Intuitive image analyzing on plant data-High throughput plant analysis with LemnaTec Image Processing

Stefan Paulus, Tino Dornbusch, Marcus Jansen

Pascalstr. 59, 52076 Aachen, Germany

Stefan.Paulus | Tino.Dornbusch | Marcus.Jansen@lemnatec.de

A paper from the Proceedings of the
13th International Conference on Precision Agriculture
July 31 – August 4, 2016
St. Louis, Missouri, USA

Abstract. *For digital plant phenotyping huge amounts of 2D images are acquired. This is known as one part of the phenotyping bottleneck. This bottleneck can be addressed by well-educated plant analysts, huge experience and an adapted analysis software. Automated tools that only cover specific parts of this analysis pipeline are provided. During the last years this could be changed by the image processing toolbox of LemnaTec GmbH. An automated and intuitive tool for the automated analysis of huge amounts of 2D data. Various image processing pipelines like edge detectors or background foreground separators are available as well as machine learning routines for more sophisticated problems. Segmentation of single plant parts is possible for plant images on different scales from microtiter plates, petri dishes, and single plants in the greenhouse or on field scale. Single modules can be attached to build an adapted analysis pipeline for a specific dataset and then repeatedly used for datasets of a similar plant. This enables the extraction of parameters like convex hull, height, and diameter or leaf area.*

For applications like the geometric parameterization of the complete plant, the classification of the ears from cereal field images using RGB cameras or 3D laser scans, or the segmentation of leaves by using hyperspectral images are possible in high throughput. Once created parameterization pipelines can be easy adapted to different plant species.

Two application scenarios using this software are described in detail within this publication. An automated running analysis pipeline for the parameterization of geometric plant parameters by using RGB photos is shown on greenhouse scale. This is based on an automated acquisition using

LemnaTec conveyor systems and an adapted measuring booth. Furthermore we show the localization of plant organs by using radiometric features on images coming from crane based measuring platform on field scale.

The image analysis software LemnaGrid (LemnaTec GmbH, Aachen) provides a professional tool that enables the intuitive connection of different image processing algorithms. It is adaptable for different plant types and on different scales. In this process the data processing can use different sensor data coming from RGB, 3D, hyperspectral or fluorescence imaging.

Keywords. *Image analysis, automated plant parameterization, plant phenotyping, growth monitoring*

Background:

Automated and non-invasive phenotyping of plants and other organisms relies on applying sensors with sensitivity to different ranges of the electromagnetic spectrum (Berger et al., 2010; Berger et al., 2012; Fiorani and Schurr, 2013; Großkinsky et al., 2015; Rahaman et al., 2015). Although use of visible light sensitive cameras is still most prevalent, sensors recording non-visible wavelengths, generating 3D point clouds or recording spectra are strongly gaining importance. In particular combining different sensors in sensor-fusion-approaches enables collecting physiological data together with structural information. Repeated measurements during the plants' lifetime deliver data with temporal resolution whereas the image-generating mode of action of the sensors yields spatially resolved information. For adequate use in phenotyping it is essential to translate the measured data into meaningful parameters that allow biological information. Generating biological knowledge out of collected data is a challenge that requires interactive work of information- and data-scientists together with biologists or agronomists (Großkinsky et al., 2015; Rahaman et al., 2015). To achieve such goals, it is essential to identify those feature within the captured data that link to biological properties of the measured organisms.

All externally measurable phenotypes originate from physiological activities (Großkinsky et al., 2015). The physiology itself is controlled by the genome as well as by influences from the environment, which either provide essential inputs for the physiology such as light or water, or create stress factors such as toxic substances or missing essential factors. As consequence, measuring phenotypic properties and their dynamic changes enables explaining gene functions and activities as well as responses to the environment. The environment interacts with the basic physiology by providing light, CO_2, water, and nutrients, but at the same time, stress can arise from improper levels of the needed supply or from unneeded factors that act on the organisms. Due to these complex interactions, measuring phenotypic properties and their changes requires careful interpretation of the measured data and linkage of measured parameters to biological processes.

In the first step of phenotypic measurements, data is collected with sensors of different mode of operation. Following steps analyze captured data with dedicated software procedures and derive parameters from the measurement that relate to biological properties of the plants (Dhondt et al., 2013; Lobet et al., 2013; Klukas et al., 2014). Many of such sensors are cameras and related instruments that give image-based representation of the data and thereby allow spatially resolved

interpretation. Cameras capture reflected or fluorescent light in visible wavelength ranges and beyond, comprising infrared and near infrared. Depending on their sensor-chip properties, they

collect data amounts between 1 and 20 MB per exposure. In the simplest case, reflectance of visible wavelengths refers to size and shape of the plants, and image processing delivers data that can be interpreted as growth processes and orientation of e.g. leaves or branches in the space. Growth processes are strongly governed by the genomic growth program, but they are changing upon environmental constraints. Visible light reflectance can convey color information, too, and this can relate to physiological states of the plants originating from developmental processes or environmental stress, for instance senescence that either can be part of a developmental program or response to a stress event. An alternative technique to capture information on size and shape of the plants is generating a 3D point cloud with a laser scanner that typically records 300 MB per run. In particular, this method gives access to single organs and structural details of the plant that frequently are not accessible from processing RGB-images. Laser scanners facilitate studying single leaves or branches, give access to leaf or branch angles and allow measuring surface structures such as leaf hairs. Other camera types capture non-visible wavelength ranges, e.g. infrared radiation, that relates to surface heat emission and can be used to assess plant and canopy temperatures. Such assessments can provide information related to plant transpiration processes. Combinations of cameras, illumination and filters, are sensitive to fluorescent light emitted from molecules in the plants after excitation with appropriate wavelengths. A common example is the auto-fluorescence of chlorophyll that can be used for assessments of photosynthetic parameters when using chlorophyll fluorometers. Spectral imaging collects reflectance information for a range of separately measured wavelengths resulting in data cubes consisting of projections of the objects stacked by different wavelengths. Several GB of data are recorded when using hyperspectral imaging setups. Spectra recorded with these instruments carry information on plant constituents, comprising pigments, structural substances, or water content. They provide static or dynamic information on the physiology of the plant or the canopy. By selecting single wavelengths out of such spectra, many vegetation indices can be calculated.

Depending on type and operation mode of the sensors, and how many sensors are combined in one experimental setup, large amounts of data come together in one event of data acquisition. Usually, experimenters take data on many samples subsequently and repeat this in time series so that data amounts rapidly increase. As mentioned initially, recording data is only the first step and a clever transfer of sensor data into plant information is the core of sensor-based plant phenotyping. Data analysis and processing however increasingly provides a bottleneck in the workflow. The more sophisticated sensors and their combination can operate, the more data demand for proper processing. Methods for data storage, access, processing and classification have to keep pace with data collection in order to provide comprehensive datasets on phenotypic properties of the measured plants.

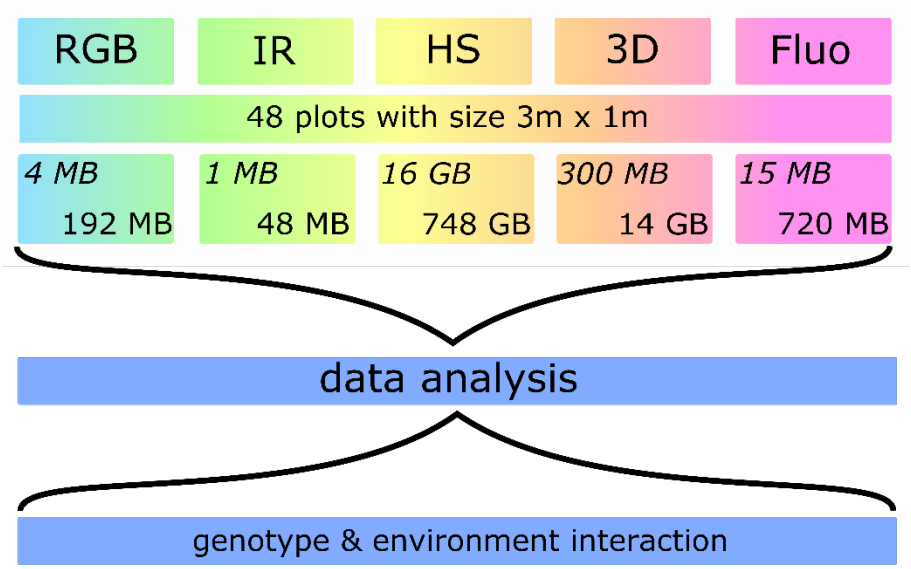

Figure 1: Phenotyping Sensor Bottleneck, data coming from the LemnaTec Scanalyzer Field multi sensor system. Coming from different sensors from a 100x10 meter field about 800 gigabytes of data are recorded.

Material and Methods:

Two experiments have been conducted to show the analysis opportunities given by the LemnaTec software. The first experiment was conducted within a green house where the plants were transported to the sensor, the second one was conducted using a field sensor platform where the sensor moves automatically over the field to specific plots.

In a first experiment using a 3D Scanalyzer, a conveyor belt based system where the plants were stored in a greenhouse, transported to different cabins for imaging, watering and spraying, the analysis of tomato plants is shown. Ten plants separated into two groups for two different treatments, one watered with 400 ml water every day and the other one with 600 ml per day. These plants were transported into the RGB imaging booth and four different images in accordance to four rotations of 90° were taken. These images were analyzed by counting the green and non-green (mostly yellowish) pixels within the plant and by taking into account only pixels belonging to the plant. The measurements were taken every day for 33 consecutive days, resulting in 10x4x33 = 1320 images that have to be analyzed. These four images for five plants of every group were averaged (mean) to result in one measurement per day and per group. The image analysis was performed using LemnaGrid the LemnaTec image processing solution. An analysis pipeline using visual programming was adapted to the analysis problem (see Figure 2). The image processing can be adapted within the single modules. By this, the creation of a complex image analysis pipeline, applicable for hundreds of images has been made easy, repeatable, efficient and comparable.

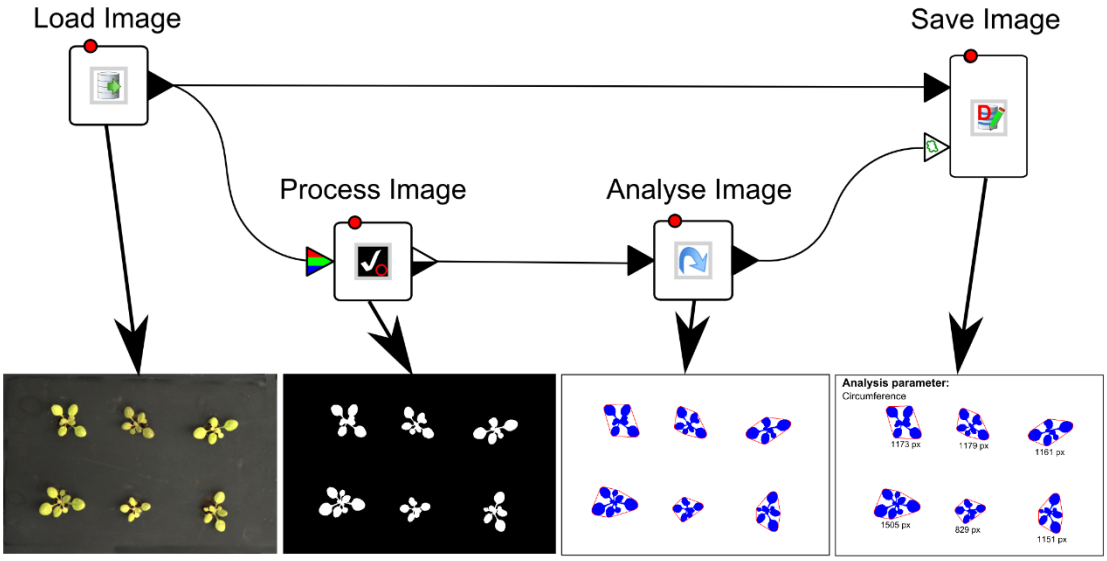

Figure 2: The LemnaTec Image Analysis Pipeline shows the connection between different image processing steps to a complete image processing pipeline. It provides parameters that can directly be used for the biological interpretation.

In a second experiment the new LemnaTec Field Scanalyzer, a rail-mounted multi-sensor platform that is able to collect data on a field of the size 10 m x 100 m was used. This sensor platform contains RGB cameras, hyperspectral-, and fluorescence imaging units, as well as 3D laser scanner and environmental sensors. Inside the area monitored by the Field Scanalyzer there were 72 test plots of 3 m x 1 m dimension arranged in 12 rows and 6 columns. Plots were sown with six wheat cultivars (Avalon, Cadenza, Crusoe, Gatsby, Soissons, and Widgeon) that were fertilized at four different Nitrogen levels (0, 100, 200, or 350 kg ha-1). Hyperspectral data cubes were recorded with Headwall Inspector VNIR and Headwall Inspector EXVNIR cameras mounted on the Field Scanalyzer and moved across the plots with the gantry crane positioning system. Data cubes were processes and specific wavelength bands were extracted from the data cubes in order to calculate vegetation indices.

Both experiments were analyzed using the LemnaGrid image processing software, a graphical programming interface. Single processing steps were visualized by processing nodes (so called devices). The connection results in a complete processing pipeline that can be exemplary adapted for best results using a representative sample dataset. This can be applied to the complete dataset independent from the data origin, greenhouse or field. Comparable methods and procedures are highly demanded as results of experiments not only depend on the analyzed plant material and the environment, but are strongly influenced by handling and experimental conditions, too (Massonnet et al., 2010; Junker et al., 2015; Parent et al., 2015).

Results:

In a greenhouse experiment plant sizes of tomato plants at two levels of irrigation were recorded during 35 days. Sizes were calculated as averages of pixels classified as belonging to the plant body out of images acquired from four horizontal views of each plant and mean values were calculated for five plants per irrigation group. Stronger irrigation with 600 ml water per day resulted in larger plants compared to an irrigation with 400 ml per day. Already from day 6 on, the plants at higher irrigation were larger that the less irrigated plants and reached nearly 1.5 times the size of the less irrigated group at the end of the experiment. However, in the stronger irrigated group a substantial fraction of non-green plant parts occurred, which became obvious from day 20 onwards and caused even a

decline of green area during days 22 to 25. At the end of the measurement period, about one quarter of the plant area of the stronger irrigated group was not green, whereas the fraction of not green parts in the less irrigated group was about 10%.

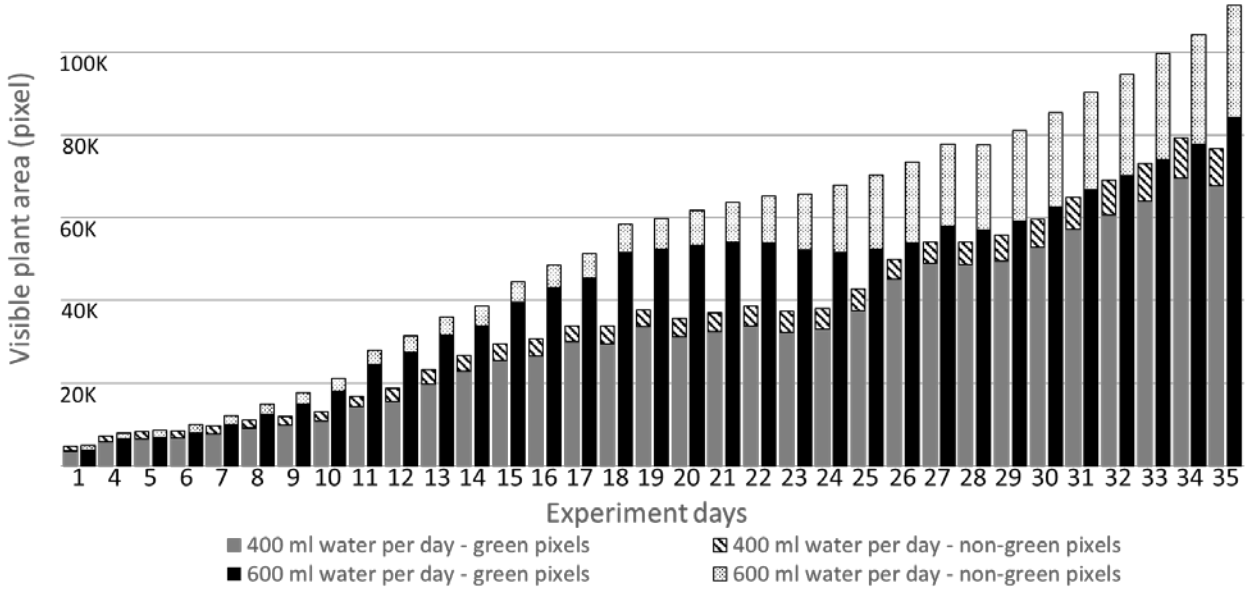

Figure 3: Measured sizes of tomato plants at 400 ml water per day (grey and hatched bars) and 600 ml water per day (black and dotted bars). Size calculated as average amount of pixels of four different viewpoints from five plants per treatment. Non-green pixels (shown as hatched or dotted parts of the bars) refer to damaged parts of the plants.

In a field experiment, plots with wheat plants were monitored with a hyperspectral camera and data were processed for NDVI values. Agricultural vegetation, particularly closes canopies, frequently are characterized by the use of vegetation indices. Among them, the NDVI – normalized difference vegetation index **(Rouse, 1973)** – is very common.

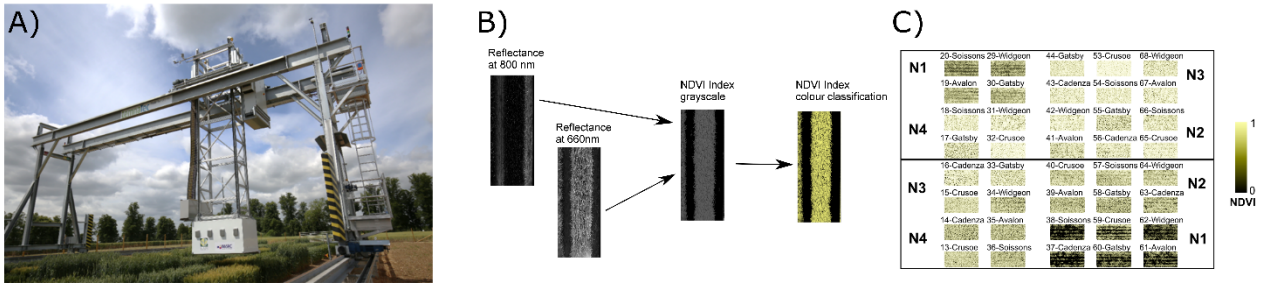

Figure 4: Data acquisition with Field Scanalyzer (A); data processing for NDVI representation (B); color-coded map of field plots, color-coding denotes NDVI values between 0 and 1 (C).

The index gives a ratio out of the canopy reflectance in the red and near-infrared wavelength range and we used NDVI= (R800-R680)/(R800/R680). The NDVI is used in remote sensing of vegetation and reported to correlate with chlorophyll content at leaf level and in the end with the biomass cover of the measured area. Different from classical air-borne remote sensing, Field Scanalyzer measurements are a proximal sensing approach with 2 m distance between sensor and vegetation. In a case study, a subset of 20 plots out of the Field Scanalyzer measuring area at Rothamsted Research, UK, was evaluated for cultivar- and fertilization- specific NDVI responses. Overall speaking, non-fertilized plots exhibited low NDVI values whereas fertilizer application increased NDVI

values in accordance to stronger growing crops (see Table 1).

Table 1: Mean NDVI data of wheat cultivars at different nitrogen fertilization levels during a growth period from 20th to 28th June 2015.

N (kg/ha)	Cadenza	Crusoe	Gatsby	Soissons	Widgeon
0	0.46	0.56	0.51	0.45	0.41
100	0.75	0.71	0.75	0.74	0.72
200	0.82	0.80	0.74	0.76	0.79
350	0.83	0.74	0.79	0.76	0.79

Providing 100 kg/ha of fertilizer substantially increased NDVI from values around 0.5 to values around 0.7 for all cultivars. Adding more fertilizer, 200 or 350 kg/ha, promoted values towards 0.8, whereas NDVI values as response to these two fertilizer levels did not differ substantially from each other. Besides this overall trend, cultivars responded differently. Cadenza and Widgeon had an increase in NDVI from 100 towards 200, but not from 200 to 350 kg/ha. In Gatsby, the increase in NDVI occurred between 200 and 350 kg/ha. For Crusoe, there was an increase in NDVI when raising fertilization from 100 to 200 kg/ha, but a decline when providing more fertilizer. Soissons had only minor changes in NDVI when elevating fertilizer from 100 over 200 to 350 kg/ha.

For all cultivars and fertilizer levels, there were fluctuations, but no trends for decrease or increase in NDVI values from day to day throughout the measuring period.

Discussion:

Phenotyping largely relies on non-invasive methods that enable analyzing properties of organisms without altering or influencing the organism as such and moreover allow repeated measurements on individuals at subsequent developmental stages (Dhondt et al., 2013). Thereby, sensors that operate with or without image-generating modes capture data on the organisms. Captured data usually needs dedicated processing to translate the measured signals into parameters that provide biological information (Rahaman et al., 2015). The transfer from measurement data to biologically interpretable data demands for phenotyping software that extracts relevant information out of the signals coming from the sensors. For instance, when taking photographs, the image processing pipelines of the phenotyping software have to separate the plant from the background by deciding for each pixel whether it belongs to the plant surface area or not. Having taken such decisions, the resulting pixel group can be analyzed further for size, geometry, or color. These information in turn brought together with other information on the measured plant and its growth environment provides a phenotypic description of the plant. Similar steps as exemplary described for photographic sensing apply when using sensors that record fluorescence signals, infrared radiation, spectral scans, or 3D point clouds. In the current case studies, photographs from greenhouse-cultivated tomato plants and hyperspectral data-cubes from field-grown wheat plants were analyzed with the phenotyping software LemnaGrid. Image processing revealed that tomato plants at higher irrigation level grew stronger than those at lower irrigation, but the fraction of non-green area, which refers to damaged or senescent plant parts, was larger. Although growth promoting, the higher irrigation level led to stress for the growing plants. In the second case study, processing hyperspectral data cubes for NDVI-values showed that the 100 kg/ha fertilization strongly improved crop growth, however increasing the fertilization to 200 kg/ha only led to minor improvements of canopy growth and 350 kg/ha gave only minor increase, but even could cause reduced growth as it appeared in Crusoe. There was a differential effect among the cultivars pointing at low responsiveness to fertilizer in Soissons. Cadenza and Widgeon were more responsive to fertilizer, but reached a plateau of responsiveness when applying the highest amout. In

Gatsby, ther was weak responsiveness to intermediate amounts of fertilizer, but a stronger response to the highest amount of fertilizer.

As the LemnaTec software package operates across different sensor platforms and application cases, it contributes to achieving comparability in phenotyping experiments. The possibility to implement analysis pipelines by graphical programming with pre-defined functions enables the users to establish re-usable data processing, which can be interchanged between sensor platforms at different locations. Broader speaking, comparability in experimental handling contributes to good practice and standards in phenotyping which is highly desired in the scientific community (Fiorani and Schurr, 2013).

The software shows a general approach for automated analysis of data independent of the origin. Data from laboratory or greenhouse can be processed as well as data from current field phenotyping systems using standard data formats. It provides processing of huge amounts of images by using one simple pipeline adaption step as preparatory work. To provide a generalized further processing a link to R, the statistical programming language (Ihaka and Gentleman, 1996), is given. This will give the user a great portfolio of possibilities and a huge database of ready-made scripts for statistical information.

Future development will show the integration of 3D data into the visual workflow and the connection of single 3D point cloud processing tools to complete analysis pipeline. This will work on greenhouse data as well as on field data. First experiments with data coming from the Scanalyzer Field System are planned and focus on the detection of single organs on field level and the evaluation of the new developed field laser scanners as well as on the generation of height and growth maps (Friedli et al., 2016). The integration of steps like the classification of single organs (Paulus et al., 2013) the automated parameterization of organs within time series (Paulus et al., 2014).

Conclusion:

Visual programming has shown to be an objective and adequate tool for high throughput analysis of huge amounts of plant data acquired with imaging systems. Independent of the used sensor like RGB cameras or hyperspectral devices the processing using intuitive connections between complex image processing subroutines can be performed by non-specialized scientist. The objective and repeatable analysis can easily be adapted to new parameters, experiments and plants. Sharing analysis pipelines and processing steps enable the comparison of experimental results and depict a milestone in publication of analysis algorithm and tool at the same time.

We showed the analysis of a tomato and wheat experiment using RGB cameras and hyperspectral devices. Differences among the treatments can be shown and quantified on greenhouse and field scale. On field scale we also showed the generation of maps showing the connection between nitrogen access and the NDVI reflectance.

An easy to handle and intuitive way of visual image processing was shown and evaluated within experiments on different scales. We conclude that this analysis software is ready for use within scientific and productive context. It will move the focus from image processing and recognition problems among big sets of data to the biological interpretation of extracted parameters and help to overcome the restrictions of the data bottleneck in phenotyping.

References:

Berger B, Parent B, Tester M (2010) High-throughput shoot imaging to study drought responses. Journal of Experimental Botany 61: 3519–3528

Berger B, de Regt B, Tester M (2012) High-Throughput Phenotyping of Plant Shoots. *In* J Normanly, ed, High-Throughput Phenotyping in Plants. Humana Press, Totowa, NJ, pp 9–20
Dhondt S, Wuyts N, Inzé D (2013) Cell to whole-plant phenotyping: the best is yet to come. Trends in Plant Science 18: 428–439
Fiorani F, Schurr U (2013) Future Scenarios for Plant Phenotyping. Annual Review of Plant Biology 64: 267–291
Friedli M, Kirchgessner N, Grieder C, Liebisch F, Mannale M, Walter A (2016) Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot crop species under field conditions. Plant Methods. doi: 10.1186/s13007-016-0109-7
Großkinsky DK, Svensgaard J, Christensen S, Roitsch T (2015) Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. Journal of Experimental Botany erv345
Ihaka R, Gentleman R (1996) R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5: 299–314
Junker A, Muraya MM, Weigelt-Fischer K, Arana-Ceballos F, Klukas C, Melchinger AE, Meyer RC, Riewe D, Altmann T (2015) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. Frontiers in Plant Science. doi: 10.3389/fpls.2014.00770
Klukas C, Chen D, Pape J-M (2014) Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping. Plant Physiology 165: 506–518
Lobet G, Draye X, Périlleux C (2013) An online database for plant image analysis software tools. Plant Methods 9: 38
Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisec J, Beemster GTS, Meyer RC, Messerli G, Gronlund JT, et al (2010) Probing the Reproducibility of Leaf Growth and Molecular Phenotypes: A Comparison of Three Arabidopsis Accessions Cultivated in Ten Laboratories. Plant Physiology 152: 2142–2157
Parent B, Shahinnia F, Maphosa L, Berger B, Rabie H, Chalmers K, Kovalchuk A, Langridge P, Fleury D (2015) Combining field performance with controlled environment plant imaging to identify the genetic control of growth and transpiration underlying yield response to water-deficit stress in wheat. Journal of Experimental Botany. doi: 10.1093/jxb/erv320
Paulus S, Dupuis J, Mahlein A-K, Kuhlmann H (2013) Surface feature based classification of plant organs from 3D laserscanned point clouds for plant phenotyping. BMC bioinformatics 14: 238
Paulus S, Dupuis J, Riedel S, Kuhlmann H (2014) Automated Analysis of Barley Organs Using 3D Laser Scanning: An Approach for High Throughput Phenotyping. Sensors 14: 12670–12686
Rahaman MM, Chen D, Gillani Z, Klukas C, Chen M (2015) Advanced phenotyping and phenotype data analysis for the study of plant growth and development. Frontiers in Plant Science. doi: 10.3389/fpls.2015.00619
Rouse JW (1973) Monitoring Vegetation Systems in the Great Plains with ERTS. Third ERTS Symposium, NASA SP-351. pp 309–317