Automatic wheat ear counting using machine learning based on RGB UAV imagery

Jose A. Fernandez-Gallego[1,2,3], Peter Lootens[4], Irene Borra-Serrano[4,5], Veerle Derycke[6], Geert Haesaert[6], Isabel Roldán-Ruiz[4,7], Jose L. Araus[1,2,*], Shawn C. Kefauver[1,2,*]

[1] Plant Physiology Section, Department of Evolutionary Biology, Ecology and Environmental Sciences, Faculty of Biology, University of Barcelona, Diagonal 643, 08028, Barcelona, Spain; jfernaga46@alumnes.ub.edu, jaraus@ub.edu, sckefauver@ub.edu

[2] AGROTECNIO (Center for Research in Agrotechnology), Av. Rovira Roure 191, 25198, Lleida, Spain

[3] Programa de Ingeniería Electrónica, Facultad de Ingeniería, Universidad de Ibagué, Carrera 22 Calle 67, 730001, Ibagué, Colombia

[4] Plant Sciences Unit, Institute of Agricultural, Fisheries and Food Research (ILVO), Caritasstraat 39, 9090 Melle, Belgium; peter.lootens@ilvo.vlaanderen.be, irene.borra-serrano@ilvo.vlaanderen.be, isabel.roldan-ruiz@ilvo.vlaanderen.be

[5] Department of Biosystems, MeBios, KU Leuven, Leuven, Belgium

[6] Ghent University, Department Plants and Crops, Faculty of Bioscience Engineering, Valentin Vaerwyckweg1, 9000 Ghent, Belgium; veerle.derycke@ugent.be; geert.haesaert@ugent.be

[7] Ghent University, Department of Plant Biotechnology and Bioinformatics, Technologiepark 71, 9052 Ghent, Belgium

*Corresponding authors:

José L. Araus, Tel. + 34 934021469; E-mail address: jaraus@ub.edu

Shawn C. Kefauver, Tel. +34 934021465; M. +Fax +34 934112842

E-mail address: sckefauver@ub.edu

SUMMARY

In wheat and other cereals, the number of ears per unit area is one of the main yield determining components. An automatic evaluation of this parameter may contribute to the advance of wheat phenotyping and monitoring. There is no standard protocol for wheat ear counting in the field, and moreover it is time-consuming. An automatic ear counting system is proposed using machine learning techniques based on RGB images acquired from an unmanned aerial vehicle (UAV). Evaluation was performed on a set of 12 winter wheat cultivars with 3 nitrogen treatments during the 2017-2018 crop season. The automatic system uses a frequency filter, segmentation, and feature extraction with different classification techniques to discriminate wheat ears in micro-plot images. The relationship between the image-based manual counting and the algorithm counting exhibited high accuracy and efficiency. In addition, manual ear counting was conducted in the field for secondary validation. The correlations between the automatic and the manual in-situ ear counting with grain yield were also compared. Correlations between both ear counting systems were strong, particularly for the lower N treatment. Methodological requirements and limitations are discussed.

Keywords: Digital image processing, ECOFE, Field phenotyping, Machine learning, Wheat

SIGNIFICANCE STATEMENT

Ear density (ears/m^2) is one of the main agronomical yield components of wheat. This study represents a novel contribution to the field of RGB image processing for plant phenotyping using unmanned aerial vehicle (UAV) platforms. By combining high resolution RGB imagery with an automatic ear classification and counting system, we have shown that it is possible to assess ear density with high precision from an aerial platform. This is the first study successfully deploying this approach.

INTRODUCTION

High throughput plant phenotyping (HTPP) relies on the availability of advanced sensors, suitable image analysis and data mining tools (Araus and Cairns, 2014; Deery *et al.*, 2014). In recent years, research in this area has been growing exponentially, but field phenotyping is still perceived as a bottleneck for crop breeding due to the need for massive data collection and processing (Araus and Kefauver, 2018), image analysis tasks (Kelly *et al.*, 2015; Minervini *et al.*, 2015; Kefauver *et al.*, 2018), science community adaptation to new technologies (Singh et al., 2016), and the need to adapt sensors, algorithms and data management to the wide array of traits needed for plant phenotyping (Qiu *et al.*, 2018).

Grain weight, number of grains per ear and ear density (understood as the number of ears or spikes per unit ground area) are the most important yield components in wheat (García del Moral *et al.*, 2003; Slafer *et al.*, 2014; Simane *et al.*, 1993). An appropriate quantification of these components is therefore essential for wheat breeders to be able to assess the yield potential of breeding material in early generations. Traditionally, ear density is determined manually in-situ, by counting the number of ears present in a given area, which is time-consuming. In addition, as only a small subsection of the plot is usually considered, intra-plot heterogeneity might result in inaccurate estimations.

As an alternative to this approach, on-ground automatic ear counting systems have been developed, based on RGB (Red/Green/Blue), thermal, multispectral and laser images. In the case of thermal, multispectral and laser sensors, few image processing techniques have been developed, for instance, color thermal maps and contrast limited adaptive histogram equalization (CLAHE) (Fernandez-Gallego *et al.*, 2019a); threshold segmentation and denoising based on morphological filters (Zhou *et al.*, 2018a) for multispectral images; and in case of a laser sensor, voxel-based tree detection and mean shift approach (Velumani *et al.*, 2017). Nevertheless, RGB sensors have been widely used as proximal and remote sensing tools in many phenotyping tasks (Araus *et al.*, 2018) due to their relatively low cost (Qiu *et al.*, 2018; Araus *et al.*, 2018), high resolution (Deery *et al.*, 2014; Minervini *et al.*, 2015), and a fast adaptation to natural light conditions (Cointault *et al.*, 2008; Fernandez-Gallego, *et al.*, 2019c) that allows RGB sensors to acquire a faithful representation of an original scene even mounted on aerial platforms with continuous and unforeseen movements.

Different image processing techniques have been developed for ear counting using RGB sensors. These image processing techniques include (i) hybrid spaces with texture parameters (Cointault *et al.*, 2008); (ii) decorrelation stretching, scale-invariant feature transform (SIFT) descriptors and support vector machine (SVM) (Sadeghi-Tehran *et al.*, 2017; Zhu *et al.*, 2016); (iii) multi-feature extraction using color, texture and histogram, kernel principal component analysis (KPCA) and the twin-support-vector-machine (TWSVM) model (Zhou, *et al.*, 2018b); (iv) Laplacian frequency filter, median spatial filter and local peak segmentation (Fernandez-Gallego, *et al.*, 2018b), including a simulation and implications of lower resolution (Fernandez-Gallego, *et al.*, 2018a); and (v) convolutional neural networks (CNNs) using fast region-based CNN (Madec *et al.*, 2019). However, to date all automatic ear counting systems have been implemented only from the ground for resolution reasons, using zenithal RGB images acquired at less than one meter (Cointault *et al.*, 2008), around one meter (Fernandez-Gallego *et al.*, 2018b) or at most, a limited height above the crop: 2.5 m (Madec *et al.*, 2019), 2.9 m (Sadeghi-Tehran *et al.*, 2017), 3.5 m (Zhou, *et al.*, 2018a) and even 5 m (Zhu *et al.*, 2016).

To the best of our knowledge, there is no information in the literature regarding the use of RGB images acquired at further distances above the crop, for example from an Unmanned Aerial Vehicle (UAV), for ear counting. Resolution is a key factor for image processing: higher resolution allows for extracting more features from the input image compared to lower resolution images (Syrris *et al.*, 2015), which is critical for the detection of ears from any aerial platform. While this limited application in the past, the increasing availability of high resolution RGB imaging devices that provide higher pixel density and thus improved ground sampling distance (GSD) from a greater distance, now make it feasible to use UAV platforms for ear counting.

In this study, we propose an automatic wheat ear counting system using RGB images acquired from an UAV. A field trial comprising 12 modern wheat varieties tested under three fertilization conditions in four replicates was used for method development and validation. Orthophotos with a moderately high GSD of 0.24 cm/pixel were analysed with an image processing pipeline using filtering, segmentation, feature extraction and machine learning techniques. Manual in-situ and image-based counting were conducted for validation purposes along with grain yield.

RESULTS

Algorithm development and validation

The true positive and true negative classification accuracies of each classifier was calculated (Table 1). The classification accuracies of the cross-validation demonstrated a high percentage of correct prediction with a low standard error for k-nearest neighbors (kNN), support vector machine (SVM), decision trees (DT) and random forest (RF) (Table 1). The RF classifier reached the highest percent of accuracy of true positives (TP) and true negatives (TN) for both dates (June 4: TP = 98.0%, TN = 96.9%; June 19: TP = 98.8%, TN = 95.8%), while the generalized linear models (GLM) (TP = 65.2%) and native Bayes (nB) (TN = 78.5%) showed the lowest accuracy in terms of TP and TN, respectively, for June 4; and (nB) (TP = 87.9%) and discriminant analysis (DA) (TN = 75.3%) showed the lowest accuracy of TP and TN, respectively, in the case of June 19.

The *manual image-based counting* and the *algorithm counting* demonstrated high determination coefficient (R^2) cross-validation results with low standard error for SVM, DT and RF under further subplot inputs (Table 2). The RF classifier achieved the highest R^2 cross-validation values for both dates (June 4: R^2 = 0.82; June 19: R^2 = 0.87), while GLM showed the lowest R^2 cross-validation values for both dates (June 4: R^2 = 0.33; June 19: R^2 = 0.36).

The relationship between the *manual image-based counting* and the *algorithm counting* for the best classifier (RF) also showed a high determination coefficient for both dates (June 4: R^2 = 0.83, June 19: R^2 = 0.89) when using a linear regression without cross-validation (Figure 1).

Relationship between manual in-situ, algorithm counting and grain yield

The relationship between *manual in-situ counting* and *algorithm counting* with grain yield were assessed using the R^2 of the cross-validation. In the case of *algorithm counting*, the mean of nine subplots from the RF classifier were used (Table 3).

Using all data, determination coefficients showed no correlation between the *manual in-situ counting* and the *algorithm counting* for June 4 with grain yield (R^2 = 0.02 and R^2 = 0.04). In the case of *algorithm counting* for June 19 and June 4 + June 19 together, low correlations (R^2 = 0.14, R^2 = 0.28; respectively) were observed. Including G effects, the results showed low correlation for all four input data (R^2 = 0.06, R^2 = 0.11, R^2 = 0.16 and R^2 = 0.20; respectively). By contrast including N effects (i.e. combining data of the three different nitrogen fertilization trials) the correlation increased (R^2 = 0.34–0.36). In the case of G+N effects, the determination coefficient also increased (R^2 = 0.41–0.46). Grouping by N treatments, the best correlations were achieved for N.std and N+30 (R^2 = 0.14 and R^2 = 0.20; respectively) on June 4; while for N-50 (R^2 = 0.42) it was on June 19. Additionally, using two input data together (June 4 + June 19), the correlation increased for N-50 (R^2 = 0.46). Moreover, the *manual in-situ counting* did not improve the strength of the correlations of the algorithm counting against grain yield when this variable was added in a multiple linear regression model (data not shown).

The relationship between the *manual in-situ counting* and the *algorithm counting* were calculated using the determination coefficient of the cross-validation and the complete data set for each date of measurement (June 4, June 19) individually and also combining both dates (June 4 + June 19). No significant relationships were noted between the two counting techniques for either date individually or the combined dates. Moreover, the genotype (G), nitrogen fertilization (N), G+N and G*N effects on the relationship between *manual in-situ counting* and *algorithm counting* were tested with no interactions observed.

Additionally, we also grouped the data by N treatments in order to analyze the correlation between the *manual in-situ counting* and the *algorithm counting*, but no correlations were observed ($R^2 \approx 0.0$) for all cases.

DISCUSSION

Agronomical yield components are key to dissecting how wheat responds to growing conditions as well as forming the basis for the genetic advancement of grain yield (Slafer *et al.*, 2014). In our study, the classification showed high accuracy for TP and TN in the training and classifying step; four classification techniques showed results above 90% for TF and TN, which means relevant information contributed by the feature extraction step to classification (Kumar and Bhatia, 2014) (Table 1). Across all machine learning techniques, RF achieved the highest classification accuracy for both dates of measurement for *Class1* and *Class2* (June 4: TP = 98.0%, TN = 96.9%; June 19: TP = 98.8%, TN = 95.8%) (Table 1). In the case of validation using *manual image-based counting*, RF also achieved the highest cross-validation results (June 4: R^2 = 0.82, June 19: R^2 = 0.87), but the other classifiers achieved much lower results (Table 2). Particularly, the best three classifications techniques using *manual image-based counting* and *algorithm counting* cross-validation were SVM, DT and RF (Table 2). In the linear regression, RF also achieved the highest determination coefficient on both dates (June 4: R^2 = 0.83, June 19: R^2 = 0.89 in Figure 5), in this case cross-validation was not performed. In our case, RF has performed better than the other classifiers for our shape and statistic features; this may be due to RF often showing higher performance in classification tasks when multi-dimensional data is used (Belgiu and Drăgu, 2016), as it is usually does in other remote sensing applications. RF robustness to outliers and noise (Breiman, 2001) and bootstrapping aggregations together with the many tree learners used in the RF classifier proved less sensitive to the quality of the training samples than other machine learning classifiers (Belgiu and Drăgu, 2016). These characteristics allowed for an effective prediction capacity and also resulted in less overfitting (Berk, 2013). Moreover, for remote sensing approaches, parametric classifiers such as DA, GLM, and nB have shown limitations dealing with multimodal distributions (Liu *et al.*, 2011), while by contrast, nonparametric classifiers such as kNN, NN, SVM, DT and RF have shown better results under multimodal distributions (Marsum *et al.*, 2018; Maulik and Chakraborty, 2017).

To date, automatic ear counting systems, regardless of the acquisition equipment, have been evaluated from ground, using only a portion of the area of the plot (Cointault *et al.*, 2008; Zhu *et al.*, 2016; Sadeghi-Tehran *et al.*, 2017; Velumani *et al.*, 2017; Zhou, *et al.*, 2018a,b; Fernandez-Gallego, *et al.*, 2018a,b,2019a,b; Madec *et al.*, 2019). Although the use of a UAV platform allows the acquisition of the complete area of the phenotyping micro-plots, multispectral and thermal sensors that have fairly low spatial resolution from aerial platforms and laser sensors are still relatively costly. RGB sensors are not without their limitations; images taken on June 6, June 25 and July 3 under direct sunlight conditions (sunny days) were discarded. Due to the sunlight reflections on bending leaves, it was hard to differentiate between ears and reflections on the leaves, thus making it impossible to do a correct visual (i.e. manual) ear detection on the orthomosaic images for validation. Therefore, the relatively low spatial resolution from the UAV combined with sunny day restrictions (blurring and degraded orthophotos due to sunlight reflections) did not permit precise reconstruction of the orthophoto at the canopy scale (Ortega-Terol *et al.*, 2017). As a result, these resolution and light conditions affect the number of matching features found for the Structure from Motion (SfM) process used to build the orthophoto

(Aasen *et al.*, 2018). Previous studies have shown reduction in the ear recognition accuracy due to lower resolution (Fernandez-Gallego, *et al.*, 2018a). Nonetheless, under cloudy sky conditions, RGB orthophotos can precisely reconstruct the ears, leaves and soil for recognition purposes.

On the other hand, although the ear density forms part of the main yield components, previous studies in ear recognition have not used this information in order to understand the relationship between (automatic) ear counting systems with grain yield. In this study using the complete plot size area, the automatic ear counting system and cross-validation technique overall results showed no correlation with grain yield at June 4 (R^2 = 0.04) and low correlation at June 19 (R^2 = 0.14). In the same way, for *manual in-situ counting* no correlation with grain yield (R^2 = 0.02) was observed. Nevertheless, when including G and N effects the determination coefficient increased. Furthermore, the correlation for N effects is higher (R^2 = 0.34–0.36) compared with G effects (R^2 = 0.06–0.20) for manual and algorithm counting as input data, which suggests that the relationship between the *manual in-situ counting* and the *algorithm counting* with grain yield is more supported by the nitrogen treatment factors than genotype differences. In fact, Slafer et al. (2014) also concluded that the effect of nitrogen fertilization may affect ear density far more than the genotypic differences across cultivars. In addition, the determination coefficient including G+N effects also increased (R^2 = 0.41–0.46) in all cases. In general, for grain yield assessment, *algorithm counting* performed better in terms of correlation with grain yield than *manual in-situ counting*. Regarding this relationship, no correlations were observed between manual and algorithm counting. This may be due to the limited reference measurement of two half linear meters used for *manual in-situ counting* compared to the complete plot area footprint acquired from the UAV platform. On the other hand, automatic counting from zenithal images only considers the upper (i.e. exposed) ears, which usually correspond to the main and primary tillers, while manual counting considers all the ears, including those from secondary and tertiary tillers, which frequently are placed to lower levels within the canopy. Since the contribution of secondary and tertiary tillers to grain yield is usually minor if not negative (Ishag and Taha, 1974), this might explain the fact that in-situ ear counting correlated even weaker with grain yield than the values of the automatic counting.

Grouping by nitrogen treatments, the best correlations of ear density against grain yield were achieved for June 4 and June 19 in N+30 and N-50 (R^2 = 0.20 and R^2 = 0.42, respectively). June 4 + June 19 achieved the best correlation in the N-50 treatment (R^2 = 0.46). In this way, the number of ears may be assumed to be affected by genotype characteristics as well as the N treatments, and therefore these considerations should be added as variables in the model for grain yield assessment. A higher correlation between ear density and grain yield was observed at the lower N treatment (N-50). This could be explained by less hidden ears as less nitrogen fertilization diminishes tillering capacity (Power and Alessi, 1978), which decreases hidden ears out of the reach of the automatic counting system. Fewer ears in total may also result in less instances of overlapping ears. Considering the massive amount of data acquired from an aerial platform and the lower spatial resolution due to increased distance between object and camera, the image processing systems combined with machine learning techniques demonstrated an effective data management and image interpretation capacity.

CONCLUSIONS

We have presented details for an automatic system for ear counting using RGB aerial images captured from a UAV platform that includes a pipeline for employing machine learning techniques for image classification and ear counting. The ear counting system was able to identify wheat ears with relatively high accuracy considering the reduction in image resolution when using a UAV platform (compared with ground-acquired images). Though similar techniques have been implemented previously from the ground, adaptation to UAVs should provide improved throughput and more complete plot coverage making automatic ear counting feasible to large phenotyping studies. Moreover, our approach demonstrated higher accuracy than the already published studies from the ground. In this way, the system may be used for targeted trait breeding in cereal phenotyping that could be translated into increasing yield gains through indirect selection (Araus *et al.*, 2018). Nevertheless, in scaling from ground to aerial platforms, sensitivity to light conditions increased and should be investigated further.

The automatic ear counting demonstrated better correlations with grain yield compared to the manual in-situ measurements and thus may provide for improved direct selection of higher performance varieties. Including the effects of G+N factors increased the R^2 with grain yield and the R^2 with grain yield was higher when including N than G factors, indicating that the ear counting relationship with grain yield was more supported by the N treatment factors than by G differences. Finally, the highest correlation between automatic ear counting and grain yield were achieved at the lowest N treatment, where fewer hidden ears and lower tillering capacity occurs, which could indicate more applicability in real growing conditions in farmer fields under rainfed or low N conditions. Future potential for increased image resolution and processing and 3D imaging, along with ear size/volume estimation, could be next steps to achieve higher correlations with grain yield for direct selection.

EXPERIMENTAL PROCEDURES

Plant material and growing conditions

A field trial with twelve winter wheat (*Triticum aestivum* L.) varieties (Benchmark, Bologna, Nara, Chambo, Henrik, Hondia, Diego, Julius, Lilli, Siskin, RGT Reform and Sobervio) and 3 nitrogen levels was established on a sandy loam soil at the experimental farm of Ghent University in Bottelare, Belgium (lat. 50.96 N, long. 3.78 E). Nitrogen fertilization levels included the standard recommended in the area (N.std), standard+30% (N+30) and standard-50% (N-50) (Figure 2). The trial was part of a multi-location field trial run by the European Consortium for Open Field Experimentation (ECOFE, https://www.ecofe.eu) (Stützel *et al.*, 2016). The field trial was set-up as a split-plot design with varieties grown in plots of 1.5 m by 12 m at a sowing density of 350 seeds m^{-2}, and with four replicates. Nitrogen fractions were given on March 22, 2018, on April 27, 2018 and on May 25, 2018 respectively. Nitrogen fertilizer used was ammonium nitrate 27%. The accumulated rainfall during the growing season was 513.7 mm and the average temperature was 10.2 ºC (Figure 2). Plots were mechanically harvested on July 14, 2018.

Orthophotos, plot and subplot images

RGB images were acquired using a 12-rotor UAV (Model Onyxstar Hydra-12, Altigator, Belgium) flying a predefined flight plan with 70% front and side overlap of the images. Flight altitude was 25 m and flight speed 2.5 m/s. The camera was triggered based on the waypoints of the flight plan. The images were taken with a Sony α6000 (Sony Corporation, Japan), which is a 24.5-megapixel resolution camera with a 23.5 x 15.6 mm sensor size.

The camera has a native resolution of 6000 x 4000 pixels and was equipped with a 35 mm focal length lens. All images were taken in manual mode to avoid different settings in successive images. Trigger speed, aperture and sensitivity to light (ISO) were adjusted in the field before the start of the flight, and the focus was set to automatic mode. Files were stored in RAW and JPG format. Images were acquired under diffuse light conditions (cloudy days) at two dates: June 4 and June 19, 2018 corresponding to 61 and 75 growth stages (GS) of the Zadoks scale (Zadoks *et al.*, 1974), respectively. Images were also acquired under direct sunlight conditions (sunny days) at three dates: Jun 6, June 25 and July 3, 2018 corresponding to 61, 85 and 90 GS. The images acquired at these dates contribute to understand the low resolution + sunlight reflections issues in the reconstruction of the orthophoto at canopy scale from aerial platforms. Agisoft Photoscan software (version 1.2.3, Agisoft LLC, St. Peterburg, Russia) was used to build geo-referenced orthophotos using nine ground control points (Figure 2). The coordinates of those points were determined with an RTK GPS (Stonex S10 GNSS, Stonex SRL, Italy). The spatial resolution was defined automatically by the software based on the camera parameters and flight altitude. In practice it ranged from 0.23 cm/pixel to 0.24 cm/pixel. For comparison reasons the orthomosaics of both dates were exported at the lowest spatial resolution, which was 0.24 cm/pixel. Halcon Image Analysis software (version 11, MVTec Software GmbH, Munich, Germany) was used to delineate each plot avoiding borders and to divide it into nine subplots (Figure 3). The resulting plots had a footprint size of 0.96 m x 8.64 m; therefore, each subplot had a footprint size of 0.96 m x 0.96 m. Images from the center of each plot (subplot #5, Figure 3) were selected for training and validation purposes in order to avoid possible errors due to the distortion or perspective

(Jaud *et al.*, 2018). The complete plots (from subplot #1 to #9) were used for the automatic wheat ear counting system. In total 2592 subplot images were processed.

Automatic wheat ear counting system using UAV imagery

The algorithm for ear counting is based on the pipeline developed by Fernandez-Gallego *et al*. (2018b) which includes three main steps: Laplacian frequency filter, median filter and Find Maxima. For the case presented in this study based on UAV imagery, we have used the Laplacian frequency filter and Find Maxima steps (Fernandez-Gallego, *et al.*, 2018b), and have included two additional steps: (a) feature extraction and (b) training and classifying (Figure 4). In this adaptation of the original algorithm, the median filter step was excluded in order to maintain the high frequency information of the canopy after the Laplacian frequency filter step, considering the greater distance between the sensor and canopy. The algorithms were developed in ImageJ (version 2.0.0-rc-69, NIH, Bethesda, MD, USA) and MATLAB (version R2014b, Mathworks, Inc., MA, USA). Therefore, the final pipeline algorithm consists of four steps: (1) Laplacian frequency filter, (2) Find Maxima, (3) feature extraction and (4) training and classifying. The Laplacian filter was applied as a wide frequency filter to avoid unwanted objects such as awns, leaves and soil; this isotropic filter responds independently of image discontinuities detects and changes in the different directions of the image. Find Maxima was then used for local peak detection in order to define image areas where ears could be located. This step creates a binary image (segmentation) using the pixel value for each local peak and its nearest neighbor pixels; in addition, this step reduces the overlapping ear errors by first locating the center of the ears that contributes to isolating neighboring ears. We developed a feature extraction step in order to obtain numerical characteristics related to shape, color and statistical measurements (such as mean and standard deviation) for each local peak detected in the previous step. Finally, a training and classifying step was developed to decide between two classes; *Class1*: Ear, *Class2*: Non-Ear. The image processing system proposed uses as inputs all of subplot images in batch (Figure 4). This means that the estimation of the number of ears per plot is the sum of each subplot (from the subplot #1 to #9 not including the buffer area, which was excluded in the preliminary plot delineation). The sequence of steps implemented is described in Figure 4. Laplacian frequency filter and Find Maxima steps are thoroughly discussed in Fernandez-Gallego *et al*., 2018 (Fernandez-Gallego, *et al.*, 2018b).

Feature extraction

The feature extraction was developed using Analyze Particles after Find Maxima (Schneider *et al.*, 2012). The binary areas from the Find Maxima step were used as masks to calculate features (Figure 4). We have extracted shape descriptors and statistical information from the original RGB image and its color channels, such as area, height, width, Feret, circularity, mean, standard deviation, mode and more measurements, totaling 30 features (*m* = 30 in Figure 4). An overview of the complete set of features extracted and their definition can be found in Supplementary Table 1 (Table S1) and Supplementary Table 2 (Table S2). A feature selection was developed in order to reduce the dimensionality of the data in preference to feature reduction by transformation. This allowed us to keep the units and meaning of all variables (Tripathy and Sahoo, 2015), thereby losing less of the information contained in the original features space (Khalid *et al.*, 2014). For this purpose the Sequential Feature Selection (SFS) (Kohavi and John, 1997) was used with forward direction and stop criterium of 20 features (*s* = 20 in Figure

4). Supplementary Table 3 (Table S3) also shows the selected features.

Training and classifying

The training and classifying steps were developed using diverse machine learning techniques in order to compare the capability of each technique to discriminate between the *Class1* and the *Class2* labeled objects using the manual marks as a reference.

For labeling purposes (*Class1* or *Class2*), red marks were used to delineate all the ears manually. We marked the complete area covered by each single ear in order to maximize the information gained in this step. Figure 3 shows in the dotted line rectangle a sample of the manual marks. For each date of measurement (June 4 and June 19), 16 subplots images were manually marked for training and classifying; totaling 32 subplots. Different subplots were used for each date of measurement. The overlapping areas between automatic and manual selection were represented in white color and correspond to the *Class1*; the non-overlapping areas were represented in blue color and correspond to the *Class2* (Figure 4). The same data was used for training and validation purposes for each classification technique. While the use of supervised classification techniques requires time inputs, the robustness of the models produced should make them repeatedly applicable across similar growth stages and for similar varieties as the original training and validation data.

The supervised machine learning techniques implemented for classification were: (i) discriminant analysis (DA) (Box, 1949) using linear discriminant, (ii) generalized linear models (GLM) (Dobson and Barnett, 2008) using binomial distribution, (iii) k-nearest neighbors (kNN) (Mitchell, 1997) using Euclidian distance, (iv) native Bayes (nB) (Mitchell, 1997) using normal distribution, (v) neural feedforward neural networks (NN) (Beale *et al.*,

2015) with ten hidden layers, (vi) support vector machine (SVM) (Cristianini and Shawe-Taylor, 2000) with a Gaussian radial basis kernel (Cristianini and Scholkopf, 2002; Liu *et al.*, 2012), (vii) decision trees (DT) (Sheppard, 2017) using a binary tree and (viii) random forest (RF) (Breiman, 2001) using bootstrapping aggregation. The selected parameters of each machine learning technique can be found in Supplementary Table 4 (Table S4). The same data from the feature extraction step were used to train each classifier. The classification accuracy was calculated using cross-validation in terms of true positives (TP) and false positives (FP) and true negatives (TN) and false negatives (FN) based on the confusion matrix (Tso and Mather, 2009). The TP and TN correspond to the *Class1* and *Class2* correctly classified by the automatic system, respectively.

Algorithm validation

In addition to the training and classification section, the performance of the image processing system using an UAV platform was also tested at anthesis (June 4, GS = 61) and early grain filling (June 19, GS = 75) growth stages using additional subplots from the total dataset of subplot images. These additional subplots were not used at the training and classifying procedure (previous section). In order to further validate the automatic ear counting system, the algorithm results were compared with the manual image-based marks on the same images. The number of ears automatically detected by the image processing system is referred to as the *algorithm counting* and the number of ears manually marked is referred to as the *manual image-based counting*. For each N treatment and date of measurement we have used 24 subplots, totaling 144 subplots. For this *manual image-based counting*, only one red dot was marked per ear in the original image with the same color value, circular shape and size; then, we used a simple algorithm to search the same color and shape marks and count them. The corresponding *manual image-based counting* and *algorithm counting* numbers were expressed in terms of ears per square meter in order to use standard units. To determine the prediction power of the automatic ear counting, we calculated cross-validation R^2 values between the *manual image-based counting* and the *algorithm counting* for each discrimination technique and date of measurement.

A *manual in-situ counting* at crop maturity was carried out. For each plot, two half linear meter counts were used as a reference for the number of ears. Two different rows near the centre of the plot were selected. The *manual in-situ counting* was calculated as the sum of the number of ears counted in each half linear meter divided by the ratio between the plot width and the number of rows per plot.

STATISTICAL ANALYSIS

Data analyses were performed using R Studio (version 1.2.135, R Foundation for Statistic Computing, Vienna 2018) and MATLAB (version R2014b, Mathworks, Inc., MA, USA). Determination coefficients of linear regressions (LR) and multiple linear regression (MLR), as well as the root mean square error (RMSE), were calculated. The effects of genotype (G), nitrogen (N) and genotype plus nitrogen (G+N) fertilization factors on grain yield (GY) were also calculated using LR and MLR. The G by N interactions (G*N) were analysed using analysis of variance (ANOVA). To validate the robustness of the classification and validation a five-fold cross-validation (CV) was performed. In total, 100 CV runs (20 times five-fold CV) were performed. We have not included the model equation per each factor as we have performed a five-fold cross-validation (20 times five-fold cross-validation), totaling 100 different model equation per each factor and input data. The data was plotted using SigmaPlot (version 12, Systat Software, Inc., San Jose California USA).

ACKNOWLEDGMENTS

The authors want to thank the technical staff of UGent for the setup and maintenance of the trial and the technical staff of ILVO, especially to Thomas Vanderstocken, for the UAV flights and the pre-processing of the data and the technical assistance of Viviana Espejo for the in-situ measurements in the field and the image-based validation. The multilocation trial was executed in the framework of the ECOFE network (www.ecofe.eu). This work was established during an internship of Jose A. Fernandez-Gallego at ILVO (B). This work was supported by the project “Formación de Talento Humano de Alto Nivel” [project number BPIN 2013000100103] approved by the “Fondo de Ciencia, Tecnología e Innovación”, from the “Sistema General de Regalías”; and “Gobernación del Tolima-Universidad del Tolima, Colombia” as the sole funding source of the first author Jose A. Fernandez-Gallego. Contribution of Jose L. Araus was supported in part by ICREA Academia, Generalitat de Catalunya, Spain.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

SUPPORTING INFORMATION

Additional supporting information can be found in the online version of this article.

OPEN RESEARCH BADGE

DATA AVAILABILITY STAMENT

The image pre-processing in this study was initially completed using Halcon due to the very large field file size, but the first step of micro-plot segmentation from an orthophoto may also be conducted using, for instance, the MosaicTool, as described in Gracia-Romero *et al*. (2019). The authors are working to incorporate the new Technical Advancements presented here into the pre-existing CerealScanner plug-in, which was produced as open source software to share previous ground-based cereal ear counting and trait-based phenotyping tools. Meanwhile, the FIJI and MATLAB code developed as part of this study, as well as a subset of the original training and validation data of the data presented here have been made available in an open-access folder with the title of this publication within the GitLab of the CerealScanner, link. For video methods instructions on the use of the CerealScanner itself see Fernandez-Gallego, *et al*. (2019), JoVE.

Table S1: Shape descriptors: In the feature extraction step, shape descriptors from the original Red/Green/Blue (RGB) image, including Area, Feret distance, Min Feret distance, Width, Height, Raw integrated distance, Circularity, Aspect ratio, Solidity and Round were extracted. The visual representation corresponds to one area segmented at the Find maxima step into the image processing system. The shape descriptors were calculated using the complete color image. In summary, 10 shape descriptors.

Table S2. Statistical descriptors: In the feature extraction step, statistical descriptors from the original Red/Green/Blue (RGB) color channels, including mean (Mean), standard deviation (Std), mode (Mode), maximum (Max), and minimum (Min). These descriptors were extracted from the image using the value of all of the pixels of each individual area segmented during image processing pipeline. The statistical descriptors were calculated using the complete color image (5 descriptors). In addition, the statistical descriptors were also calculated for each color channel separately (R, B and G color channels, 15 descriptors). In summary, 20 statistical descriptors.

Table S3. Feature selection: The Sequential Feature Selection (SFS) was used to select the 20 features from the complete 30 features extracted as detailed in Tables S1 and S2. The complete shape descriptors and statistical descriptors (for the color image) were selected (15 features). Moreover, the statistical descriptors of Mean (in G color channel), Std (in R, G, B color channels), Max (in G color channel) were also selected (5 features) for each color channel, totaling 20 features selected per segmented area.

Table S4. Machine learning parameters: The following parameters were used for each machine learning technique in the training and classifying steps.

REFERENCES

Aasen, H., Honkavaara, E., Lucieer, A. and Zarco-Tejada, P.J. (2018) Quantitative remote sensing at ultra-high resolution with UAV spectroscopy: a review of sensor technology, measurement procedures, and data correctionworkflows. *Remote Sens.*, **10**, 1–42.

Araus, J.L. and Cairns, J.E. (2014) Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.*, **19**, 52–61.

Araus, J.L. and Kefauver, S.C. (2018) Breeding to adapt agriculture to climate change: affordable phenotyping solutions. *Curr. Opin. Plant Biol.*, **45**, 237-247.

Araus, J.L., Kefauver, S.C., Zaman-Allah, M., Olsen, M.S. and Cairns, J.E. (2018) Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.,* **23**, 451-466.

Beale, M.H., Hagan, M.T. and Demuth, H.B. (2015) *Neural Network Toolbox™ User's Guide*, The MathWorks, Inc.

Belgiu, M. and Drăgu, L. (2016) Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.*, **114**, 24–31.

Berk, R.A. (2013) Random forests. In *Statistical Learning from a Regression Perspective*. New York, NY: Springer New York, pp. 1–63.

Box, G.E.P. (1949) A general distribution theory for a class of likelihood criteria. *Biometrika*, **36**, 317-346.

Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, pp. 5–32.

Cointault, F., Guerin, D., Guillemin, J. and Chopinet, B. (2008) In-field *Triticum aestivum* ear counting using colour-texture image analysis. *New Zeal. J. Crop Hortic. Sci.*, **36**, 117–130.

Cristianini, N. and Scholkopf, B. (2002) Support vector machines and kernel methods: the new generation of learning machines. *AI Mag.*, **23**, 31-41.

Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge: Cambridge University Press.

Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X. and Furbank, R. (2014) Proximal remote sensing buggies and potential applications for field-based phenotyping. *Agronomy*, **4**, 349–379.

Dobson, A.J. and Barnett, A.G. (2008) *An Introduction to Generalized Linear Models,* *Third Edition*, Boca Ratón, FL: Chapman and Hall/CRC.

Fernandez-Gallego, J.A., Buchaillot, M.L., Aparicio Gutiérrez, N., Nieto-Taladriz, M.T., Araus, J.L. and Kefauver, S.C. (2019a) Automatic wheat ear counting using thermal imagery. *Remote Sens.*, **11**, 751.

Fernandez-Gallego, J.A., Buchaillot, M.L., Gracia-Romero, A., et al. (2019b) Cereal crop ear counting in field conditions using zenithal RGB images. *J. Vis. Exp.*, **e58695**, 10.

Fernandez-Gallego, J.A., Kefauver, S. C., Aparicio Gutiérrez, N., Nieto-Taladriz, M.T. and Araus, J. L. (2018a) Automatic wheat ear counting in-field conditions: simulation and implication of lower resolution images. In *Proc. SPIE 10783, Remote Sensing for Agriculture, Ecosystems, and Hydrology XX, 107830M*. p. 23.

Fernandez-Gallego, J.A., Kefauver, S. C., Aparicio Gutiérrez, N., Nieto-Taladriz, M.T. and Araus, J. L. (2018b) Wheat ear counting in-field conditions: high throughput and low-cost approach using RGB images. *Plant Methods*, **14**, 22.

Fernandez-Gallego, J.A., Kefauver, S.C., Vatter, T., Aparicio Gutiérrez, N., Nieto-Taladriz, M.T. and Araus, J.L. (2019c) Low-cost assessment of grain yield in durum wheat using RGB images. *Eur. J. Agron.*, **105**, 146–156.

García del Moral, L.F., Rharrabti, Y., Villegas, D. and Royo, C. (2003) Evaluation of Grain Yield and Its Components in durum wheat under mediterranean conditions. *Agron. J.*, **95**, 266-274.

Gracia-Romero, A., Kefauver, S.C., Fernandez-Gallego, J.A., Vergara-Díaz, O., Nieto-Taladriz, M.T. and Araus, J.L. (2019) UAV and ground image-based phenotyping: a proof of concept with durum wheat. *Remote Sens.*, **11**, 1244.

Ishag, H.M. and Taha, M.B. (1974) Production and survival of tillers of wheat and their contribution to yield. *J. Agric. Sci.*, **83**, 117–124.

Jaud, M., Passot, S., Allemand, P., Dantec, N. Le, Grandjean, P. and Delacourt, C. (2018) Suggestions to limit geometric distortions in the reconstruction of linear coastal landforms by SfM photogrammetry with PhotoScan® and MicMac® for UAV surveys with restricted GCPs pattern. *Drones*, **3**, 2.

Kefauver, S.C., Araus-Serret, I., Vergara-Díaz, O., Bort, J., El-haddad, G., Nieto-Taladriz, M.T., Aparicio, N. and Araus, J.L. (2018) Challenges and bottlenecks in UAV phenotyping. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 8240–8243.

Kelly, D., Vatsa, A., Mayham, W., Ngô, L., Thompson, A. and Kazic, T. (2015) An opinion on imaging challenges in phenotyping field crops. *Mach. Vis. Appl.*, **27**, 1–14.

Khalid, S., Khalil, T. and Nasreen, S. (2014) A survey of feature selection and feature extraction techniques in machine learning. *Proc. 2014 Sci. Inf. Conf. SAI 2014*, 372–378.

Kohavi, R. and John, G.H. (1997) Wrappers for feature subset selection. *Artif. Intell.*, **97**, 273–324.

Kumar, G. and Bhatia, P.K. (2014) A detailed review of feature extraction in image processing systems. *Int. Conf. Adv. Comput. Commun. Technol. ACCT*, 5–12.

Liu, K., Shi, W. and Zhang, H. (2011) A fuzzy topology-based maximum likelihood classification. *ISPRS J. Photogramm. Remote Sens.*, **66**, 103–114.

Liu, Z., Zuo, M.J. and Xu, H. (2012) Parameter selection for Gaussian radial basis function in support vector machine classification. *Proc. 2012 Int. Conf. Qual. Reliab. Risk, Maintenance, Saf. Eng. ICQR2MSE 2012*, 576–581.

Madec, S., Jin, X., Lu, H., Solan, B. De, Liu, S., Duyme, F., Heritier, E. and Baret, F.

(2019) Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.*, **264**, 225–234.

Marsum, M.A., Arsa, D.M.S., Hermawan, I., Jatmiko, W. and Nurhadiyatna, A. (2018) Multicodebook neural network using intelligent k-means clustering based on histogram information for multimodal data classification. *2018 Int. Work. Big Data Inf. Secur*, 129–135.

Maulik, U. and Chakraborty, D. (2017) Remote sensing image classification: a survey of support-vector-machine-based advanced techniques. *IEEE Geosci. Remote Sens. Mag.*, **5**, 33–52.

Minervini, M., Scharr, H. and Tsaftaris, S.A. (2015) Image analysis: the new bottleneck in plant phenotyping [applications corner]. *IEEE Signal Process. Mag.*, **32**, 126–131.

Mitchell, T.M. (1997) *Machine learning*, New York, NY, USA.: McGraw-Hill, Inc.

Ortega-Terol, D., Hernandez-Lopez, D., Ballesteros, R. and Gonzalez-Aguilera, D. (2017) Automatic hotspot and sun glint detection in UAV multispectral images. *Sensors*, **17**, 1–16.

Power, J.F. and Alessi, J. (1978) Tiller development and yield of standard and semidwarf spring wheat varieties as affected by nitrogen fertilizer. *J. Agric. Sci.*, **90**, 97-108.

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G. and Li, M. (2018) Sensors for measuring plant phenotyping: a review. *Int. J. Agric. Biol. Eng.*, **11**, 1–17.

Sadeghi-Tehran, P., Sabermanesh, K., Virlet, N. and Hawkesford, M.J. (2017) Automated method to determine two critical growth stages of wheat: heading and flowering. *Front. Plant Sci.*, **8**, 1–14.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH image to imageJ: 25 years of image analysis. *Nat Meth*, **9**, 671–675.

Sheppard, C. (2017) *Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting*, Austin, Texas, USA.: CreateSpace Independent Publishing Platform.

Simane, B., Struik, P.C., Nachit, M.M. and Peacock, J.M. (1993) Ontogenetic analysis of yield components and yield stability of durum wheat in water-limited environments. *Euphytica*, **71**, 211–219.

Singh, A., Ganapathysubramanian, B., Singh, A.K. and Sarkar, S. (2016) Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.*, **21**, 110–124.

Slafer, G.A., Savin, R. and Sadras, V.O. (2014) Coarse and fine regulation of wheat yield components in response to genotype and environment. *F. Crop. Res.*, **157**, 71–83.

Stützel, H., Brüggemann, N. and Inzé, D. (2016) The future of field trials in Europe: establishing a network beyond boundaries. *Trends Plant Sci.*, **21**, 92–95.

Syrris, V., Ferri, S., Ehrlich, D. and Pesaresi, M. (2015) Image enhancement and feature extraction based on low-resolution satellite data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **8**, 1986–1995.

Tripathy, S. and Sahoo, P.L. (2015) A Survey of different methods of clustering for anomaly detection. *Int. J. Sci. Eng. Res.*, **6**, 351–357.

Tso, B. and Mather, P. (2009) *Classification Methods for Remotely Sensed Data, Second Edition*, CRC Press.

Velumani, K., Oude Elberink, S., Yang, M.Y. and Baret, F. (2017) Wheat Ear Detection in Plots by Segmenting Mobile Laser Scanner Data. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. pp. 149–156.

Zadoks, J., Chang, T. and Konzak, C. (1974) A decimal growth code for the growth stages of cereals. *Weed Res.*, **14**, 415–421.

Zhou, C., Liang, D., Yang, X., Xu, B. and Yang, G. (2018a) Recognition of wheat spike from field based phenotype platform using multi-sensor fusion and improved maximum entropy segmentation algorithms. *Remote Sens.*, **10**, 246.

Zhou, C., Liang, D., Yang, X., Yang, H., Yue, J. and Yang, G. (2018b) Wheat Ears Counting in Field Conditions Based on Multi-Feature Optimization and TWSVM. *Front. Plant Sci.*, **9**, 1024.

Zhu, Y., Cao, Z., Lu, H., Li, Y. and Xiao, Y. (2016) In-field automatic observation of wheat heading stage using computer vision. *Biosyst. Eng.*, **143**, 28–41.

TABLES

Table 1. Classification accuracy of the automatic wheat ear counting system using UAV imagery for each classification technique and date. Training and classifying step used cross-validation to calculate the confusion matrix. Standard error (se) was calculated for each result for each true positives (TP) and true negatives (TN). For details about the classification techniques assayed see Material and Methods. Discriminant analysis (DA), generalized linear models (GLM), k-nearest neighbors (kNN), native Bayes (nB), neural feedforward neural networks (NN), support vector machine (SVM), decision trees (DT) and random forest (RF).

Classification technique	June 4				June 19			
	TP (%) *Class1*	se	TN (%) *Class2*	se	TP (%) *Class1*	se	TN (%) *Class2*	se
DA	90.7	0.09	80.2	0.16	93.2	0.07	75.3	0.17
GLM	65.2	0.13	93.7	0.08	92.2	0.08	77.5	0.17
kNN	96.4	0.17	94.1	0.28	97.0	0.15	93.9	0.27
nB	86.1	0.11	78.5	0.15	87.9	0.11	75.6	0.16
NN	90.5	0.09	79.7	0.16	93.1	0.08	79.8	0.15
SVM	94.0	0.14	93.5	0.10	95.3	0.14	91.7	0.13
DT	94.3	0.13	93.0	0.16	94.3	0.13	91.6	0.20
RF	98.0	0.10	96.8	0.16	98.8	0.06	95.8	0.20

Table 2. Determination coefficient (R^2) of the cross-validation results between the *manual image-based counting* and the *algorithm counting* for each classification technique and date. Standard error (se) was calculated for each result. Discriminant analysis (DA), generalized linear models (GLM), k-nearest neighbors (kNN), native Bayes (nB), neural feedforward neural networks (NN), support vector machine (SVM), decision trees (DT) and random forest (RF).

Classification technique	cross-validation R^2 value			
	June 4	se	June 19	se
DA	0.59	0.02	0.63	0.02
GLM	0.33	0.03	0.36	0.03
kNN	0.58	0.02	0.71	0.02
nB	0.44	0.02	0.38	0.02
NN	0.60	0.02	0.64	0.02
SVM	0.80	0.01	0.71	0.01
DT	0.77	0.01	0.76	0.01
RF	0.82	0.01	0.87	0.01

Table 3. Determination coefficient (R^2) of the cross-validation results between the *manual in-situ counting* and the *algorithm counting* (June 4, June 19 and June 4 + June 19) with grain yield (GY) using the linear regression (LR) for all data (n = 144) and also for data grouping by N fertilization (N.std, N+30 and N+50) are shown (n = 48). The results are also shown for the same data and dates with GY including the effects of G (all data + G) and N (all data + N) and also both combined (all data + G + N) factors (n = 144) using multiple linear regression (MLR). G by N interactions were not observed. The root mean square error (RMSE) was calculated for each model.

Input data	n = 144								n = 48					
	all data		all data + G		all data + N		all data + G + N		N.std		N+30		N-50	
	R^2	RMSE	**R^2**	RMSE	**R^2**	RMSE	**R^2**	RMSE	**R^2**	RMSE	**R^2**	RMSE	**R^2**	RMSE
Manual in-situ counting	**0.02**	806.0	**0.06**	807.5	**0.34**	664.8	**0.41**	631.0	**0.11**	733.6	**0.14**	735.1	**0.08**	442.4
Algorithm counting (June 4)	**0.04**	798.5	**0.11**	781.8	**0.34**	662.9	**0.45**	609.4	**0.14**	740.0	**0.20**	737.4	**0.17**	447.2
Algorithm counting (June 19)	**0.14**	766.2	**0.16**	766.2	**0.36**	764.8	**0.41**	630.1	**0.11**	752.3	**0.17**	732.2	**0.42**	376.2
Algorithm counting (June 4 + June 19)	**0.28**	737.7	**0.20**	745.5	**0.35**	656.6	**0.46**	606.8	**0.10**	766.0	**0.16**	727.9	**0.46**	368.5

FIGURE LEGENDS

Figure 1. Linear regression for *algorithm counting* vs. *manual image-based counting* on the same image using the RF classifier. Two dates of measurement with the whole data were used. The dotted line indicates the 1:1 slope. The root mean square error (RMSE) was calculated for each date.

Figure 2. A) Schematic overview of the field trial with the nine ground control points (GCPs), the four blocks division, plots delimited and the treatments applied; B) Zoom of one plot (central part indicated in red in A); C) overview of the field trial and D) average air temperature (°C) and cumulative precipitation (mm) for the entire growing period (November 1, 2017 – July 15, 2018) of the winter wheat trial.

Figure 3. Schematic representation of the image acquisition system and image-based validation. Each plot was divided in nine subplots. The central subplot (subplot #5) was selected for training and validation, to avoid possible errors due to the distortion or perspective. Red marks were placed manually on each ear. The dotted line rectangle shows a zoom-in of subplot #5 including the manual marks corresponding to the complete area covered by each single ear. These marks were used for training and classification purposes. For algorithm validation, a single dot per ear was used, as the final purpose of this work was to develop a methodology for wheat ear counting and not to determine their shape or size.

Figure 4. Image processing system proposed. The input is a subplot images in batch. Laplacian filter and Find Maxima are used for filtering and segmentation tasks respectively. Numerical characteristics such as shape, color and statistical measurement for each segmented area are calculated such that each row of the feature extraction matrix contained the features of each particular area detected per subplot image. Inside the dotted line rectangle, each matrix row was labelled automatically *Class1* or *Class2*. Areas in color were used for training and classifying purposes: (i) red: manual image-based ears; (ii) white (*Class1*: Ear): overlap between the areas automatically selected and the areas manually marked as ear; (iii) blue (*Class2*: Non-Ear): no overlap between the areas automatically selected and the areas manually marked as ear; these blue areas corresponding to soil, leaves and unwanted objects that were wrongly identified by ear by the automatic counting algorithm. The *m* features (columns) per *n* segmented areas (rows) were calculated to obtain the feature matrix. Feature selection reduced the dimensionality of the data and then the classifier was trained. The *m* features per each *n* area were used for training and classifying, and *s* features were selected. The same data from the feature extraction step was used to train each classifier. Classification accuracy was calculated using cross-validation.