Machine-learning assisted determination of best acquisition protocols in variety testing

Hadhami Garbouge[1], Salma Samiei[1,2], Pejman Rasti[1,2], David Rousseau[1]

[1] Université d'Angers, LARIS, UMR INRAe IRHS, Angers, France
[2] CERADE, ESAIP, Angers, France
david.rousseau@univ-angers.fr

Abstract

In this communication, we propose an attempt of pipeline to determine the best acquisition protocols automatically from ensemble of images acquired in different conditions. This is discussed to be specially useful in the field of variety testing where similar traits are measured in non standardized ways with help of images. Illustrations are given on distinct traits including detection of head wheat and sugar beet leaves. We discuss the methodological questions opened by this pilot proposal.

Introduction

As a consequence of climate change, there is an urgent need to develop new varieties capable of facing new climatic scenarios. However, the process of variety selection is rather long (10 years). To commercialize a new variety of an agricultural or vegetable species, a plant breeder has to follow a process managed by a national authority and delegated to an examination office (EO) that will describe and evaluate the variety for its registration on the national list. Evaluation results including variety descriptions may also serve for the granting of Plant Variety Rights (PVR). Currently a large majority of these tests are based on manual measurements performed from visual inspection. This method has consequences in terms of efficiency due to the time consuming nature of these tests. It is also an issue for the reproducibility of these tests when some characteristics are based on qualitative characteristics which may suffer from subjectivity in their assessment. Improving efficiency and reproducibility of these observations would be extremely useful for EOs that are continuously seeking for optimized testing methods implemented in testing protocols. It could also provide means to assess new characteristics developed in response to new agricultural constraints, particularly in the perspective of climate change. In addition, more efficient measurement methods would assist in addressing the challenge of the constant increase in the number of varieties that have to be tested. The described challenges encourage to head toward the use of sensors and numerical practices to progressively replace classical manual methods of examination whenever there is a need to speed up measurement or increase their reproducibility and objectiveness (Garbouge et al. 2020). The trend of using more and more imaging for plant science has started some decades ago and has been extensively reviewed (Li, Zhang, and Huang 2014; Qiu et al. 2018) for most recent ones, including with cost-effective strategies (Reynolds et al. 2019). While imaging modalities used in plant science and variety testing may be similar, the types of measures in plant science and variety testing differ either by their nature and technical aspects. So far, few attention from the academic imaging community focus on these specific aspects of variety testing. This ongoing numerical transition is currently encouraged at the European level via collaborative networking projects (including https://www.h2020-invite.eu/).

There are several challenges to address in order to reach common numerical practices in variety testing. One of them lays right at the level of image acquisition. How to define optimal protocols of acquisition which would be shared and strictly followed by several countries? A Top-down approach would consists in letting engineers propose a strict protocol including the brand and set up of a camera, lighting mode, vector on which to fix the camera, position of the imaging setup toward the targeted crops, ... Such a rigid approach would by sure normalize the practices, but would run the risk to face non-compliant behaviors among the local experts in charge of image acquisition since it may not systematically be applicable due to local environmental constraints not envisioned before-hand. Another bottom-up approach would consists in letting the local experts of all interested nations discuss before-hand with engineers to define a common protocol. A risk here is to have a low convergence of these discussions. We believe that another option is possible to help this process of selection of best acquisition protocol.

We propose in this communication to consider the situation where existing datasets gathered in several places for the same purposes are fed to an algorithm capable of identifying automatically the best images for a final measurement. This methodology is illustrated on three datasets. We finally discuss the perspectives opened by this first pilot trial which could be extended and enriched in many ways.

Method

We assume a dataset constituted of raw images is acquired with various acquisition protocols for the same purpose and the associated ground truth (binary masks for segmentation for instance (e.g. binary masks for segmentation)) exists. We

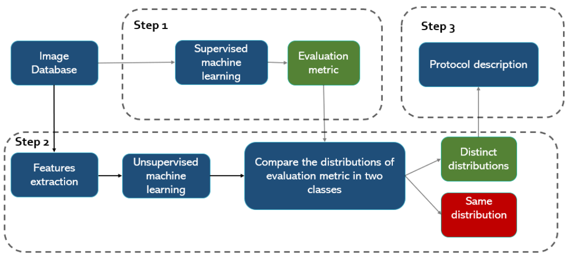

Figure 1: Proposed generic pipeline proposed to select best acquisition protocol in variety testing.

propose the following method to automatically detect the best imaging conditions for acquisition protocols inside this dataset (See Fig. 1). At the first step, we split the dataset to the train and the test. These datasets are composed of balanced (uniform) images from the different acquisition protocols. In the second step, handcrafted features corresponding to the expected optical quality of the acquired images are computed and a clustering method is applied on these features. The clustering method includes two classes for the expected good and bad quality of images. A statistical test is then made to decide if the distribution of the quality metric inside each cluster can be considered distinct or not. Finally in step 3, based on the results of the statistical test, a recommendation setting of the optical parameters are generated for the users.

We did not identified clear most related work from the computer vision community on this problem. Ideally, we would like to come up with a caption associated to an image were the expected quality of the image would be directly indicated to the technician in the field if acquisition parameters (focal, focus, angle, light, ...) are not in agreement with the reference dataset.

Dataset

We implemented the generic pipeline of Fig. 1 and tested it on three datasets shown in Fig. 2. The first dataset includes 213 images (150 in training and 63 in test) of a sugar beets acquired under various illuminations including overexposed (i.e. where the sensor is saturating) conditions. The purpose of this dataset is the segmentation of the leaves from the soil. The percentage of coverage of the soil at a given date is an important trait in variety testing. The second dataset includes 190 images(160 in training and 30 in test) of wheat observed for side view. The task is the segmentation of the spikes from the first raw of the micro-parcel. The last dataset is taken from the global wheat data challenge (David et al. 2021) with a subset of 3422 images (2758 images in training and 664 in test) of wheat ears observed from top view in the field. The task is to segment the ears. Here again the angle of view may vary from top view (90 degrees) to 45 degrees from top view. Images from these three datasets have been manually annotated to produce binary masks of the objects to be segmented.

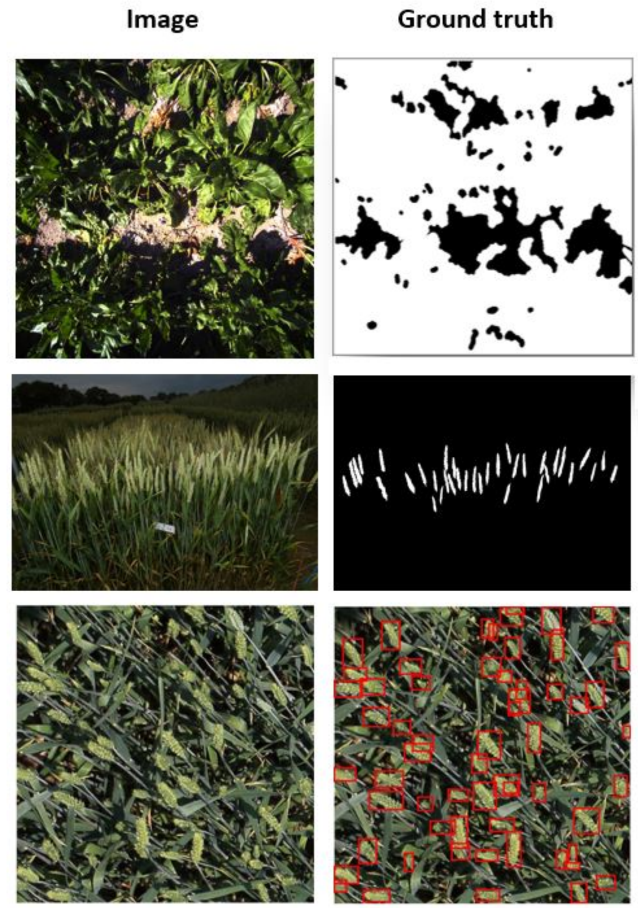

Figure 2: Datasets and ground truth used to test the pipeline of Fig. 1. Top raw: sugar beets observed from top view with various illuminations; Middle raw: wheat observed from the side view with various angles of the cameras; Bottom raw: hear observed from top view with various angles of the various angles of the cameras .

Algorithms

We now provide more details about the specific algorithms used in the generic pipeline of Fig. 1. The three considered datasets being dedicated to segmentation, we used a standard U-Net neural network architecture (Ronneberger, Fischer, and Brox 2015) for the image processing algorithm of step 1. The evaluation metric was chosen as the Sørensen-Dice coefficient D of the segmentation

$$D = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

where X is the predicted segmentation and Y the ground truth.

The features extracted were selected to test the impact of variations of acquisition conditions on the final result. The sugar beet dataset were acquired under various spatial illumination including risks of image saturation and low exposure. We proposed for this dataset to simply count the percentage of pixels having low values, arbitrarily chosen from

0-30 after RGB to gray conversion, and the pixels close to saturation level, arbitrarily chosen from 227-255. An image with correct exposition is expected to have low percentage of pixels in these saturation part of its input-output characteristic. Wheat from side view were acquired under various angles of the camera toward the ground. To probe this optical parameter, we included an estimation of the depth from RGB monocular view (arbitrarily chosen from (Liu et al. 2015) among many deep learning variants from the literature) and simply computed the standard deviation of the estimated depth map. An image with low standard deviation in this depth map is expected to be acquired with an angle of 90 degree from the main vertical axis of the wheat heads. Last, to also probe the angle of view, the percentage of vegetation was computed from a standard semantic segmentation such as the one used in (Samiei et al. 2018). A high percentage of vegetation indicates a side or top view with low part due to the sky or additional non plant items (humans, tractors, ...). These three simple features were applied on each images to feed the clustering method.

Image quality control by binary clustering (K-means with K=2) is applied to test the hypothesis of the quality of images based on the defined acquisition protocol. All features were normalized to 1 to avoid distortion effects when using Euclidean distance in the K-Means algorithm. A Wilcoxon rank-sum test (Kanji 2006) was applied on the distribution of the Dice coefficient inside each cluster. The null-hypothesis was chosen as the equality of the medians. This null hypothesis is validated at the default 5% on the P-Value. A recommendation of specific care about the tested optical parameter is finally recorded based on the result of this test.

Results

The distribution of the Dice coefficient in each cluster produced by the K-Means algorithm are displayed in Fig. 3 for the three tested data sets. The P-value indicates in all these cases that hypothesis H_0 can be rejected. This indicates that the optical parameters tested (Illumination for dataset 1 and 3, Orientation for dataset 2) have an impact on the quality of the segmentation performance. Interestingly, when gazing at the image in each cluster (see Fig. 4) the clustering indeed corresponds to uniform optical conditions, i.e. saturated or well exposed images in dataset 1 and 3 and uniform angle of view is dataset 2. On could use the result of such an experiment to identify the most important optical parameters and define in a data driven way the best practices. Here the experiment indicates to avoid saturation and favor side view or 45 degree view rather than top view. One can also notice that the distribution of the Dice coefficients are overlapping in the three conditions. This means that despite a statistically grounded difference in the performance in each cluster the difference is limited and could probably be reduced again by extending significantly the size of the training data sets with optical parameters in the range of what was included in the first. With both analysis our pipeline of Fig. 1 provides fruitful feedback and strategy to define the best acquisition protocol depending on the size of the dataset and the associated effort of image annotation.

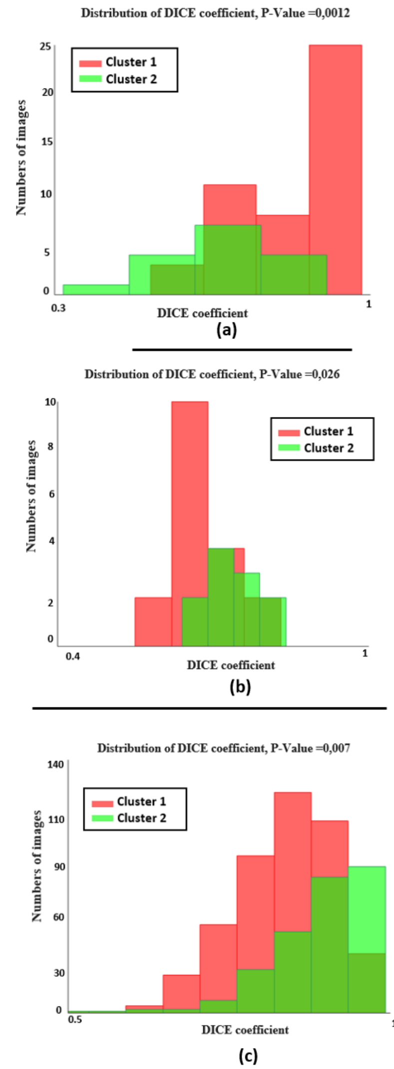

Figure 3: Distribution of Dice coefficient in each cluster for the three datasets processed in the study.

Conclusion and perspective

In this communication, we have introduced the problem of normalization of acquisition protocol in variety testing. We believe that machine learning can help to define the best protocol in a reverse engineering mode. In this pilot study, first we proposed a supervised approach where handcrafted features correlated to optical parameters were used to cluster images. The approach was then successfully illustrated on datasets dedicated to segmentation tasks.

The work could be extended in many ways. While the problem appears to us original and challenging for computer vision some clear limitations can be underlined on the way we tackled it so far. Because we have chosen a supervised approach, we have to deliver a similar amount of data for all the tested variants of the protocol. This may seem problematic since we especially do not completely specify the

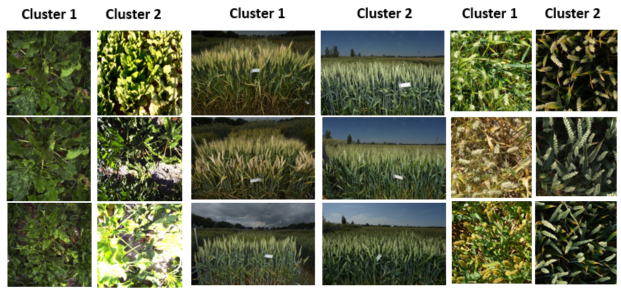

Figure 4: Instances of each cluster in each of three datasets processed in this study.

protocol itself but rather propose to dive into the dataset to select the best practices. Also, annotation of the images has to be done on the whole dataset while we suspect that some of these data has insufficient quality. This appears as a loss of time. We can expect that expert that will do the annotation, will, by common sense, be able to identify the quality of the images by themselves and may not in the end have to wait for the answer of our algorithm to sort out the good from the bad quality images. One could envision heading toward a fully unsupervised and end-to-end approach. Variational auto-encoders (VAE) (Girin et al. 2020) could be used to produce a latent space where the clustering would operate. A possible limitation is that this latent space would still depend on the composition of the initial dataset. What would happen if among all the protocols, the best one was represented with few images only. This last remark rely on the fact that in the implementation presented in this communication the datasets were limited. A direction would be to bet on unsupervised algorithms trained on huge dataset purposely acquired in diverse conditions in order to ensure from the data rather than from the protocol itself sufficient robustness.

Another direction would be to investigate the possible use of synthetic plants positioned in virtual environment such as the one used for video gaming conception. There are models of virtual plants for almost all crops of interest and the libraries are continuously growing. The production of these models benefit from extensive use of L-System grammars (Room, Hanan, and Prusinkiewicz 1996; Mishra and Mishra 2007; Boudon et al. 2012) to simply but very realistically produce in-silico plant models. Optical parameters such as lighting, angle, optics, depth of field, exposure, resolution of the cameras can automatically be simulated in virtual environment. Annotation of the plants themselves can also be automated since the ground truth is created by the computer directly. The selection of the optimal acquisition protocols would in this case be more direct since the optical parameters would directly be known and not only correlated with handcrafted features. Our group has expertise in this field of digital twin (Douarre et al. 2019; Sapoukhina et al. 2019) and we are working in this direction to overcome some of the mentioned limitations of our proposed approach.

References

Boudon, F.; Pradal, C.; Cokelaer, T.; Prusinkiewicz, P.; and Godin, C. 2012. L-Py: an L-system simulation framework for modeling plant architecture development based on a dynamic language. *Frontiers in plant science*, 3: 76.

David, E.; Ogidi, F.; Guo, W.; Baret, F.; and Stavness, I. 2021. Global Wheat Challenge 2020: Analysis of the competition design and winning models. *arXiv preprint arXiv:2105.06182*.

Douarre, C.; Crispim-Junior, C. F.; Gelibert, A.; Tougne, L.; and Rousseau, D. 2019. Novel data augmentation strategies to boost supervised segmentation of plant disease. *Computers and electronics in agriculture*, 165: 104967.

Garbouge, H.; Nicolas, M.; Peller, J.; Vermeulen, P.; Demilly, D.; Van de Zedde, R.; Collonnier, C.; Mueller-Linow, M.; and Rousseau, D. 2020. Toward Numerical Practices in Variety Testing: a Rationale to Select the Most Promising Traits. *Report of the annual meeting of UPOV*.

Girin, L.; Leglaive, S.; Bie, X.; Diard, J.; Hueber, T.; and Alameda-Pineda, X. 2020. Dynamical variational autoencoders: A comprehensive review. *arXiv preprint arXiv:2008.12595*.

Kanji, G. K. 2006. *100 statistical tests*. Sage.

Li, L.; Zhang, Q.; and Huang, D. 2014. A review of imaging techniques for plant phenotyping. *Sensors*, 14(11): 20078–20111.

Liu, F.; Shen, C.; Lin, G.; and Reid, I. 2015. Learning depth from single monocular images using deep convolutional neural fields. *IEEE transactions on pattern analysis and machine intelligence*, 38(10): 2024–2039.

Mishra, J.; and Mishra, S. 2007. *L-system Fractals*. Elsevier.

Qiu, R.; Wei, S.; Zhang, M.; Li, H.; Sun, H.; Liu, G.; and Li, M. 2018. Sensors for measuring plant phenotyping: A review. *International Journal of Agricultural and Biological Engineering*, 11(2): 1–17.

Reynolds, D.; Baret, F.; Welcker, C.; Bostrom, A.; Ball, J.; Cellini, F.; Lorence, A.; Chawade, A.; Khafif, M.; Noshita, K.; et al. 2019. What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Science*, 282: 14–22.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Room, P.; Hanan, J.; and Prusinkiewicz, P. 1996. Virtual plants: new perspectives for ecologists, pathologists and agricultural scientists. *Trends in Plant Science*, 1(1): 33–38.

Samiei, S.; Rasti, P.; Daniel, H.; Belin, E.; Richard, P.; and Rousseau, D. 2018. Toward a Computer Vision Perspective on the Visual Impact of Vegetation in Symmetries of Urban Environments. *Symmetry*, 10(12): 666.

Sapoukhina, N.; Samiei, S.; Rasti, P.; and Rousseau, D. 2019. Data augmentation from RGB to chlorophyll fluorescence imaging Application to leaf segmentation of Arabidopsis thaliana from top view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 0–0.