costa, J.-P. 2017. Potential of on-board colour imaging for in-field detection and counting of grape bunches at early fruiting stages. Advances in Animal Biosciences, 8(2), 505-509. doi:10.1017/S2040470017001030.

Abdelghafour F., Rosu R., Keresztes B., Germain C. and Da Costa J.-P. 2018. Joint structure and colour based parametric classification of grapevine organs from proximal images through several critical phenological stages. 14th Int. Conf. on Precision Agriculture, June 24-June 27, 2018 Montreal, Canada.

Arsigny, V., Fillard, P., Pennec, X. and Ayache, N. 2006. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56, 411-421. https://doi.org/10.1002/mrm.20965

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. 2007. Geometric means in a novel vector spacestructure on symmetric positive-de ?nite matrices. SIAM journal on matrix analysis and applications, 29(1):328-347

Besag, J.E.1986. On the Statistical Analysis of Dirty Pictures. Journal of the Royal Statistical Society, Series B, 48 (3): 259-302, JSTOR 2345426

Bigun, J., Granlund, G., and Wiklund, J. 1991. Multi-dimensional orientation estimation with applications to texture analysis and optical flow. IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, no. 8, pp. 775-789, Aug. 1991.

Bramley, R.G.V. 2010. Precision Viticulture: Managing vineyard variability for improved

quality outcomes. In: Managing Wine Quality Volume 1. Viticulture and wine quality. A.G. Reynolds (Woodhead Publishing: Cambridge) pp. 445-480.

De Luis-Garcia, R., Deriche, R. and Alberola-Lopez, C. 2008. Texture and color segmentation based on the combined use of the structure tensor and the image components. Signal Processing 88, 776-795. https://doi.org/10.1016/j.sigpro.2007.09.019

Demestihas, D., Debuuisson, S and Descotes, A. (2018). Decomposing the notion of vine vigour with a proxydetection shoot sensor: Physiocap. E3S Web Conf. Volume 50, XII Congreso Internacional Terroir. https://doi.org/10.1051/e3sconf/20185003003

Keresztes, B., Abdelghafour, F., Randriamanga, D., Da Costa, J.-P. and Germain, G. 2018. Real-time Fruit Detection Using Deep Neural Networks. In Proceedings of the 14th International Conference on Precision Agriculture. Monticello, IL: International Society of Precision Agriculture.

Knutsson, H.1989. Representing local structure using tensors. in Proc. 6th Scand. Conf. Image Anal., Oulu, Finland, Jun. 1989, pp. 244-251.

Liu, S., Whitty, M., and Cossell, S. 2015. A Lightweight Method for Grape Berry Counting based on Automated 3 D Bunch Reconstruction from a Single Image. Workshop on Robotics in Agriculture, Seattle, USA, May 2015.

Liu, S., Cossell, S., Tang, J., Dunn, G. and Whitty, M. 2017. A computer vision system for early stage grape yield estimation based on shoot detection. Computers and Electronics in Agriculture, 137, 88-101. doi:10.1016/j.compag.2017.03.013

Mathews, A.J. and Jensen, J.L.R. 2013. Visualizing and quantifying vineyard canopy LAI using an Unmanned Aerial Vehicle (UAV) collected high density structure from motion point cloud. Remote Sens. 2013 (5), 2164-2183. http://dx.doi.org/10.3390/rs5052164.

Nuske, S., Gupta, K., Narasimhan, S. and Singh, S. 2012. Modeling and Calibrating Visual Yield Estimates in Vineyards. Field and Service Robotics 343-356.

Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U. and Weber, E. 1995. Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L.

ssp. vinifera)-Codes and descriptions according to the extended BBCH scale.

Pfeiffer, S. A., Guevara, J., Cheein, F. A. and Sanz, R. 2018. Mechatronic terrestrial LiDAR for canopy porosity and crown surface estimation. Computers and Electronics in Agriculture, 146, 104-113. doi:10.1016/j.compag.2018.01.022.

Rosu, R.G., Da Costa, J.-P. and Donias, M. 2016. Structure tensor Log-Euclidean statistical models for texture analysis. IEEE, pp. 3553-3557. https://doi.org/10.1109/ICIP.2016.7533021

Rosu, R., Donias, M., Bombrun, L., Said, S., Regniers, O. and Da Costa, J.-P. 2017. Structure Tensor Riemannian Statistical Models for CBIR and Classification of Remote Sensing Images. IEEE Transactions on Geoscience and Remote Sensing 55, 248-260. https://doi.org/10.1109/TGRS.2016.2604680

Serra J. 1986. Introduction to mathematical morphology. Computer Vision, Graphics, and Image Processing. Volume 35, Issue 3, September 1986, Pages 283-305. https://doi.org/10.1016/0734-189X(86)90002-2.

Titterington, D. M., Smith, A. F. M. and Makov, U. E. 1985. Statistical analysis of finite mixture distributions. John Wiley & Sons Ltd..

Taylor, J.A., McBratney, A.B. and Whelan, B.M. 2007. Establishing management classes for broadacre agricultural production. Agronomy Journal 99, 1366-1376.

Tisseyre, B., Ojeda, H. and Taylor, J. 2007. New technologies and methodologies for site-specific viticulture. Journal international des Sciences de la Vigne et du Vin 41, 63-76.

Vieri, M., Lisci, R., Rimediotti, M., and Sarri, D. 2013. The RHEA-project robot for tree crops pesticide application. Journal of Agricultural Engineering, 44(s1), 359-362.

Yalcin, H. 2015.Phenology monitoring of agricultural plants using texture analysis. Fourth International Conference on Agro-Geoinformatics (Agro-geoinformatics), Istanbul, 2015, pp. 338-342. doi: 10.1109/Agro-Geoinformatics.2015.7248114

Zhang, C., Craine W., Davis, J.B., Khot, L.R., Marzougui, A., Brown, J., Hulbert, S. H. and Sankaran, S., 2018. Detection of canola flowering using proximal and aerial remote sensing

733 techniques. Proc. SPIE 10664, Autonomous Air and Ground Sensing Systems for Agricultural
734 Optimization and Phenotyping III, 1066409 (21 May 2018) ; doi: 10.1117/12.2304054.