Joint structure and colour based parametric classification of grapevine organs from proximal images through several critical phenological stages

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.

Abstract. *Proximal colour imaging is the most time and cost-effective automated technology to acquire high-resolution data describing accurately the trellising plane of grapevine. The available textural information is meaningful enough to provide altogether the assessment of additional agronomic parameters that are still estimated either manually or with dedicated and expensive instrumentations. This paper proposes a new framework for the classification of the different organs visible in the trellising plane. The proposed method is an implementation of a Bayesian decision process based on a joint parametric representation of Local Structure tensors and color. The purpose is to obtain a pixel-wise description of grapevine images based on joint structural and colorimetric features. In this paper, a representation of colour extended structure tensors mapped into the log-Euclidean metric space is introduced. This new feature is used for the description of the textural properties of grapevine organs in multivariate Gaussian models. The final classification is performed by Bayesian MAP estimation based on the models. The paper presents and compares different variants of the method which are applied to three key phenological stage: flowerhood falling, pea-sized and berries touching (BBCH 68, 75, 79). The resulting classification performances are measured in terms of recall and precision that reached overall between 80% and 90% depending on the stage. These results are produced with leave-one-out cross validations where models are estimated from 15 images per stage containing about 1.5e6 samples. The achievement of a reliable classification of the leaves, flowers and berries for each vinestock is an integral step toward the estimation of leaf area index, leaf porosity, fruitfulness, cluster structuration and yields. These are key parameters for the monitoring and evaluation of main field works such as fertilisation, irrigation, and trimming, defoliation, trimming and thinning. In addition the modeling of healthy grapevine organs is also preliminary to achieve a modeling and classification of grapevine major fungal diseases.*

Keywords. *Proximal sensing, parametric classification, structure tensor, grape and foliage detection.*

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). With adequate technologies and innovative cultivation systems, essential field operations such as fertilization, irrigation, pruning, defoliation, trimming, thinning and harvest can be optimized with local and Variable Rate Applications (Tisseyre et al., 2007). The efficiency of these PV applications relies mostly on the abundance, reliability and resolution of in-vivo measurements of basic agronomic parameters (Taylor et al., 2007). In this context, image processing has been proven one of the most promising automated and non-intrusive techniques. It enables the observation of local variations of vine properties for large acreage, with relatively low costs in terms of instrumentation, labour and time duty. Image processing has already been successfully applied to in-field yield estimations (Keresztes et al., 2014; Liu et al., 2013; Nuske et al., 2012). However, current methods are dedicated to late fruiting phenological stages around the ripening season. In addition, the most thorough and advanced methods rely on heavy machine learning processes and abundant databases. Authors previously proposed a colour and texture based machine learning method, able to detect inflorescences and grape clusters at early fruiting stages (Abdelghafour et al., 2017). While producing valuable results at these stages, this method has shown to be difficult to tune and adapt to other phenological stages. In this paper a new method is proposed for the classification of all organs present on the trellising plane and throughout different phenological stages thanks to on-board colour imaging. The results can serve the assessment of additional agronomic parameters such as Leaf Area Index (LAI), canopy porosity, fertility and bud fruitfulness which are essential to new PV strategies (Bramley et al., 2011). In addition, it could provide feedback for operations such as trimming and defoliation and an effective tool for the assessments of seasonal damage to the crop such as hail, frost, coulure, millerandage and fungal diseases.

The proposed method is based on parametric models describing jointly structural and colorimetric information. It combines a log-Euclidean representation of local structure tensors with RGB data into a vector representation within joint multivariate Gaussian models. The distribution models are learnt on a collection of image patches of various 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 classification map that can be further regularized both by probabilistic relaxation and by morphological filtering.

Material and methods

Experimental setup

An image database has first to be constructed for the learning and validation steps of the proposed approach. Images should preferably present homogeneous properties in terms of resolution, distance and angle of capture and also in terms of illumination. Indeed, a consistent database results in more coherent models and better performances. The achievement of such prerequisites is not so trivial in uncontrolled and highly variable outdoor environments. Therefore, instrumentation and acquisition methods are designed to minimize the impact of natural light and atmosphere on the levels and variations of illumination in images in order to preserve their intrinsic textural properties.

Image aqcuisition: plant material and instumentation

The plant material is composed of two 0.2 ha plots with 120cm row-spacing and planted with the red wine grape variety Merlot Noir in "Le Domaine de la Grande Ferrade", a public experimental facility in Bordeaux area (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, as described by Lorenz et al. (1995), from "inflorescence swelling" to "half-ripening" (BBCH 53 to 83). 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 sppelite 750ii) with short exposure time (250-300μsec), a *12* V 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 is embedded on a vineyard tractor at 70cm above ground and at 50cm 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})
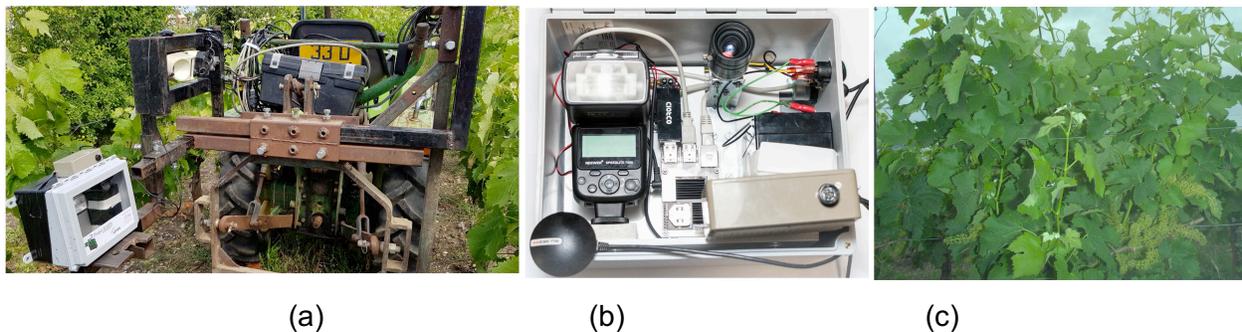

(a) (b) (c)

Figure 1- Instrumentation: embedding on a vineyard tractor (a), device compounds (b) and example of a resulting image (c)

Image processing methodology

The purpose of the proposed method is to provide a segmentation 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, stem 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 neighborhood 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 structure, anisotropy and colour. These properties are captured by an extended form of the Local Structure Tensor (LST). The following sub-parts aim at describing the different steps of the processing chain implementing the proposed method (Fig. 2).
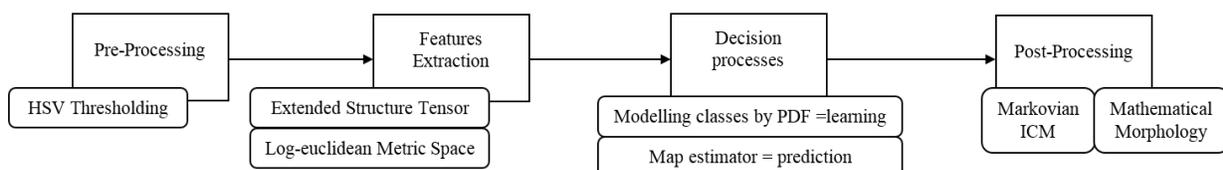

Figure 2- Processing chain

Pre-processing: thresholding non-vegetal objects

The pre-processing aims at narrowing down areas of interest by removing irrelevant parts of images, *i.e.* pixels not belonging to the canopy (sky, ground, trellising wires, poles, grass, wood stem).

Indeed, the classification process is based on a finite number of statistical models, in order to concentrate the subsequent classification on objects of interest. It is especially true when dealing with high resolution images and high dimension descriptors inducing heavy storage and computation power. Such areas can be easily discarded by simple filtering processes. Authors propose a simple thresholding in the HSV colour space followed by simple morphological operations. Hue and saturation channels enable to discard easily objects with colours which are implausible for plants, while the Value channel enables to discard objects whose illumination does not correspond to flora or where the proper structural patterns are not visible. In order to smooth results, avoid sporadic decisions and gaps, simple morphological operations (closing and opening) are implemented on raw thresholding results.

Features extraction

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

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 LST's is a two-step process, starting with estimating local gradients in the neighborhood of every pixel in an image. Given an image I of size [M*N], the gradient image ∇I is estimated as:

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

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.

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

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

Thus, for every pixel of coordinates (i,j) in the image $(i \in [1, N]; j \in [1, M])$ there is a corresponding local structure tensor $Y(i,j)$, in the form of a 2 by 2 symmetric matrix:

$$Y(i,j) = \begin{bmatrix} y_{xx}(i,j) & y_{xy}(i,j) \\ y_{xy}(i,j) & y_{yy}(i,j) \end{bmatrix}.\qquad(3)$$

Log-Euclidean metric space 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. 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 and then to apply standard probabilistic models (Rosu *et al.*, 2017) or geometric tools (Arsigny *et al.*, 2006).

Rosu *et al.* (2017) successfully applied these methods to the classification 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. Therefore these contributions motivate authors' choice of focusing 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}:

$$Y_{LE} = log_m(Y) = \begin{bmatrix} Y_{LExx} & Y_{LExy} \\ Y_{LExy} & Y_{LEyy} \end{bmatrix}, \quad (4)$$

Let consider the factorization RDR^{-1} where *D* 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*.

$$Y = RDR^{-1} \ ; \quad D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}. \quad (5)$$

The matrix logarithm of *Y* can be easily obtained by:

$$log_m(Y) = R(\log(D))R^{-1} \ ; \quad \log(D) = \begin{bmatrix} \log(\lambda_1) & 0 \\ 0 & \log(\lambda_2) \end{bmatrix}. \quad (6)$$

As mentioned in (Arsigny, 2006), a more convenient way to handle the matrix $log_m(Y)$ is to express it in the vector form:

$$\vec{Y}_{LE} = (y_{LExx}, y_{LEyy}, \sqrt{2}y_{LExy}), \quad (7)$$

which ensure that $\left\| \vec{Y}_{LE} \right\|_2 = \|Y_{LE}\|_2$.

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. The reverse mapping is obtained by applying the matrix exponential. The distance between two tensors Y_1 and Y_2 can simply be computed as the Euclidean distance between their equivalent vector representations on the LE space Y_{LE1} and Y_{LE2}.

Extended Structure Tensors, including colour information

De Luis-Garcia et *al.* (2008) proposed a method to join structural and colorimetric information into a single descriptor. The method consist 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 :

$$\vec{\nabla}I_{ce} = [I_x, I_y, R, G, B] . \quad (8)$$

The resulting structure tensor Y_{ce} is then a 5x5 SPD matrix representing the covariances of colour extended gradients. Alike the common LST, the colour extended structure tensor can be mapped 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. *LEEST* representation is a vector containing the spatial covariances between gradients and colour.

Decision process : classification of pixels

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 λ this pixel belongs to. This decision is based on a MAP estimator, a Bayesian method based on the maximization of $p(\lambda\,|Y)$, i.e. of the probability of λ knowing Y.

According to Bayes theorem:

$$\lambda_{opt} = \arg\max_{\lambda\in L} p(\lambda\,|Y) = \arg\max_{\lambda\in L} \frac{f(Y|\lambda)p(\lambda)}{f(Y)} = \arg\max_{\lambda\in L} f(Y|\lambda)p(\lambda), \tag{9}$$

where $f(Y|\lambda)$ is the likelihood resulting from the Probability Density Function (PDF) *f* describing the distribution of a subset of structure tensors Y in the class λ, and $p(\lambda)$ is *the a priori probability* of class λ. Both $f(Y|\lambda)$ and $p(\lambda)$ can be modelled thanks to representative samples of structure tensors in each class.

A priori probabilities

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

- Even distribution of classes: $p(\lambda_k) = \frac{1}{K}$, where K is the number of classes.
- Uneven distribution classes : $p(\lambda_k) = \pi_k$, the statistical frequency of the class k
- Heterogeneous distribution classes: $p(\lambda_k) = \pi_k(h)$, a function of space *i.e.* the a priori probability depends on the location of the pixel in the image.

The most genuine assumption is the latter, indeed, images of grapevine plants are spatially structured, within such images, the different type 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 judicious to consider a priori probabilities as functions of the height at which pixels are located.

Authors propose to vertically divide images into 3 parts of equal heights. Where for each part a different value of λ per class is estimated thanks to the average proportions observed on labelled images.

Probability Density Functions , parametric models

The distributions of structure tensors are represented by Gaussian multivariate distributions and Gaussian multivariate mixtures (Rosu et al., 2017).

A Gaussian multivariate distribution can be expressed with only 2 parameters, a Covariance matrix Σ and a centre of mass vector $\vec{\mu}$.

Every class λ of organs can then be described by a Gaussian multivariate model:

$$f_\lambda\big(\vec{Y}_{LE}\big|\vec{\mu}_{LE},\Sigma_{LE}\big) = \frac{1}{2\pi^{N/2}|\Sigma_\lambda|^{1/2}} \exp\left(-\frac{1}{2}(\vec{Y}_{LE} - \vec{\mu}_{LE})^t \Sigma_\lambda^{-1}(\vec{Y}_{LE} - \vec{\mu}_{LE})\right). \tag{10}$$

where Σ_k and $\vec{\mu}_k$ are respectively the covariance and the mean of LEESTs, estimated from a group of labelled images.

The chosen class λ is determined by the maximum value obtained with the MAP estimator among all *k*-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 belong to by computing the maximum likelihood that is obtained for the different possible models.

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:

$$f\left(\vec{Y}_{LE}\big|(\omega_k,\vec{\mu}_k,\Sigma_k)\right) = \sum_{k=1}^{K}(\omega_k \cdot p_k(\vec{Y}_{LE}|\vec{\mu}_k\Sigma_k)), \qquad (12)$$

the parameters $\omega_k > 0$ are the weights of sum equal to 1. The mixture model parameters $\omega_k, \mu_k, and\ \Sigma_k$ are estimated by employing the expectation-maximization (EM) algorithm (Titterington *et al.,* 1985).

Learning phase, extraction of study samples and ground-truth for models estimation

Ground-truth labelling is a process where a set of *n* sites $S_n = \{s_1(x_1,y_1),\ldots,s_n(x_n,y_n)\}$ is assigned a set of *n* labels $L = \{\lambda_1, \lambda_2,\ldots,\lambda_n\}$, the serial numbers for *K* classes, and assigned a set of *n* descriptors (LST), $D = \{\vec{Y}_{LE1},\ldots,\vec{Y}_{LEn}\}$. Eventually this process provides a database with samples whose position, class and properties are known. As there is no existing coherent database of labelled vine grapevines images, authors manually labelled images acquired for the study. LST's were then computed with a specially designed algorithm. So that models can include the various morphologies and textures and describe most accurately their variability, points of interest have been sampled with respect of the proportions and disparities encountered. The labelled samples then provide material to learn models and also material to compare decisions resulting from the proposed models to the ground-truth.

Post-processing : spatial regularisation

Markovian method : ICM algorithm (iterated conditional modes)

The classification method employed is a probabilistic decision made independently for each pixel, without considering the decisions reached for its neighbors. 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 method can produce such erratic results. In order to enhance the efficiency/veracity of the segmentation, authors propose a spatial regularisation based on an ICM algorithm. 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 neighbors (8-connectivity cliques). ICM algorithm is a simple 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 smoother segmentations.

Mathematical morphology

This regularization process aims also at discarding sparse distributions, but for larger objects (connected components), *i.e.* misclassified groups of labels whose neighbors 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 morphological opening and closing operations (Serra, 1982).

Analytical protocol

Different methods are proposed for each step of the processing chain. The purpose of this analysis is to compare different combinations of these methods in terms of classification

performances. RGB, LST and *LEEST* 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: flowerhood 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 10^5 sample sites manually selected per image per class are used for the estimation of models. There is then one model per class for each stage that is 12 probability density functions to be estimated. Performances are evaluated with a *leave-one-out cross-validation* process where, for each batch, performance metrics are computed from confusion matrices. Two metrics are used to describe performances for each class, Precision and Recall. Precision represents the fraction of relevant instances among the retrieved instances. Recall represents the fraction of relevant instances that have been retrieved over the total amount of relevant instances.

Results

Optimal scale for the extraction of structural properties

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

For example, at small scales, it is the *granular* appearance of leaves with sparse veinlets that is described, when at larger scales the structure tensors describe a smoother pattern. Similarly, small scales describe textural properties within a berry or a flower when larger scales describe a more entropic pattern of fruits, stalks and peduncle.

It is then not so obvious to determine the scale at which textural properties are robust to local noises, describe and discriminate classes to obtain the best classification performances.

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).

Fig.4 presents the relative performances obtained for varying values of scale parameter σ_t given a fixed value of gradient ($\sigma_g = 3.5$). Results are presented only for the stage berries touching (BBCH 79), they are obtained with the *LEEST* representation for mvG without any regularisation.
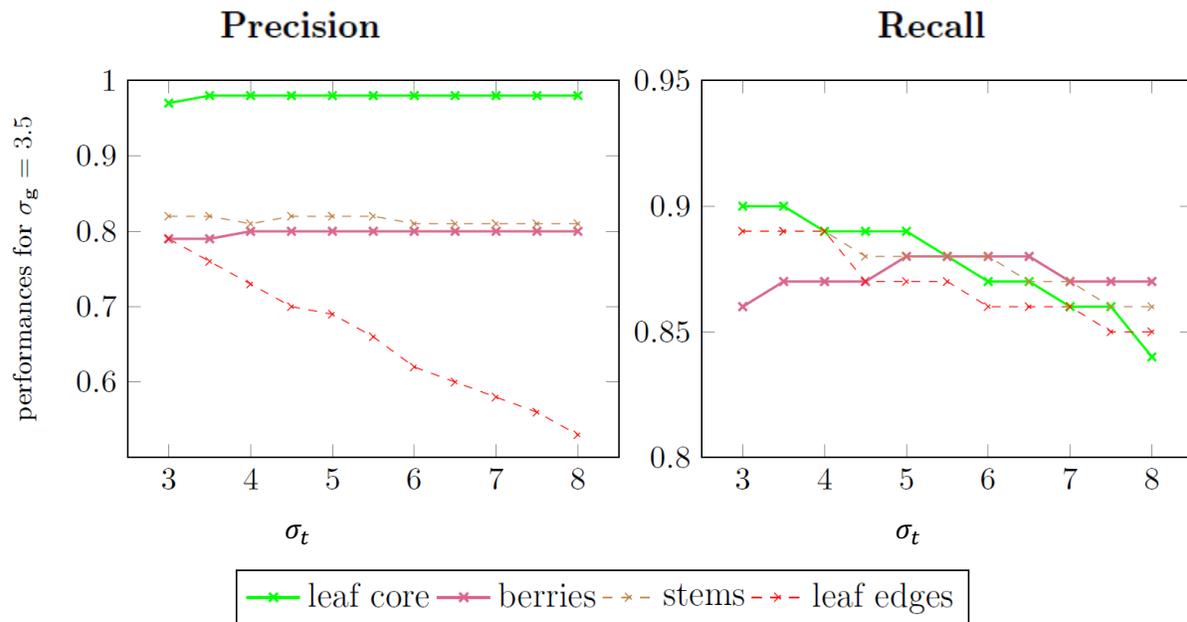

Figure 3 - Influence of σ_t scale for optimal value of σ_g on precision and recall metrics for stage BBCH 79

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 and σ_t. The chosen gradient value ($\sigma_g = 3.5$) results in precisions reaching up to 80% for berries, which is also the scale that maximize 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 to then be determined to ensure maximizes altogether recall rates for both primary classes. Such a compromise can be found at the intersections of the berry curves and the leaf core curves for $\sigma_t \in [5.0\,;5.5]$.

A similar analysis was conducted to determine the optimal value of σ_g given fixed values of σ_t. Eventually the optimal couple is found for the couple ($\sigma_g = 3.5\,;\sigma_t = 5.5$) that both maximize recall for berries (*88%*) while ensuring recall above *88%* too for leaf cores. The best optimal parameters are then the same for both precision and recall metrics. In the following, performance tests are conducted with the couple ($\sigma_g = 3.5\,;\sigma_t = 5.5$).

Comparison of features representations, decision and regularisation methods

Table 1 - comparison of features and of the different methods for LEEST representation at stage BBCH 79

Methods			Precision				Recall			
Representation	Decision	PP	leaf	berries	stems	edges	leaf	berries	stems	edges
RGB	mvG	∅	0.48	0.37	0.29	0.17	0.69	0.51	0.43	0.22
LST	mvG	∅	0.96	0.61	0.68	0.27	0.79	0.74	0.53	0.55
LEEST	mvG	∅	0.95	0.79	0.80	0.45	0.86	0.79	0.81	0.75
	mvG + sprob	∅	0.90	0.80	0.85	0.57	0.91	0.83	0.78	0.55
	mvG + sprob	ICM	0.95	0.82	0.82	0.71	0.86	0.84	0.80	0.82
	mvGM	∅	0.97	0.80	0.81	0.71	0.83	0.81	0.84	0.85
	mvGM + sprob	∅	0.93	0.83	0.87	0.77	0.90	0.96	0.81	0.81
	mvGM + sprob	ICM	0.93	0.87	0.89	0.79	0.92	0.90	0.81	0.82
	mvGM + sprob	ICM + Morph	0.95	0.91	0.81	0.86	0.92	0.90	0.85	0.79

Alone, colorimetric information is not sufficient to describe and discriminate the grapevine organs, it results in random classifications. The pure structural information provided by LST's better describes the textural properties encountered, however it is also insufficient to achieve a satisfying classification (recall rate <80%, precision <65% for berries). The colour extension of structure tensor is essential to capture the distinctive properties of the textures appearing on grapevine images. LEEST representations improve the classification performances with recall rates always above 80% for primary classes.

The proposed method for regularisation and for the management of a priori probabilities tend to improve selectively the performances for some classes while slightly decreasing performances for other classes. However the combinations of these propositions improve results in all classes. The use of multivariate Gaussian mixtures results in better performances for all classes. When combining all propositions with mvGM, performances metrics are above 90% for the primary classes.
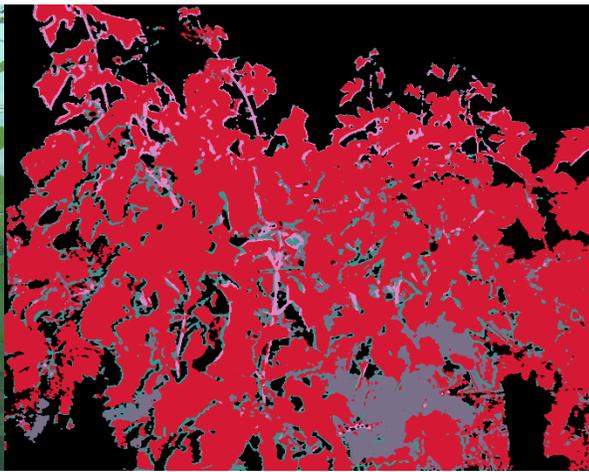
Table 2- comparison of the best performances achieved with LEEST mvGM and full regularisation for three phenological stages BBCH 68, 75 and 79

Metrics	Precision				Recall			
Phenological stages	leaf	berries	stems	edges	leaf	berries	stems	edges
Flowerhoods falling (BBCH 68)	0.90	0.86	0.89	0.84	0.92	0.87	0.83	0.79
Pea-sized berries (BBCH 75)	0.84	0.82	0.89	0.73	0.82	0.90	0.75	0.81
Berries touching (BBCH 79)	0.95	0.91	0.81	0.86	0.94	0.90	0.85	0.79

LEEST method produces satisfying performances for the 3 key phenological stages. However classification performances tend to be lower for the earlier phenological stages. It could 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* : $\pi_\lambda(h)$.
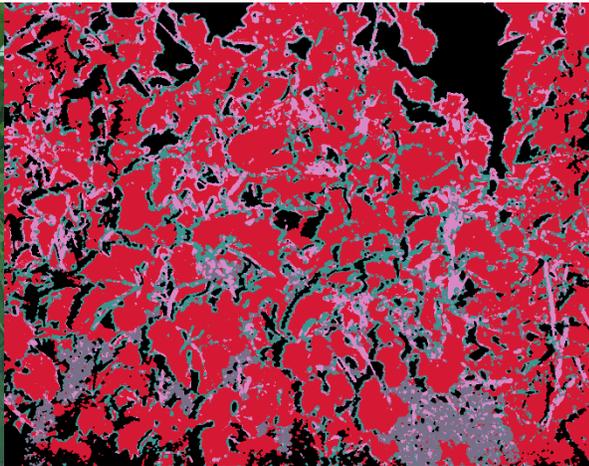
(a) – Original image BBCH 68   (b) – LEEST classification map BBCH 68
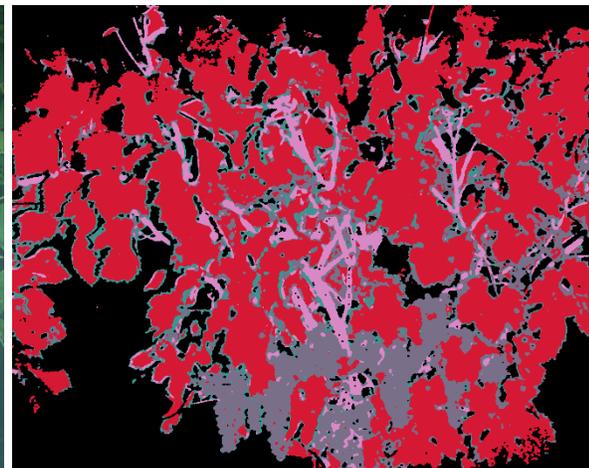
(c) – Original image BBCH 75   (d) – LEEST classification map BBCH 75

(a) – Original image BBCH 79   (b) – LEEST classification map BBCH 79

■ Leaf Cores   ■ Berries   ■ Stems   ■ Leaf edges

Figure 4- Examples of images and classification maps obtained with the *LEEST* and mvGM based method with full regularization (ICM + Mathematical Morphology) implemented for stages BBCH 68 (a-b), BBCH 75 (c-d) and BBCH 79 (e-f).

Discussion

The proposed framework was tested only for a limited number of pixel samples in a limited number of images. 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 of whole images without more abundant and thoroughly indexed database. Such a database could enable to compare the proposed method with reference classification methods such as SVM and Neural networks.

The robustness of the method and its models to varying varieties of grapevine or different cultivation systems are not tested yet either. 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 etc.) into formal agronomic parameters. This integral step toward 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-acknowledge methods.

Conclusion

In order to solve common problems regarding the classification of objects within natural outdoor images of plants in proximal sensing, a new framework has been proposed.

The proposed method is based on the parametric modelling of joint structure and colour. The purpose of this method is to overcome two major problems encountered in machine learning applications: availability of abundant and indexed data for learning processes and the determination of parameters for the settings / tunings of algorithms

To do so, several variants of the framework were proposed:

The vector representation *LEEST*, based on colour extended structure tensor mapped onto the log-Euclidean space was tested with Gaussian multivariate and Gaussian mixtures probabilistic models. In addition, a method of spatial management of a priori probabilities in the MAP estimator and two methods of spatial regularisation (ICM and mathematical morphology) were tested to improve the method.

The work contributed to develop a descriptor joining essential information of structure and colour extracted with *LEEST* representation based on the colour extension of local structure tensors priorly mapped into the log-Euclidean metric space. LEEST representation, in comparison with the usual LST approach, produces descriptors respecting the Riemannian geometric properties of structure tensors with additional correlated colorimetric information. These properties enable the estimation of consistent 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 which is based on modelling requires few manual settings or tuning from users. The only parameters to be determined by user are mainly intuitive scale parameters highly correlated to texture sized. Moreover the *LEEST* method is quite robust to these scale parameters. The method is easily applied to different phenological stages with satisfying results in each case.

The method 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. However the proposed framework could be easily transposed to crops with similar structures such as fruit trees

Acknowledgment

References

Arsigny, V., Fillard, P., Pennec, X., 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

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.

Bramley, R.G.V., Trought, M., Praat, J.P. (2011). Vineyard variability in Marlborough, New Zealand: Characterising variation in vineyard performance and options for the implementation of Precision Viticulture. Australian Journal of Grape and Wine Research. 17. 72-78. 10.1111/j.1755-0238.2010.00119.x.

De Luis-Garcia, R., Deriche, R., 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

Keresztes B., Da Costa J.-P, Grenier G, Germain C, David-Beaulieu X. (2014). A depth and color vision system for the automated detection of grape maturity, in: Equipment for Agriculture and Forestry. Presented at the RHEA, Madrid.\\

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

Leonard, G. (2011). History and Theoretical Basics of Hidden Markov Models, in: Dymarski, P. (Ed.), Hidden Markov Models, Theory and Applications. InTech. https://doi.org/10.5772/15205 \\

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

Li, P. and Wang, Q. (2012). Local Log-Euclidean covariance matrix (L2ECM) for image representation and its applications. Proc. 12th Eur. Conf. Comp. Vis., pp. 469-482.

Liu S., Marden S., Whitty M. (2013). Towards Automated Yield Estimation in Viticulture. Proceedings of Australasian Conference on Robotics and Automation, University of New South Wales, Australia

Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U., 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.

Rosu, R.G., Da Costa, J.P., 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., 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. (1982). Image analysis and Mathematical Morphology. ISBN 0-12-637210-3.

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

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.