Machine learning meets distinctness in variety testing

Geoffroy Couasnet
IRHS-UMR 1345, INRAe, France
geoffroy.couasnet@inrae.fr

Mouad Zine el Abidine
Université d'Angers, France
mouad.zineelabidine@etud.univ-angers.fr

François Laurens
IRHS-UMR 1345, INRAe, France
francois.laurens@inrae.fr

Helin Dutagaci
Eskisehir Osmangazi University, Turkey
hdutagaci@ogu.edu.tr

David Rousseau
Université d'Angers, France
david.rousseau@univ-angers.fr

Abstract

Distinctness is a binary trait used in variety testing to determine if a new plant variety can be considered distinct or not from a set of already existing varieties. Currently distinctness is mostly based on human visual perception. This communication considers distinctness with a machine learning perspective where distinctness is evaluated through an identification process based on information extraction from machine vision. Illustrations are provided on apple variety testing to perform distinctness based on color. An automated pipeline of image acquisition, processing and supervised learning is proposed. A feature space based on the 3D color histogram of a set of apples is built. This feature space is built using optimal transport, fractal dimension, mutual entropy and fractional anisotropy and it provides results in accordance with human expertise when applied to a set of varieties highly contrasted in color and another one with low color contrast. These results open new research directions for achieving higher-throughput, higher reproducibility and higher statistical confidence in variety testing.

1. Introduction

o commercialize a new variety of an agricultural species a plant breeder has to follow a very strict and framed process managed by a international authorities and delegated to examination offices (EO) that describe and evaluates the variety for its registration on an official list. Currently a large majority of these tests are based on manual measurements performed by experts through visual inspection. This method is inefficient due to its time-consuming nature. There is also the issue of the reproducibility of these tests when they involve subjective assessment of qualitative characteristics. Improving efficiency and reproducibility of these tests would be extremely useful for EOs that are continuously seeking for optimized and objective methods implemented in testing protocols. It could also provide means to assess new varieties 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. Reproducible assessment tools would also contribute to harmonizing practices between countries.

So far few attention from the academic imaging community has focus on these specific aspects of variety testing and this communication can be considered as a contribution in this direction. Several types of evaluations are performed by experts for variety testing including distinctness, homogeneity and stability. We focus here on distinctness. We developed a technique to determine if a representative set of a plant variety is sufficiently distinct from a catalogue of existing varieties so as to be registered as a new variety. We formulate this task as an identification problem to explore the potential of a supervised machine learning approach in distinguishing ensembles of plant samples.

For illustration we consider a specific use-case with apple variety testing based on their distinctness in terms of color (one of the most important characteristic for apple consumer). Machine learning approaches have been extensively used for fruit classification, e.g. separating a species from another, or a fruit of a given quality from others of lower quality. However, variety testing requires a more fine-grained approach and necessitates the exploration of the

distinct visual characteristics among varieties of the same species and the same consumer quality.

Color difference evaluation has been extensively studied by CIE which provides a set of formulae found in accordance with human perception for more or less simple uniform images with controlled background and illumination [12]. In our case we face the question of differentiation of a large set of textured images of apples which characterizes a variety from another large set of images. Comparing pairs of individual images of apples with the standard CIE approach would be a rather brute force approach which is not followed by the experts in charge of variety examinations. We rather decided to mimic their current visual approach of these experts which is to compare two images each containing a large set of sample apples and decide whether the two sets belong to the same variety or not. We thus address the problem as a statistical one and consider distinctness as equivalent to deciding if a set of 3D color histograms representing an apple variety is sufficiently "distant" from a set of other 3D color histograms corresponding to other sets of apples.

The paper is organized as follows. After positioning the most related works of the literature, materials and methods section presents the image acquisition system and protocol together with the set of features chosen to characterize the distinction between 3D color histograms in specially designed experiments. Results obtained on two datasets of apples are presented and discussed before final conclusion and perspectives.

2. Related works

3D color histograms of high resolution images are dense point clouds and are difficult to be properly visualized for assessment of the density variations [5]. As a consequence, 3D color histograms have been characterized in many ways in the literature [19]: histogram intersection, dominant color descriptor, color correlogram, color co-occurrence matrix, dominant color descriptor, chromaticity, fractal dimension [7, 6], fractional anisotropy [1]. Color histogram matching can also be performed with the help of metrics developed in information theory such as the Shannon mutual information and their variants [20, 4] or optimal transport [16, 17] to probe the discripency between two statistical distributions. For a first step of variety testing in the domain of machine learning we decided to select a small subset of these features since we do not claim to provide an exhaustive analysis of the appropriate features. Here, we rather aim at setting the machine learning scheme and give a proof of feasability in the automation of distinctness.

From the computer vision generic perspective, the applied task considered in this communication may relate to identification problems such as image retrieval or object tracking [9, 10, 11, 17]. In our case again the identification problem is related to a population of images while in image retrieval or object tracking [19, 11] the task mostly targets identifying single image or objects. Fruit characterization is a field of machine vision in itself which has received considerable attention either for species identification or quality control (see [22, 14, 15, 18] for reviews). Again the statistical situation considered here is different since the goal is not to sort each individual apple but to determine if a set of apples can be considered as distinct or not from others. Most related works to what we propose can be seen as [13, 23] where supervised machine learning are proposed to classify existing varieties. In our case, while considering other crops and testing other features, we deal with the situation where a new set of plant has to be identified as similar to existing ones or distinct up to the point where it deserves to be designated as a new variety.

3. Materials and methods

3.1. Images Acquisition

The acquisition of the images of the different varieties of apples was carried out with the help of a conveyer machine allowing to move the fruits in translation while carrying out a rotation (see Fig. 1). A camera located at the top of the conveyor belt of the machine with a perpendicular viewing direction, took pictures of the apples in rotation, which allowed us to have multiple images providing almost an entire coverage of each apple. Approximately 9 to 10 views of the same apple were captured thanks to this rotation-translation process. These multiple views are important since apple may have several major colors on their skins. With the standard visual approach experts have to manually rotate the apples to have a full perception of the variation of color on a single apple. Here the machine presented in Fig. 1 can acquire a set of 30 apples in a couple of minutes. Images were acquired in burst mode with a Canon camera (10.1 megapixels resolution) controlled by a simple Raspberry-pi minicomputer. Apples were segmented automatically from background as visible in Fig. 1 and assembled in multiple view images of 30 apples as shown in Figs. 2 and 3. This machine, developed for this study, is much simpler and lower cost (approximately 10 keuros versus 100 keuros for classical apple sorting machines) than any commercial systems since it does not need to incorporate any sorting mechanism. Also, by contrast with most commercial system, access to raw image format, i.e. uncompressed format, is possible.

3.2. Datasets

Currently, when experts of EO are performing distinctness, they observe directly with their own eyes boxes of 30 apples of each tested variety and reference varieties manually positioned in a same room and they decide form a pure

Figure 1. Acquisition system. Upper panel: Machine equipped with a conveyor belt, used for the acquisition of images of apples with a high surface coverage. Lower panel: view of the acquired images of apples after segmentation from the background.

subjective perception if these sets are distinct or not from one another. An objective of this work is to produce a step toward an automation of such examination through the use of computer vision applied on images such as Figs. 2 and 3 which are automatically produced after acquisition on the system of Fig. 1. Two datasets were produced for this study to test the proposed machine vision approach for distinctness evaluation.

3.2.1 Non-Gala Mutant varieties

We first created a dataset of images of apples with highly distinct color distributions. The dataset is composed of 1293 images of apples belonging to 8 varieties (see Fig. 2) which we refer to as non-Gala Mutants. These varieties correspond to varietes identified as distinct from each other by the official examining offices. These varieties are not named, they simply have a reference number to identify them. The number of images per variety is given in Table 1.

Variety	Images
variety30	113
variety37	127
variety40	133
variety41	99
variety42	106
variety44	312
variety46	215
variety47	188

Table 1. Number of images of 8 reference, registered varieties corresponding to Non-Gala Mutant dataset.

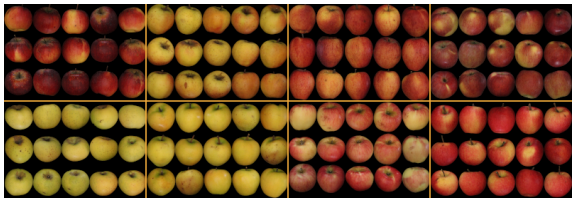

Figure 2. Images representing the 8 non-mutant Gala varieties in the order of appearance in the list of Table 1.

3.2.2 Gala mutants

As a complement to the first dataset, we built a second dataset containing 4040 images of apples belonging to 9 different mutants of the variety Gala. These mutants are similar to each others in terms of color content as shown in Fig. 3. The details of these mutants are given in Table 2 where first column gives the encrypted reference of each corresponding mutant. These mutants are also considered as distinct from each other by experts of EO but they somehow reach the limit of what they consider as distinct.

References	Images
X4111	597
X4410	438
X4712	716
X6716	684
X7440	444
X7812	259
X8125	329
X8594	343
X9214	230

Table 2. Number of images of 9 reference, registered varieties corresponding to Gala Mutant dataset.

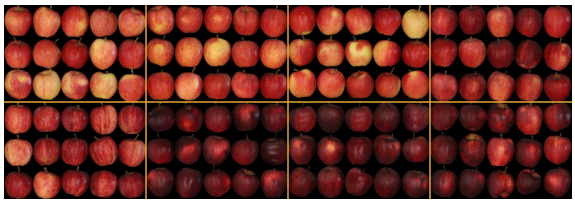

Figure 3. Images showing a subset of each Gala Mutant, in the same order of appearance as in Table 2, from left to right and by line, except for the mutant *X8594* which is completely yellow and highly distinctive.

3.3. 3D RGB Histograms

With our objective being to differentiate apples mainly based on the color distribution, we extracted features from the RGB histogram of the images represented in 3 dimensions (one axis by component color). We calculated the

RGB histogram of each image, and to obtain the RGB histogram of a variety, we simply calculated the sum of the RGB histograms of the images belonging to the variety. We can see the corresponding summed histogram of each of the non-Gala mutants in Fig. 4 and of the Gala mutants (except X8594) in Fig. 5. It is interesting to see that despite the loss of spatial localisation in RGB histograms a contrast between colors is clearly visible in this representation with the non-Gala mutants in Fig. 4. However, the contrast is much more difficult to perceive with RGB histograms in the case of the Gala mutants which represents a clearly challenging classification task.

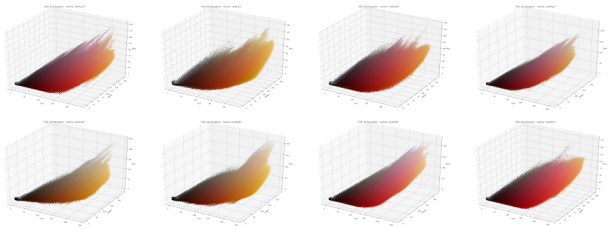

Figure 4. Images representing histograms of non-mutant Gala varieties. From left to right, by row: variety30, variety37, variety40, variety41, variety42, variety44, variety46 and variety47.

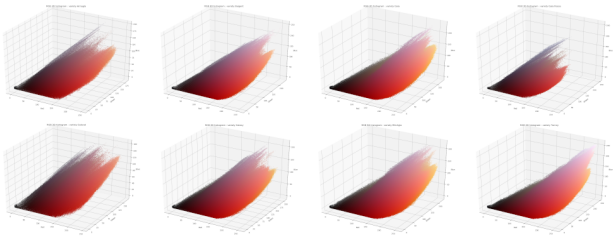

Figure 5. Images representing histograms of mutant Gala varieties. From left to right, by row: X4111, X4410, X4712, X6716, X7440, X7812, X8125, X9214.

3.4. Color features

Once the histogram of each image and each variety is built, we extract features allowing to characterize several aspects of the histograms. Same features were used for both datasets. We present the features used for this study in the rest of this section.

Average and variance of colors: The first two descriptors are the mean and the variance of the colors. It seems quite intuitive to use them since they give us respectively the average color of the variety of apples and the contrast of yellow and red regions are captured via the variance. These two values are easily calculated. Let I be any image of size $m \times n$ represented in the RGB space by an array of RGB vectors $P = (p_{i,j})$ where $p_{i,j} = (r_{i,j}, g_{i,j}, b_{i,j}) \in [\![0, 255]\!]^3$ is the color of the pixel of coordinates $(i, j) \in [\![0, m-1]\!] \times [\![0, n-1]\!]$ of image I. Denoting the mean and variance of the colors respectively by $\bar{P}$ and V the mean and the variance of the image I, we have:

$$\bar{P} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} c_{i,j}, \qquad V = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} (c_{i,j} - \bar{c})^2 \tag{1}$$

Fractional Anisotropy: Fractional anisotropy is a number in the interval $[0, 1]$ which reflects the degree of anisotropy of the shape of the point cloud formed by the 3-dimensional histogram. This scalar gives a measure of the stretch of the point cloud in various directions. If its value is 1, it means that the points would all be distributed along a perfectly linear axis. If its value is zero, it means that the points are distributed homogeneously in all directions. Thus, a sphere has a zero fractional anisotropy, an ellipse has a fractional anisotropy between 0 and 1 and a straight line has a fractional anisotropy equal to 1. After obtaining the eigenvalues λ_1, λ_2 and λ_3 of the PCA (Principle Components Analysis) on the 3-dimensional histogram of an image, we can easily calculate the fractional anisotropy, denoted as FA, via the formula

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}} \tag{2}$$

Fractal box counting dimension: The fractal box counting method [2] subdividing the 3D color cube $[\![0, 255]\!]^3$ into '*box*' of the edge length r counts the number of boxes $N(r)$ necessary to cover each color cell occupied by the point cloud making up the 3D histogram. This number of boxes $N(r)$ has been found to follow a law of the form r^{-D} where D is the fractal dimension of the histogram [6, 7]. This number ranging between 2 and 3 for natural images provides a description of the structure and density of the point cloud. A smaller value of fractal dimension indicates that although the histogram is distributed throughout the color space, there remain empty regions.

Mutual entropy: Mutual entropy [8] allows us to compute the information common to 2 histograms. We include this mutual entropy as the measure of the color similarity between an image and the target variety. To calculate the mutual entropy, also called joint entropy, between two histograms, we use the following formula

$$ME = -\sum q(x) \log(\frac{q(x)}{p(x)}) \tag{3}$$

where p and q represent the respective pixel distributions represented in the histograms of the two images.

Cost of optimal transport: As last feature we propose to include optimal transport [21] which provides a way of transporting a set of points to another set in the least expensive way possible. In our case minimizing the total distance between the two sets of points fits with the capability of optimal transport. Since we work on 3-dimensional histograms of our images, we can measure the cost in terms of the distance between the histogram of an image and the average histogram of a target variety. If we assume we have k varieties (here $k = 8$ or 9 depending on the dataset used), we get k values representing the cost of moving our image to the k varieties. These k values are treated as color features each representing a measure of the probability of the image to belong to the corresponding variety. The lower the cost, the closer the histograms are in terms of structure. Let $\mu = \sum_{i=0}^{m-1} p_i \delta_{a_i}$ and $\nu = \sum_{j=0}^{n-1} q_j \delta_{b_j}$ be two discrete measures associated with two histograms, and let c be a cost function for which we note $c_{ij} = c(a_i, b_j)$ for all $i, j \in [\![0, m-1]\!] \times [\![0, n-1]\!]$. We then try to minimize $\iint c(x, y)\, dx\, dy$. By noting $b = \{p_0, p_1, \ldots, p_{m-1}, q_0, q_1, \ldots, q_{n-1}\}^T$ and $c = \{c_{11}, c_{12}, \ldots, c_{1(n-1)}, \ldots, c_{m1}, c_{m2}, \ldots, c_{(m-1)(n-1)}\}^T$, the problem then becomes a minimization problem of $c^T x$ under the constraints

$$\begin{cases} A.x = b \\ \forall i, j, x_{ij} \geq 0\ . \end{cases} \tag{4}$$

In practice, we calculated this cost using the python package named POT (Python Optimal Transport, https://pythonot.github.io/). The cost is computed using the method based on earth mover's distance, from 3D histograms. This algorithm has 2 advantages : histograms do not need to be normalized and they do not need to be of the same size [3]. All in all the feature space is composed of features of various dimensions. The optimal transport feature is a vector for which each component is the value of the norm of the cost between two varieties. Therefore if the dataset is composed of k varieties, the optimal transport is a vector with k components. The other features can be scalars as fractal dimension or fractional anisotropy, 3 components vectors as RGB means and RGB variance, or a vector of the same dimension as the optimal transport for mutual entropy.

3.5. Classification setups

In this part, we detail the machine learning classification setups tested to assess distinctness with both apple datasets presented in the previous section.

Multi-class classification A first setup is simply to perform a multi-class classification between the varieties allowing to assess if the varieties are distinguishable between them. This is a "one versus one" approach where the tested variety is tested against all the existing ones individually. For this, we simply separated an initial dataset of images to create 2 sub-datasets: a test sub-dataset containing $\frac{1}{3}$ of the images, and a training sub-dataset containing the rest. These 2 sub-sets were used respectively to train the supervised classifiers and to test their efficiency to distinguish the different varieties. For this classification, two sets of features were used, a set containing all features and the other set containing only the most relevant features among all the tested features.

Binary classification The second classification setup was used to test if our model was able to differentiate the two apple datasets. This is a "one versus all" approach where the one tested corresponds to the variety compared with all the existing registered varieties at once. We gathered the 2 datasets presented previously, thus constituting a dataset of 5333 images, with 4040 images of Gala mutants and 1293 non Gala mutants which are our 2 classes. We separated the dataset into test and training sub-sets with a 50-50 ratio. To mimic the procedure experts currently follows for apple variety testing, the algorithm made an individual prediction on each apples and a majority voting over subsets of 30 apples.

4. Classification results

4.1. Multi-class classification between Gala mutant varieties

We first performed the multi-class classification between Gala mutant apples only, in order to verify that it was indeed possible to distinguish these 9 registered varieties between them. We first separated the data into test and training sets with a ratio of $\frac{2}{3}$ for the training set, then we used 3 different supervised classifiers: Support vector machine (SVM), Random Forest and Linear Discriminant Analysis (LDA).

The classification results visible in Tables 3 and 4 show that the Gala mutant varieties are distinguishable in terms of color. These results are of the same order of magnitude for the three tested classifiers. However, SVM gives an F1-score over 97%, and perform slightly better than others.

A forward analysis, testing the performance of each individual type of features, identified that the best features for the classification happened to those from optimal transport. As visible in Tables 5 and 6, these features alone do not allow us to obtain a fully satisfactory classification, however they are relatively efficient since they yield a classification accuracy of about 50%. Since our dataset has 9 distinct classes, a random classification of the data would give 11%. accuracy. The relative superiority of optimal transport toward the other features can be explained since all histograms share the same elongated shape centered on

red-yellow barycenter.

SVM : All Features	
Precision Score	97,13 %
Recall Score	97,10 %
F1-Score	97,07 %
Parameters	
Kernel	linear
Penalty	l2
C	2.0
Loss	squared_hinge
Dual	False
Tol	0.0001
Multi_class	crammer_singer
Class_weight	None
Max_iter	-1
Data Transformation	Normalized

Table 3. Results obtained by SVM with Gala mutant dataset and all features.

LDA : All Features	
Precision Score	93,39 %
Recall Score	93,32 %
F1-Score	93,30 %
Parameters	
Solver	SVD
Shrinkage	None
Priors	None
n_components	1
store_covariance	False
tol	0.0001
Data Transformation	Normalized

Random Forest : All Features	
Precision Score	94,13 %
Recall Score	93,99 %
F1-Score	93,92 %
Parameters	
n_estimators	250
criterion	entropy
max_depth	None
max_leaf_nodes	None
bootstrap	False
Data Transformation	Normalized

Table 4. Results obtained by LDA and Random Forest with Gala mutant dataset and all features.

SVM : OT Features	
Precision Score	50,48 %
Recall Score	50,26 %
F1-Score	46,34 %
Parameters	
Kernel	linear
Penalty	l2
C	8.9
Loss	squared_hinge
Dual	True
Tol	0.0001
Multi_class	None
Class_weight	balanced
Max_iter	1,00E+06
Data Transformation	Normalized

Table 5. Results obtained by SVM with Gala mutant dataset and optimal transport only.

LDA : OT Features	
Precision Score	48,39 %
Recall Score	50,56 %
F1-Score	46,56 %
Parameters	
Solver	lsqr
Shrinkage	auto
Priors	None
n_components	1
store_covariance	False
tol	0.0001
Data Transformation	Normalized

Random Forest : OT Features	
Precision Score	52,35 %
Recall Score	53,38 %
F1-Score	52,49 %
Parameters	
n_estimators	40
criterion	entropy
max_depth	None
max_leaf_nodes	None
bootstrap	True
Data Transformation	Normalized

Table 6. Results obtained by LDA and Random Forest with Gala mutant dataset and optimal transport only.

4.2. Multi-class classification between non-Gala mutant varieties

In a second step, we performed the same classification method as in the previous section, this time using the dataset composed of the non-Gala mutant varieties.

For this dataset, the results obtained in Tables 7 and 8 are also very satisfactory, with F1-scores close to 90%. Once again, the SVM with polynomial kernel gives the best results with a precision score of 93.76%. For the Non-Gala mutants, these results show that the varieties composing this dataset are clearly distinguishable as also confirmed by the experts from EO since these are registered as official varieties. The precision score is logically found a bit lower than for the non-Gala mutant datasets since the contrast in color between varieties is lower.

Again we performed a forward analysis which established optimal transport as providing the most significant features. As visible in Tables 9 and 10, the classification only based on these optimal transport features reached their best results with SVM with polynomial kernel of degree 3, which gives a precision score of 71.25%. Globally the result observed with the well contrasted dataset non-Gala mutant are robustly conserved when the method is transposed to less contrasted apples such a the one of the Gala mutants. This demonstrates the high potential of a machine learning framework equipped with color features for variety testing.

SVM : All Features	
Precision Score	93,76 %
Recall Score	93,50 %
F1-Score	93,51 %
Parameters	
Kernel	poly
Degree	2
Penalty	l2
C	12
Gamma	auto
Coef0	0.5
shrinking	False
tol	0.0001
Class_weight	None
Max_iter	-1
Data Transformation	Normalized

Table 7. Results obtained by SVM with non Gala mutant dataset and all features.

LDA : All Features	
Precision Score	89.50%
Recall Score	89.10%
F1-Score	88.96%
Parameters	
Solver	lsqr
Shrinkage	auto
Priors	None
n_components	1
store_covariance	False
tol	0.0001
Data Transformation	Normalized

Random Forest : All Features	
Precision Score	89,35 %
Recall Score	89,10 %
F1-Score	88,81 %
Parameters	
n_estimators	90
criterion	entropy
max_depth	None
max_leaf_nodes	None
bootstrap	True
Data Transformation	Normalized

Table 8. Results obtained by LDA and Random Forest with non Gala mutant dataset and all features.

SVM : OT Features	
Precision Score	71,25 %
Recall Score	70,53 %
F1-Score	67,22 %
Parameters	
Kernel	poly
Degree	3
Penalty	l2
C	20
Gamma	auto
Coef0	0.5
shrinking	False
tol	0.0001
Class_weight	None
Max_iter	-1
Data Transformation	Normalized

Table 9. Results obtained by SVM with non Gala mutant dataset and optimal transport only.

LDA : OT Features	
Precision Score	52,96 %
Recall Score	61,72 %
F1-Score	55,20 %
Parameters	
Solver	svd
Shrinkage	None
Priors	None
n_components	1
store_covariance	False
tol	0.0001
Data Transformation	Normalized

Random Forest : OT Features	
Precision Score	57,78 %
Recall Score	59,40 %
F1-Score	58,12 %
Parameters	
n_estimators	40
criterion	gini
max_depth	None
max_leaf_nodes	None
bootstrap	True
Data Transformation	Normalized

Table 10. Results obtained by LDA and Random Forest with non Gala mutant dataset and optimal transport only.

4.3. Binary classification with the 2 collected datasets

Once we observed that both datasets were well distinguishable, we focus on the most difficult dataset and explore the potential of our framework to determine whether a set of test images corresponds to a certain Gala mutant or not. To mimic the way experts perform their scoring, we decided to focus not only on individual classification of apples but also on a group classification from the same variety. To do so, we selected images from the test data and by a random draw without replacement of apples of the same variety, to create subsets of 30 apples. This number exactly corresponds to the size of the group of apples chosen by the experts when they observe groups of apple for distinctness. Once our model is trained on classification of individual apple images, we tested its efficiency on the subsets through majority voting.

As can be seen in Tables 11, 12 and 13, we get 100% F1-score with all classifiers when all features are employed. Consistent with the results of the previous section optimal transport again appeared as the most important features in a forward analysis. With optimal transport only, Random Forest gives the best results in individual classification with a precision of 88.13% and an F-score of 75.67%.

5. Conclusion and perspectives

In this communication, we have considered, for the first time to the best of our knowledge, a variety testing problem

2*SVM – Individual classification – OT only		True labels		Scores	
		0	1	Precision	51,16 %
2*Predictions	0	1014	117	Recall	73,88 %
	1	316	331	F1-Score	60,46 %
2*SVM – Subsets classification – OT only		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	0	Recall	100,00 %
	1	0	14	F1-Score	100,00 %
2*SVM – Individual classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	1330	0	Recall	100,00 %
	1	0	448	F1-Score	100,00 %
2*SVM – Subsets classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	0	Recall	100,00 %
	1	0	14	F1-Score	100,00 %

Table 11. Results obtained by SVM on individual data and subsets of apples, with optimal transport only and all features.

2*LDA – Individual classification – OT only		True labels		Scores	
		0	1	Precision	82,26 %
2*Predictions	0	1286	244	Recall	45,54 %
	1	44	204	F1-Score	58,62 %
2*LDA – Subsets classification – OT only		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	10	Recall	28,57 %
	1	0	4	F1-Score	44,44 %
2*LDA – Individual classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	1330	0	Recall	100,00 %
	1	0	448	F1-Score	100,00 %
2*LDA – Subsets classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	0	Recall	100,00 %
	1	0	14	F1-Score	100,00 %

Table 12. Results obtained by LDA on individual data and subsets of apples, with optimal transport only and all features.

2*RF – Individual classification – OT only		True labels		Scores	
		0	1	Precision	88,13 %
2*Predictions	0	1291	151	Recall	66,29 %
	1	40	297	F1-Score	75,67 %
2*RF – Subsets classification – OT only		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	0	Recall	100,00 %
	1	0	14	F1-Score	100,00 %
2*RF – Individual classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	1330	0	Recall	100,00 %
	1	0	448	F1-Score	100,00 %
2*RF – Subsets classification – All features		True labels		Scores	
		0	1	Precision	100,00 %
2*Predictions	0	14	0	Recall	100,00 %
	1	0	14	F1-Score	100,00 %

Table 13. Results obtained by Random Forest on individual data and subsets of apples, with optimal transport only and all features.

with a machine learning approach. We have introduced on a use-case dedicated to apples a supervised learning scheme to identify if a new candidate for variety registration could be considered as distinct or not from an existing set of varieties. Two datasets corresponding to highly contrasted varieties and varieties contrasted at the limit of what would be considered as distinct have been tested. Distinctness was found in perfect accordance with the human expert. This demonstrates the possibility to introduce more objective and higher-throughput approaches in the domain of variety testing. We found that among the tested features optimal transport was producing the most adapted features, i.e. which contributed the most in the correct decision making. It is specially important to notice that all these results were obtained based on color histogram, i.e. with a total loss of

spatial information.

This first step opens various ways of further investigations. A limit of the result presented so far stands in the absence of negative data, i.e. non registered varieties in our dataset. Access and diffusion of such historical data is complex from a legal point of view when dealing with EO. A workaround approach could consist in simulating fake non registered varieties from an existing dataset. This requires to enlarge the datasets used in this article and we are currently investigating this direction. On the machine learning side, several alternatives could be considered. We selected classical shallow learning algorithms (SVM, random forest and LDA). We produced binary decisions in accordance with the essence of distinctness which is a binary trait. All the tested algorithms could also provide probabilities and confidence intervals which would provide more insights. Such output, although not currently in practice in variety testing would nonetheless be very useful specially to provide arguments to the breeders when new variety candidates are rejected by EO. The set of hand crafted features could be extended to additional color features mentioned in the related work section. Also, all the analysis were performed in the native RGB color space and other color spaces more suitable to fit with the human perception could also be tested with the approach introduced in this paper. Alternatively deep learning approaches could be considered. An obvious match would be with generative adversarial networks (GAN) [24] where the discriminator network could serve to decide if a variety is distinct from another after the GAN would have been trained to reproduce images of already registered varieties.

Varieties are registered based on a large set of parameters. In this communication we considered only color, but it would be interesting to extend the scheme to incorporate more parameters. On apple it could be color texture (stripes on apple skin) as well as shape or resistance to diseases. With such extended features, the apple variety would be represented as point clouds similarly to what was found here for color histogram. In this sense, the illustration provided in this report on color would actually be extendable without any effort to any kind of characteristics to be tested for automation in variety testing. The proposed approach specially with optimal transport can be adapted to higher dimensional feature spaces and thus offers a generic framework to extend the quest for automation in variety testing.

6. Acknowledgement

Authors are grateful to Angers Loire Metropole, Région des Pays de la Loire and CPVO. This work is supported by the Horizon 2020 Framework Program of the European Union under grant agreement No 817970 (INVITE : https://www.h2020-invite.eu/). Authors thank Roland Robic from INRAe for assistance in image acquisition.

References

[1] Jing Bai, Xiaohua Wang, and Licheng Jiao. Image retrieval based on color features integrated with anisotropic directionality. *Journal of Systems Engineering and Electronics*, 21(1):127–133, 2010.

[2] Michael F Barnsley. *Fractals everywhere*. Academic press, 2014.

[3] I. Bloch and Jamal Atif. Deux approches pour la comparaison de relations spatiales floues. transport optimal et morphologie mathématique. *Rev. d'Intelligence Artif.*, 29:595–619, 2015.

[4] Surina Borjigin and Prasanna K Sahoo. Color image segmentation based on multi-level tsallis–havrda–charvát entropy and 2d histogram using pso algorithms. *Pattern Recognition*, 92:107–118, 2019.

[5] Antoni Buades, Jose-Luis Lisani, and Jean-Michel Morel. Dimensionality of color space in natural images. *JOSA A*, 28(2):203–209, 2011.

[6] Julien Chauveau, David Rousseau, and François Chapeau-Blondeau. Fractal capacity dimension of three-dimensional histogram from color images. *Multidimensional Systems and Signal Processing*, 21(2):197–211, 2010.

[7] Julien Chauveau, David Rousseau, Paul Richard, and François Chapeau-Blondeau. Multifractal analysis of three-dimensional histogram from color images. *Chaos, Solitons & Fractals*, 43(1-12):57–67, 2010.

[8] Thomas M Cover and Joy A Thomas. Information theory and statistics. *Elements of Information Theory*, 1(1):279–335, 1991.

[9] Ju Han and Kai-Kuang Ma. Fuzzy color histogram and its use in color image retrieval. *IEEE Transactions on image Processing*, 11(8):944–952, 2002.

[10] Zulfiqar Hasan Khan, Irene Yu-Hua Gu, and Andrew G Backhouse. Robust visual object tracking using multi-mode anisotropic mean shift and particle filters. *IEEE transactions on circuits and systems for video technology*, 21(1):74–87, 2011.

[11] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple object tracking: A literature review. *Artificial Intelligence*, page 103448, 2020.

[12] Manuel Melgosa, Alain Trémeau, and Guihua Cui. Colour difference evaluation. In *Advanced color image processing and analysis*, pages 59–79. Springer, 2013.

[13] Dexter I Mercurio and Alexander A Hernandez. Classification of sweet potato variety using convolutional neural network. In *2019 IEEE 9th international conference on system engineering and technology (ICSET)*, pages 120–125. IEEE, 2019.

[14] Sapan Naik and Bankim Patel. Machine vision based fruit classification and grading-a review. *International Journal of Computer Applications*, 170(9):22–34, 2017.

[15] Rashmi Pandey, Sapan Naik, and Roma Marfatia. Image processing and machine learning for automated fruit grading system: A technical review. *International Journal of Computer Applications*, 81(16):29–39, 2013.

[16] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019.

[17] Julien Rabin, Sira Ferradans, and Nicolas Papadakis. Adaptive color transfer with relaxed optimal transport. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4852–4856. IEEE, 2014.

[18] Manali R Satpute and Sumati M Jagdale. Color, size, volume, shape and texture feature extraction techniques for fruits: a review. *Int. Res. J. Eng. Technol*, 3:703–708, 2016.

[19] Divya Srivastava, Rajesh Wadhvani, and Manasi Gyanchandani. A review: color feature extraction methods for content based image retrieval. *International Journal of Computational Engineering & Management*, 18(3):9–13, 2015.

[20] Junding Sun, Ximin Zhang, Jiangtao Cui, and Lihua Zhou. Image retrieval based on color distribution entropy. *Pattern Recognition Letters*, 27(10):1122–1126, 2006.

[21] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.

[22] Jana Wäldchen and Patrick Mäder. Machine learning for image based species identification. *Methods in Ecology and Evolution*, 9(11):2216–2225, 2018.

[23] Xiaoling Yang, Hanmei Hong, Zhaohong You, and Fang Cheng. Spectral and image integrated analysis of hyperspectral data for waxy corn seed variety classification. *Sensors*, 15(7):15578–15594, 2015.

[24] Chika Yinka-Banjo and Ogban-Asuquo Ugot. A review of generative adversarial networks and its application in cybersecurity. *Artificial Intelligence Review*, 53(3):1721–1736, 2020.