Pattern Recognition Letters 81 (2016) 80–89

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Finely-grained annotated datasets for image-based plant phenotyping☆

Massimo Minervini[a,*], Andreas Fischbach[b], Hanno Scharr[b], Sotirios A. Tsaftaris[a,c]

[a] Pattern Recognition and Image Analysis Research Unit, IMT Institute for Advanced Studies, 55100 Lucca, Italy
[b] Institute of Bio- and Geosciences: Plant Sciences (IBG-2), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
[c] Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh EH9 3JL, UK

ARTICLE INFO

Article history:
Available online 12 November 2015

Keywords:
Image processing
Machine vision and scene understanding
Plant biology
Annotated datasets

ABSTRACT

Image-based approaches to plant phenotyping are gaining momentum providing fertile ground for several interesting vision tasks where fine-grained categorization is necessary, such as leaf segmentation among a variety of cultivars, and cultivar (or mutant) identification. However, benchmark data focusing on typical imaging situations and vision tasks are still lacking, making it difficult to compare existing methodologies. This paper describes a collection of benchmark datasets of raw and annotated top-view color images of rosette plants. We briefly describe plant material, imaging setup and procedures for different experiments: one with various cultivars of Arabidopsis and one with tobacco undergoing different treatments. We proceed to define a set of computer vision and classification tasks and provide accompanying datasets and annotations based on our raw data. We describe the annotation process performed by experts and discuss appropriate evaluation criteria. We also offer exemplary use cases and results on some tasks obtained with parts of these data. We hope with the release of this rigorous dataset collection to invigorate the development of algorithms in the context of plant phenotyping but also provide new interesting datasets for the general computer vision community to experiment on. Data are publicly available at http://www.plant-phenotyping.org/datasets.

1. Introduction

The study of the phenotype expressed by cultivars (or mutants) of the same plant species under different environmental conditions is central to our understanding of plant function. Identifying and evaluating a plant's actual phenotype, is relevant to, e.g., seed production and plant breeders. The phenotype relies on the fine-grained categorization of a plant's properties: e.g., how many leaves, of which architecture, visual age or maturity level, to which cultivar a plant is similar.

Previously, such categorization was annotated manually by experts, but recently image-based approaches are gaining momentum. In the last decades several approaches have been proposed to measure visual traits of plants in an automated fashion [22,29,43,58] together with customized image processing pipelines to analyze aspects of the acquired image data [2,11,23,26,42,58,62]. However, several of the computer vision tasks encountered (for example leaf segmentation and counting) are particularly challenging and remain unsolved. In fact, most experts now agree that lack of reliable and automated algorithms to analyze these vast datasets forms a new bottleneck in our understanding of plant biology and function [41]. We must accelerate the development and deployment of such computer vision algorithms, since according to the Food and Agriculture Organization of the United Nations (FAO), large scale experiments in plant phenotyping are a key factor in meeting agricultural needs of the future, one of which is feeding 11 billion people by 2050.

One of the factors that could accelerate the development of better algorithms and their consistent and systematic evaluation is the availability of benchmark data focusing on typical imaging situations and tasks in plant phenotyping. Till now, despite the 20 year history of imaging plants, a comprehensive collection of benchmark datasets for image-based non-destructive plant phenotyping is still lacking. While plant related datasets exist for leaf or flower recognition [20,46,54,61], these datasets were obtained in an uncontrolled or destructive manner and not in a phenotyping context.

Here, we present a collection of raw and annotated images of the most frequently used rosette model plants (Arabidopsis and tobacco). We briefly describe plant material, environmental conditions, and imaging setup and procedures, which led to the collection of color images showing top-down views. Shape variability in these images is high, cmp. Fig. 1.

They show compact plants, plants with long petioles, overlapping and non-overlapping leaves, different leaf numbers, etc., such that a wide range of shapes typically found in young dicotyledon plants is

☆ This paper has been recommended for acceptance by Concetto Spampinato.

* Corresponding author.
E-mail address: m.minervini@imtlucca.it, minervini.massimo@gmail.com (M. Minervini).

http://dx.doi.org/10.1016/j.patrec.2015.10.013
0167-8655/

Fig. 1. Closeups of examples plants. Left column: young plants with few leaves. Middle: rosettes with elongated petioles (Arabidopsis). Right: compact rosettes with or without rich structure within single leaves (tobacco).

covered (cmp. e.g. growth stages '1.' or '30' of oilseed rape, sunflower, or beet in [35]). Using these raw images we describe a collection of datasets, with appropriate expert annotations and metadata, for a series of computer vision tasks. We emphasize tasks such as plant/leaf detection, segmentation, and tracking, leaf counting, boundary estimation, and general regression and classification. Fig. 2 shows examples of such annotations. For consistency in evaluation methodologies we also discuss appropriate criteria. To offer exemplary use cases we also show results on plant and leaf segmentation and counting. While not all raw images have been annotated, our annotation is continuing and the datasets size is increasing. We devised and released an annotation tool [39] in order to facilitate future annotation. This article will serve as the reference point for describing the process and tasks for which the data can be used.

This article is organized as follows: Section 2 presents data collection; Section 3 presents datasets and annotations for a variety of vision tasks; Section 4 outlines exemplary use cases and results; and Section 5 offers discussion and conclusions.

2. Imaging setup and data

Non-invasive plant investigations are performed using various modalities with spatial scales varying from the microscopic to large

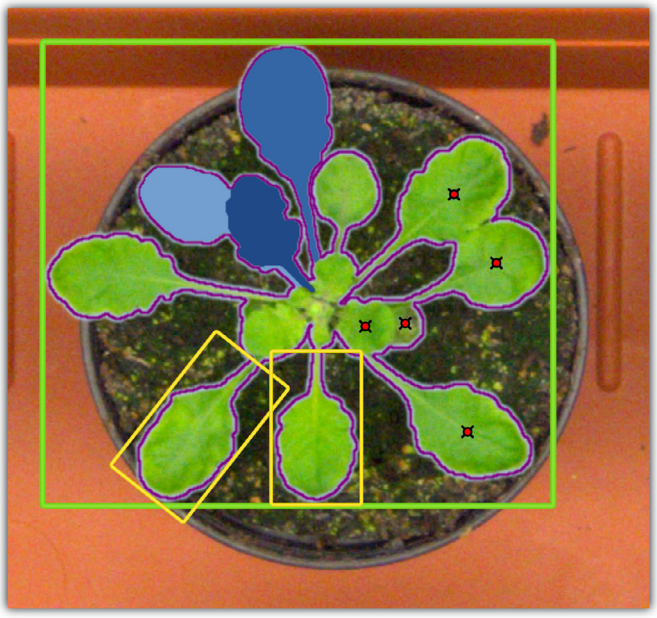


Fig. 2. Example annotations superimposed on an image of Arabidopsis (Col-0, wild-type), acquired 17 days after germination. Shown are: the plant bounding box (green), leaf bounding boxes (yellow), the plant mask (outlined in purple), three leaf masks (blue overlays), and five leaf centers (red dots) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

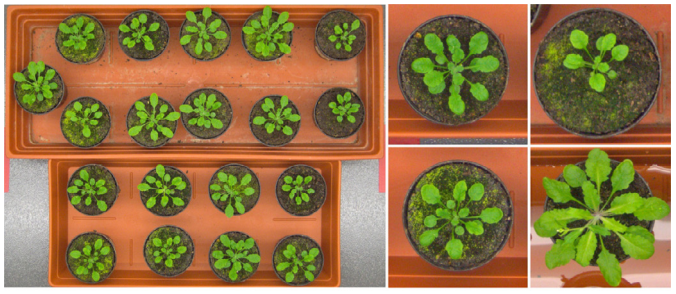
outdoor fields [4,7,42]. Typical problems in measuring a plant's visible properties comprise measuring size, shape, color or spectral reflection, and other structural and functional traits of whole plants, their organs, or plant populations. Plants are not static, but self-changing systems with complexity in shape and appearance increasing over time. For example, a single leaf appears in the scene and can grow several cm^2, i.e. orders of magnitude change in size till it stops growing. Leaves can grow rapidly, in the range of hours, with the whole plant changing over days or even months, in which the surrounding environmental (as well as measurement) conditions may also change. Biologists grow model plants, such as Arabidopsis (*Arabidopsis thaliana*) and tobacco (*Nicotiana tabacum*), in controlled environments and monitor and record the phenotype. Such experiments are fundamental and ubiquitous, and recovering the phenotype implies that computer vision algorithms must deal with the complexity of the plant, the experiment and the environmental conditions.

Starting from such typical experiments, we devised imaging apparatuses and setups to acquire three imaging datasets (Fig. 3), which form the basis of this work. They were acquired in two different labs with highly diverse equipment, as shown in Figs. 4 and 5. Arabidopsis images were acquired using two setups for investigating affordable hardware for plant phenotyping. Tobacco images were acquired using a robotic imaging system for the investigation of automated plant treatment optimization by an artificial cognitive system. Images differ in resolution, fidelity and scene complexity, with plants appearing in isolation or in trays, and with subjects belonging to different cultivars (mutants), or undergoing different treatments (see examples in Fig. 3). The vision tasks required to estimate several phenotyping parameters include detection, segmentation, and tracking across time, for plants and individual leaves.

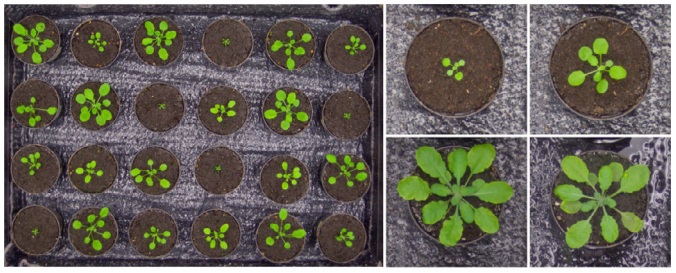
Prior to discussing the datasets (Section 3) and the computer vision tasks they can be used for (Section 4), we briefly describe below how plants were grown and imaged, with additional details available in [52].

Table 1
Summary of information of the Arabidopsis and tobacco raw imaging data.

Experiment	Subjects	Wild-types	Mutants	Period	Total images	Image resolution (MPixel)	Field of view	Plant resolution
Arabidopsis (Ara2012)	19	Col-0	No	3 weeks	150	7	large	0.25 MPixel
Arabidopsis (Ara2013, Canon)	24	Col-0	Yes (4)	7 weeks	4186	7	large	0.25 MPixel
Arabidopsis (Ara2013, Rasp. Pi)	24	Col-0	Yes (4)	7 weeks	1951	5	large	0.06 MPixel
Tobacco (23.01.2012)	20	Samsun	No	18 days	34,560	5	single	same
Tobacco (16.02.2012)	20	Samsun	No	20 days	38,400	5	single	same
Tobacco (15.05.2012)	20	Samsun	No	18 days	34,560	5	single	same
Tobacco (10.08.2012)	20	Samsun	No	30 days	57,600	5	single	same



(a) Ara2012



(b) Ara2013 (Canon)

(c) Tobacco

Fig. 3. Example raw images (left) and close-ups showing individual plant subjects (right) of: (a) Arabidopsis (Col-0, wild-type), (b) 5 different cultivars of Arabidopsis, and (c) tobacco undergoing different treatments. (a) and (b) Arabidopsis and (c) tobacco were imaged with the setups shown in Figs. 4 and 5, respectively.

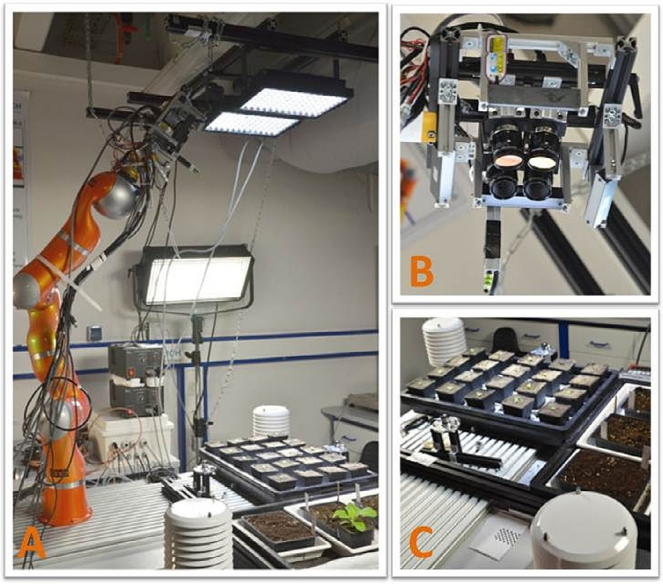


Fig. 5. Hardware setup of the GARNICS robot gardener. (A) The robot arm, in the lab environment. (B) Camera head with illumination and watering system. (C) Workspace with light, temperature, and humidity sensors.

2.1. Arabidopsis imaging with an affordable setting

Arabidopsis images were acquired in two data collections: in June 2012 and in September-October 2013, hereafter named *Ara2012* and *Ara2013*, respectively, both consisting of top-view time-lapse images of *A. thaliana* rosettes. Experiment composition is summarized in Table 1. Plants were grown in individual pots with randomized arrangement to eliminate possible bias in the results due to variations in watering or lighting conditions. No treatments were performed.

The imaging setup (cf. Fig. 4) consisted of a growth shelf and an automated affordable sensing system [56], to acquire and send images of the scene via a wireless connection to a receiving computer

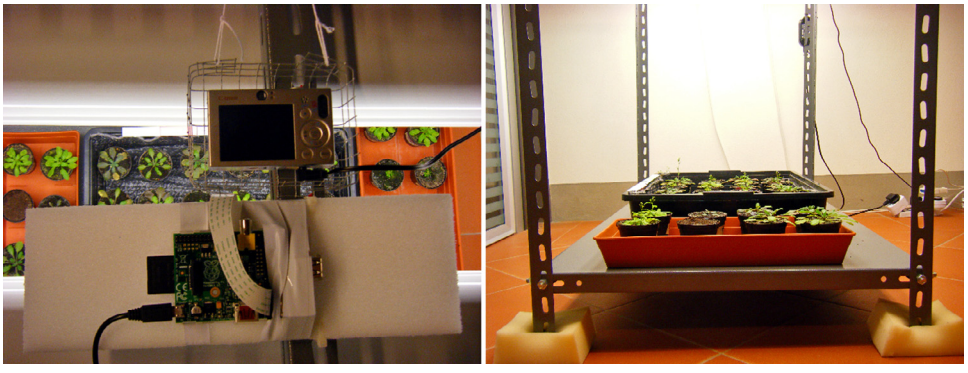


Fig. 4. Affordable acquisition setup for 'Ara2012' and 'Ara2013'.

(http://www.phenotiki.com). Example images are shown in Fig. 3a–b, illustrating plant arrangement and scene complexity.

Plants were illuminated artificially and controllably to emulate daylight imaging in a fixed day cycle. Camera sensors were positioned between the lights, approximately 1 m above the plants, and operated as intervalometers, i.e. acquiring images at preset times with preset imaging conditions (such as focus, exposure, field of view). Images were captured during day time only for a period of time (cf. Table 1). Two cameras were used: a 7 Mpixel consumer grade camera (Canon PowerShot SD1000), shorthanded as *Canon*, and an even lower cost system based on the Raspberry Pi (http://www.raspberrypi.org), Model B, with the Raspberry Pi 5 Mpixel camera module, shorthanded as *Rpi*. Ara2012 used only Canon but Ara2013 used both cameras, permitting the evaluation of algorithms on images of the same scene but with different image quality. Additionally for Ara2013, multiple focus images were acquired. All acquired images (width × height for Canon: 3108 × 2324 pixels; for Rpi: 2592 × 1944 pixels) were stored as lossless PNG files.

2.2. Tobacco image dataset

The tobacco images were acquired in the context of the European project 'Gardening with a cognitive system' (GARNICS, http://www.garnics.eu). The GARNICS project aimed at 3D sensing of plant growth and building perceptual representations for learning the links to actions of a robot gardener. This robot must deal with the inherent complexity of plants, which changes over time. Actions performed at plants (like watering), will have strongly delayed effects, making monitoring and controlling plants a difficult perception-action problem.

2.2.1. Imaging setup

The setup allowed for looking at plants from different poses (see Fig. 5). We release top views of the plants only, as top view images are used in many plant screening applications. Notice, however, that a setup performing only this type of acquisition can be considerably simpler and more affordable than the setup used here. Beneath the robot and environment sensors the implemented system featured: (a) a watering and nutrient solution dispensing system for treatment application; (b) high power white LED illumination (switched off for imaging); and (c) low power white fluorescence illumination used for imaging.

The robot head consisted of two stereo camera systems (4 × Point-Grey Grashopper, 2448 × 2048, pixel size 3.45 μm), black-and-white and color, and high quality lenses (Schneider Kreuznach Xenoplan 1.4/17-0903). We added lightweight white and NIR LED light sources to the camera head. Using this setup, each plant was imaged separately from different but fixed poses. In addition, for each pose small baseline stereo image pairs were captured using each single camera by a suitable robot movement, allowing for 3D reconstruction of the plant.

Data released here stems from experiments aiming at acquiring training data for the robot gardener. Images were acquired every hour in a 24/7 manner for up to 30 days. More details on the four experiments can be found in Table 1. Overview images of the plants from these experiments are shown in Fig. 3c.

2.2.2. Plant material and growing conditions

Tobacco plants (*N. tabacum cv. Samsun*) were grown in 7 × 7 cm pots under constant light conditions with a 16 h/8 h day/night rhythm. Water was provided by the robot system every 2 h. Nutrients were applied either manually twice a week or every other hour by the robot system. In GARNICS treatments were selected to produce training data for a cognitive system. The amounts of water and nutrient solution are therefore well adapted to the soil substrate such that the plants show distinguishable performance of generally well growing plants. Finding an optimal treatment was left to the system.

Images show growth stages from germination well into the leaf development stage, i.e. starting observations at growth stage 09 and stopping at stage 1009 to 1013 (according to the extended BBCH-scale presented in [9]), due to size restrictions. Fig. 3c shows images of the final growth stages. For further details on treatments, environmental conditions, and acquisition times we refer to [52].

3. Annotated datasets

In order to provide specialized standalone datasets for a number of computer vision tasks, parts of the raw image data described in Section 2 have been carefully annotated by experts. Note that not all data have been annotated (thousands of raw images are available) and annotation is a continuing process. However, even augmented versions of the datasets in the future will follow the structure described here. Among those that have been annotated, not all are publicly available, to permit future competitions and challenges based on data which are blind to participants. Other computer vision competitions, such as PASCAL VOC [15], and biologically inspired challenges follow this strategy. In addition, for each specific task the datasets are considered standalone and as such the user is unaware of other information (e.g., mutant type for leaf segmentation). This limits domain and prior knowledge and should lead to methodologies that are more robust to the changing and complex morphology of plants.

In the following paragraphs we describe first the semantic hierarchy considered, the manual annotation procedure, and then proceed in detailing available subsets of the data for each computer vision and machine learning task, defining annotations and appropriate metadata, and suitable evaluation criteria. Dataset size refers to the current state of annotation as to the current date.

3.1. Overview of semantic hierarchy

Each experiment has generated a vast amount of imaging data with Arabidopsis experiments showing tray images whilst tobacco individual plants. Our internal database and annotation strategy follows the hierarchy visible in Fig. 6. These original images are higher in our semantic hierarchy.

Gray boxes in Fig. 6 denote related annotated metadata: experiment type, mutant type, camera used, acquisition time, experimental treatment, segmentation difficulty, etc. Non-shaded boxes denote imaging and image level annotations. Note that an experiment may contain both tray and individual plant images such as Arabidopsis for example. However, this is not a rule: for example, for tobacco datasets tray images are not available, and for Arabidopsis experiments no treatment was performed. To construct each of the standalone datasets described below, we trace information in this hierarchy and provide related metadata and annotations wherever appropriate.

3.2. Expert segmentations

A significant number of object-based annotations, e.g., bounding boxes, can be obtained computationally on the basis of pixel-level segmentation masks of plants and leaves, respectively, which have been manually annotated by experts. Here we describe how we obtained the latter and next we detail the level of annotation for each task.

Annotation consisted of three steps. First, we obtained a binary segmentation of the plant objects in the scene in a computer-aided fashion. For Arabidopsis, we used the approach based on active contours described by Minervini et al. [38], while for tobacco, a simple color-based approach for plant segmentation was used. The result of this segmentation was manually refined using raster graphics editing

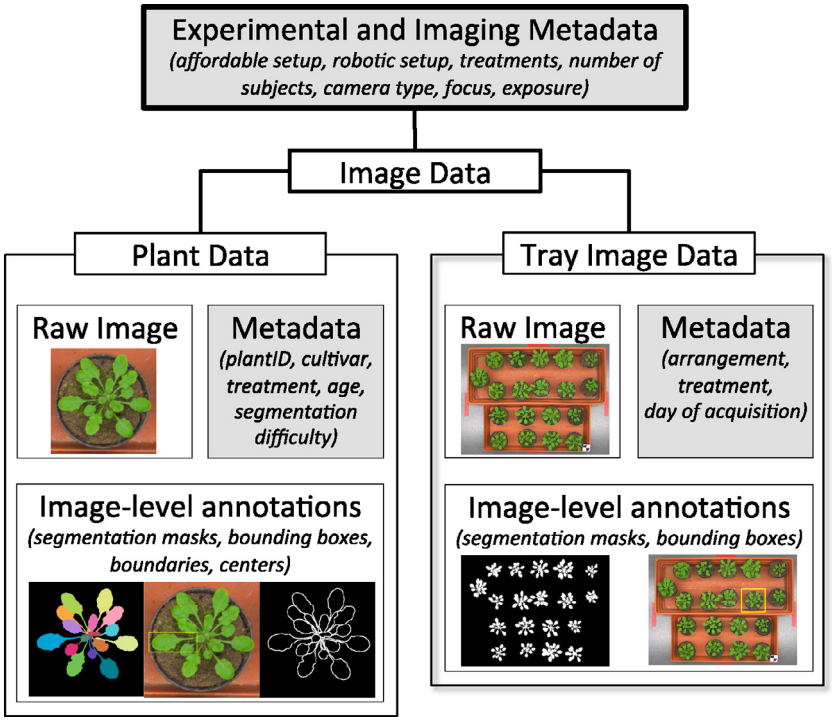


Fig. 6. Hierarchy of relationships among data, metadata, and annotations. In parentheses we provide examples of annotation variables, and we also provide pictorial examples of imaging data and annotations such as segmentation masks, bounding boxes, and leaf boundaries. Gray boxes denote metadata.

software, to ensure that all the visible part of the shoot is included in the plant mask and that the background (earth, moss, etc.) is excluded. Next, within the binary mask of each plant, we delineated individual leaves (including both the petiole and the blade) completely manually. A pixel with black color denotes background, while all other colors are used to uniquely identify leaves of the plants in the scene. Across the frames of the time-lapse sequence, we consistently used the same color code to label occurrences of the same leaf. To reduce observer variability and increase accuracy, the labeling process involved always two annotators: one annotating the dataset and one inspecting the other. For future extensions of the datasets, the annotation of additional images is supported by a tool that we recently released for semi-automated leaf segmentation and annotation [39]. Fig. 7 shows examples of plant images from the datasets, with corresponding pixel level annotation masks.

On a secondary inspection of the data, additional categorical *qualitative annotations* were recorded by annotators such as: estimate of segmentation difficulty (in the 1[easy]–5[hard] scale), plant appears in focus, leaves appear in vertical positions which is typical in tobacco (due to the so called nastic movements), plant is occluded by another one (when pots are placed close by), and scene contains complexities (water in the background, green moss on soil, debris or damage on leaves).

3.3. Computer vision tasks and datasets

Fine-grained information to be extracted from images is ubiquitous in plant phenotyping, since we do have to deal with how different mutants or treatments affect plant shape and characteristics. Even the same plant will have leaves of different shape and size according to their maturity, as it can also be seen in Fig. 7. In some cases the phenotype of a mutant is not known, and typically researchers assign qualitative characteristics, which is simple when gross phenotype differences are evident (e.g., radically different leaf shape). However, it is when such differences are subtle that the ability to extract fine-grained information directly from images would make a tremendous impact: it would permit biologists to identify small traits to be explored further. Computer vision and machine learning could enable biologists to assess a new mutant's phenotype by evaluating how similar it is to known lines of cultivars and mutants in a quantitative fashion.

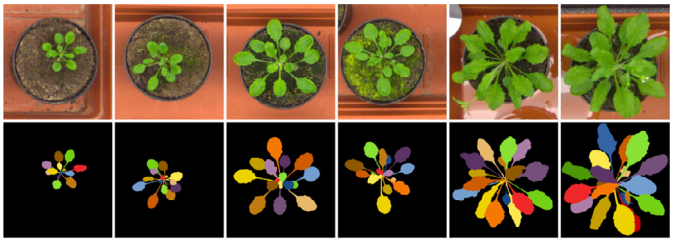
Doing this in a fully automated fashion is fertile ground for a series of interesting vision tasks, which we outline below together with descriptions of appropriate for the task datasets and evaluation criteria. To test and develop robust solutions for a realistic range of conditions, we constructed the datasets by carefully choosing images according to our qualitative annotations (Section 3.2), and also allowing several challenging situations to occur by design in the experiments. For the evaluation criteria we took inspiration from standard criteria used in computer vision and also devised some that are relevant to phenotyping. Since each task is different and the number of available annotated images is increasing, we refrain from supposing a certain split into training, validation and testing sets –a procedure typically adopted when evaluating learning-based approaches. However, our recent experience in organizing the leaf segmentation (Section 3.3.3) and leaf counting (Section 3.3.5) contests (see Section 4) using a subset of the annotated data presented here showed that the number of available images and criteria used allowed computer vision scientists to work successfully on these problems (see e.g. [19,48,49,53]).

Our annotations were initialized by computationally post-processing the expert-delineated leaf masks as discussed in Section 3.2. Afterwards, everything was verified (and if necessary corrected) by experts to ensure high quality and integrity. Note that for each task we outline the size (as in current number of annotated images) of the corresponding dataset.

Hereafter, to ease presentation we will denote as *leaf mask* the pixel-level leaf masks (including the visible part of a leaf blade and the corresponding petiole) and as *plant mask* the pixel-level binary mask obtained by the union of all individual leaf labels within a plant. To reduce storage all images are lossless compressed PNG files.

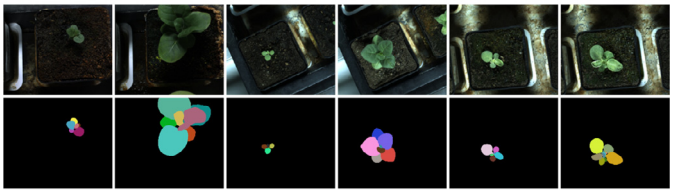
3.3.1. Plant detection and localization

Task: As Fig. 3 shows, plants can be arranged in a grid, either fixed in position in specialized trays or not and can even be touching each



(a) Ara2012

(b) Ara2013 (Canon)



(c) Tobacco

Fig. 7. Examples of single plant images at different developmental stages with the corresponding ground truth leaf labeling denoted by color. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

other, the latter leading to a non trivial detection problem. Thus, a first vision task encountered is plant detection and localization of individual plants, in the form of bounding boxes. Excitingly, this problem is also encountered even outside the context of phenotyping, e.g., in precision agriculture for detecting crops [28] albeit in a more complicated setting. Here learning-based object detection approaches are ideally suited to help.

Dataset: We derived bounding boxes for each plant as the smallest bounding box enclosing the plant mask, processing each plant individually in a tray image. An additional 5% of the box size was considered to include a larger aspect of the scene. The dataset consists of 70 tray images, as 16 from Ara2012, 27 from Ara2013 (Canon), and 27 from Ara2013 (Rpi). For each tray image, a comma-separated value (CSV) file is available, reporting for each plant the corner pixel coordinates of its bounding box.

Evaluation criteria: Following upon the ubiquitous PASCAL VOC challenge [15], we suggest the bounding box overlap ratio criterion $a_o = \text{area}\,(B^p \cap B^{gt})\,/\,\text{area}\,(B^p \cup B^{gt})$, between predicted B^p and ground truth B^{gt} bounding boxes.

3.3.2. Plant segmentation

Task: Plant biomass is an important plant breeding trait because it reflects overall plant performance. In images of rosette plants usually it is measured as projected leaf area (PLA), i.e. effectively the number of plant pixels. Finding PLA translates to the segmentation of plant from background. In simple cases this can be solved by color thresholding and other unsupervised segmentation approaches [11,58], but when scene complexity is high (non-smooth background, non-uniform lighting, plant overlap, presence of moss on soil) sophisticated learning-based algorithms are necessary [38,50].

Dataset: The dataset consists of 16 tray images from Ara2012, 27 from Ara2013 (Canon), and 27 from Ara2013 (Rpi). For each image a corresponding black (background) and white (foreground) mask encoded as an indexed image provides pixel-level information on the location of plant objects.

Evaluation criteria: Several segmentation criteria exist and we suggest Dice coefficient, precision, and recall, since they have been used throughout image analysis and are common in plant imaging as well [38]. Among those, the Dice Similarity Coefficient, DSC (%) $= 2|P^{gt} \cap P^{ar}|/(|P^{gt}| + |P^{ar}|)$, measures the degree of overlap among ground truth P^{gt} and algorithmic result P^{ar} binary segmentation masks, where $|\cdot|$ denotes set cardinality. We also suggest the Modified Hausdorff Distance (MHD) [14]:

$$\text{MHD} = \max\left\{ \frac{1}{|P^{gt}|}\sum_{a\in P^{gt}} \min_{b\in P^{ar}} \|a-b\|,\ \frac{1}{|P^{ar}|}\sum_{b\in P^{ar}} \min_{a\in P^{gt}} \|a-b\| \right\},$$

where $\|\cdot\|$ denotes Euclidean distance. The MHD has large discriminatory power, is robust to noise and is easy to interpret (it can be expressed in units of length, e.g., millimeter).

3.3.3. Leaf segmentation

Task: In rosette plants when leaves are highly overlapping, PLA may not be an accurate estimator of plant biomass anymore and individual leaf segmentation is necessary. When individual leaves are segmented, distributions of leaf size can highlight the rate of growth of new leaves with respect to old ones. However, leaf segmentation, a multi-instance segmentation problem [25], is particularly challenging since most leaves within the same plant may share appearance and shape, but can also appear severely overlapping. To complicate matters even more, plant morphology changes radically between mutants, in response to treatment, and as plants grow. Self-occlusion, shadows, leaf hairs, leaf color variations, and others add complexity. Image quality is a factor as well, so low resolution and out-of-focus (as it could occur in portions of Ara2012 and Ara2013 datasets) affect leaf segmentation accuracy. On the other hand, high resolution as in Tobacco, does introduce computational challenges.

Dataset: We used leaf masks but without temporal label consistency. 120 from Ara2012, 165 from Ara2013 (Canon), and 62 from Tobacco, images of single plants appearing centered are used. For each plant, annotations are provided in the form of indexed images (a color palette is embedded for visualization) the same dimensions of the originals. We use one label per leaf, starting from '1' up to the maximum number of leaves, with '0' denoting background.

Evaluation criteria: Several segmentation criteria are available for comparing between ground truth and algorithmic outcomes [37,57]. We suggest SymmetricBestDice, the symmetric average Dice score among all objects (leaves), where for each input label the ground truth label yielding maximum Dice is used for averaging. Best Dice (BD) is defined as:

$$\text{BD}(L^a, L^b) = \frac{1}{M}\sum_{i=1}^{M} \max_{1\le j\le N} \frac{2|L_i^a \cap L_j^b|}{|L_i^a| + |L_j^b|},$$

where $|\cdot|$ denotes leaf area (number of pixels), and L_i^a for $1 \le i \le M$ and L_j^b for $1 \le j \le N$ are sets of leaf object segments belonging to leaf segmentations L^a and L^b, respectively. SymmetricBestDice (SBD) is then:

$$\text{SBD}(L^{ar}, L^{gt}) = \min\left\{ \text{BD}(L^{ar}, L^{gt}),\ \text{BD}(L^{gt}, L^{ar}) \right\}, \quad (1)$$

where L^{gt} is the ground truth and L^{ar} the algorithmic result.

3.3.4. Leaf detection

Task: In an image analysis pipeline, image-based leaf detection could serve to initialize other processes (segmentation or tracking). Due to size differences, shape and appearance similarities, and heavy

occlusions, leaf detection is a complex task, and can benefit from approaches in computer vision of detecting overlapping objects in medicine, transportation, and surveillance [1,60].

Dataset: On the basis of leaf masks, we extracted for each individual leaf the smallest rectangular bounding box (possibly rotated with respect to the image coordinate system) enclosing the mask of that leaf. (In our definition such box would contain both leaf petiole and blade wherever visible and applicable.) The dataset consists of individual plant images, 120 from Ara2012, 165 from Ara2013 (Canon), and 62 from Tobacco, and for each image, a CSV file storing per row the leaf index and the coordinates of each bounding box, with as many rows as number of leaves. Note that our annotation does include the petiole ('leaf stalk') in Arabidopsis.

Evaluation criteria: Number of accurate detections and their accuracy evaluated with overlap measures (cf. Section 3.3.1).

3.3.5. Leaf counting

Task: From a phenotyping perspective the number of leaves is directly related to yield potential, drought tolerance, and flowering time [22,36]. From a computer vision perspective, it can also be used to constrain leaf detection or leaf segmentation algorithms [45]. To this date, user interaction is required and leaf count comes as a by-product [48] of leaf segmentation. Learning-based counting techniques could help here [17,31] and early attempts to automated leaf counting have recently appeared [19,49]. We should note that counting leaves has additional challenges w.r.t. cell/car/people counting since in these classical counting applications all objects share similar shape and size but may have different appearance (e.g., car and people have different colors) may not occlude each other (e.g., cells usually touch but do not overlap). Instead in leaf counting, leaves may heavily overlap, and have different shape and size (mature vs. younger leaves).

Dataset: On the basis of leaf masks we extracted for each leaf the distance transform weighted-center of mass and also the center of mass. When these disagree significantly (above a threshold), and if any of the centers lie outside the binary shape, it indicates a highly asymmetric leaf (e.g., due to heavy occlusion or orientation vertical to the imaging axis) and the annotator was prompted to select a center. Leaf centers in our definition are located in the middle of the visible part of leaf blades. The dataset consists of individual plant images and accompanying binary images containing the centroids for each leaf as a single pixel. (This requires larger storage but we find it more appealing than storing centers in CSV files.) Overall 120 from Ara2012, 165 from Ara2013 (Canon), and 62 from Tobacco, raw and equally numbered annotation images are provided. A CSV file listing image names and number of leaves is also provided, for convenience of approaches that solve directly the regression problem.

Evaluation criteria: Here, we suggest: a) the difference between number of leaves in algorithm's result and ground truth $\text{DiffFGLabels} = \#L^{ar} - \#L^{gt}$, and b) AbsDiffFGLabels, the absolute value of DiffFGLabels. Note that these criteria do not take into account good localization and while count maybe correct it may not correspond to actual leaves. Alternatively, count via detection measures can be adopted (see Section 3.3.4).

3.3.6. Leaf tracking

Task: Finding growth curves of individual leaves helps us understand how a plant (or a cultivar) is growing or the effects of treatments and stresses: for example, Clauw et al. [8] found that drought differentially affects leaves. This growth curve usually follows an exponential relationship with time [51,59], and frequent imaging can capture small differences. This implies the precise segmentation and temporal tracking of each leaf [62].

Dataset: Building upon the leaf masks, the dataset consists of sequentially numbered PNG files of raw individual images and annotations. We provide 4 stacks of 13 images each from Ara2012 and 8 stacks of 17 images each from Ara2013 (Canon). Leaf-level segmentations are provided with leaves having the same label index throughout the sequence to ensure temporal consistency. Also, we release for each image a corresponding CSV file, where bounding box definitions are provided as previously described. Note that this dataset can also be used for leaf or plant segmentation with additional temporal information for example for joint segmentation and tracking.

Evaluation criteria: We recommend the protocol of Nawaz and Cavallaro [44], which builds on overlap criteria, and the code available from the authors. When leaves are vertical in the imaging axis (due to severe nastic movements), overlap criteria may be unable to assign proper correspondences. This may cause lack of label consistency of a leaf across time in an algorithmic result, but it can be easily seen in individual leaf growth curves. Quantitatively, they can be detected by multiple local hypothesis testing to identify structural breaks on the growth parameters [13].

3.3.7. Boundary estimation

Task: Some approaches to multi-instance (or multi-label) segmentation rely on accurate boundary detection, which is used for example to initialize template-based models [25,62]. When image contrast and resolution are adequate, for example in Tobacco, classical edge detection works sufficiently. However, when images are partially out-of-focus and of lower resolution (due to a larger field of view) as is the case of the Arabidopsis data, learning-based methods to boundary estimation (i.e. a learned edge detector as in [34]) have been shown to perform better [49].

Dataset: Using leaf masks, we isolated each leaf label and found its perimeter, to produce an indexed labeled image where '0' is background, '1' denotes a boundary between plant and background, and '2' denotes a boundary between overlapping leaves (which can be more than 1-pixel thick.). This separation may facilitate the training of specialized boundary detectors. The dataset consists of pairs of plant images and these indexed images with boundary annotations, as 120 from Ara2012, 165 from Ara2013 (Canon), and 62 from Tobacco.

Evaluation criteria: Typical criteria such as precision and recall are suggested and those not penalizing small local misalignment, which are suited for evaluating performance of boundaries between leaves, e.g., the MHD (Section 3.3.2) and its learning-based simplification in [40].

3.3.8. Classification and regression

Task: While phenotyping typically occurs in forward hypothesis testing scenarios, recently reverse hypothesis and association studies have received attention and fine-grained categorization is particularly useful. In this case phenotyping traits are recorded and are correlated with genotyping information to identify relationships among them (data mining) [47]. This is particularly useful in the case of treatments, cross hybridizations, and other processes that may affect directly or indirectly (e.g., via silencing and other epigenetic functions) the genetic code of plants [55]. Therefore, given a plant image, it is of great interest to characterize the plant, i.e. plant age or development stage, find other cultivars possessing similar traits, what possible treatment it has undergone. This may be done at plant level or may be even possible at leaf level. When these characteristics are distinct this is considered as a recognition and classification problem. However, if we are interested in percentile similarities, likelihood and regression frameworks can be employed. A similar problem in the context of precision agriculture is weed vs. crop classification, whereupon the interest is to separate a valid crop from a weed in images, usually acquired by automated robotic mechanisms [24]. Here we consider and provide datasets for three cases: mutant recognition (classification), age regression, and treatment recognition (classification).

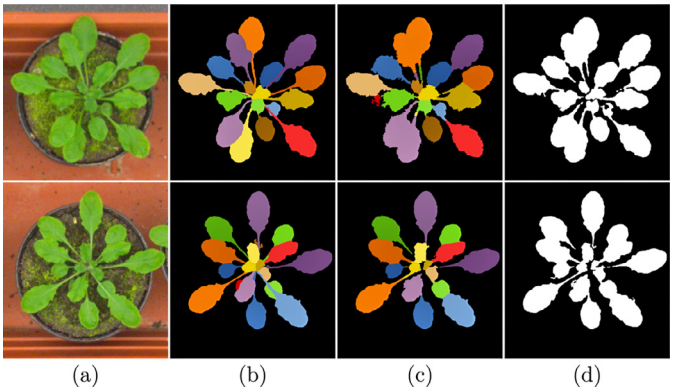


Fig. 8. Two examples of plant and leaf segmentation. (a) Original images, (b) ground truth leaf masks, (c) leaf segmentations obtained by Pape and Klukas [48], and (d) plant segmentations obtained by Minervini et al. [38]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Dataset: For mutant and treatment recognition we release individual plant images and a text list denoting per each row image name, genotype, treatment type. For treatment type classification 62 data from Tobacco are currently available. For mutant classification 165 from Ara2013 (Canon) and 165 from Ara2013 (Rpi) are currently available. We assume the only input to be image data with mutant and treatment type to be predicted values. For age regression, we release individual plant images and a text list denoting per each row image name, mutant type, treatment type, and the age in hours of the plant since germination. For this task 165 images from Ara2013 (Canon), 165 from Ara2013 (Rpi), and 62 from Tobacco are currently available. Notice that the age regression task when mutant or treatment information is not available or treatment is not fixed is extremely difficult since each mutant has different growth rates in response to treatment, so in some sense the algorithm must be able to infer by appearance the unknown mutant type. For simplicity, we assume that inputs to this learning problem are: images and mutant/treatment type information to predict age.

Evaluation criteria: Due to the diversity of problems considered in this category, for classification problems precision and recall criteria are recommended, and for regression problems mean absolute error and mean squared error between predicted and ground-truth measures are encouraged.

4. Examples of use cases

This section offers examples of how parts of the datasets outlined in the previous section have been used in developing and evaluating computer vision algorithms.

For plant segmentation from complicated background using parts of the Ara2012 dataset, Minervini et al. [38] used a new multi-channel active contour based on [6] with probabilistic priors on plant appearance. The images used included examples of heavy growth of moss in the pot, water in the tray, discoloring of plants due to temporary drought, and pot movement. On average, DSC $\approx$ 97% (cf. Section 3.3.2) was observed for segmentation accuracy when using color and texture features and a Gaussian mixture model to describe plant appearance. In the same work, rosettes were tracked over time using overlap criteria and nearest neighbor rules. Nearest neighbor rules and a *k*-means clustering on the Euclidean coordinates of plant pixels were also used to resolve overlapping plants. Examples are shown in Fig. 8d, where DSC values of 98% and 97% are reported for top and bottom images, respectively.

A specially formatted part of the data presented here was released to support the Leaf Segmentation Challenge (http://www.plant-phenotyping.org/CVPPP2014-challenge), of the Computer Vision Problems in Plant Phenotyping (CVPPP) workshops 2014 and 2015. CVPPP 2014 was held in conjunction with the European Conference on Computer Vision (ECCV), in Zürich, Switzerland, in September 2014; CVPPP 2015 in conjunction with the British Machine Vision Conference (BMVC), in Swansea, UK, in September 2015. Plant images were considered separately, ignoring temporal correspondence. Images contained instances with well separated leaves and simple background, but also more difficult examples with many leaf occlusions, complex leaf shapes, varying backgrounds, or plant objects not well in focus. Individual plant images and ground truth segmentations were used consisting of Ara2012, Ara2013 (Canon), and Tobacco datasets. The datasets were split into training and testing sets for the challenge. Respectively: for training, 128, 31, and 27 images and corresponding annotations were released; for testing, 33, 9, and 56 images were released (ground truth was available only to the organizers). Within this challenge, solutions were evaluated on the basis of plant segmentation, individual leaf segmentation, and leaf counting with the criteria outlined in Sections 3.3.3 and 3.3.5. Table 2 summarizes overall performance among participants that completed the challenge and the leading approaches of Pape and Klukas [48,49]. Although plant segmentation accuracy (DSC of plant mask) was somewhat acceptable, when considering leaf segmentation (SymmetricBestDice, Eq. (1)) and leaf count criteria poor performance was observed across the algorithms in LSC 2014, indicating the challenging nature of the problem. Some of these issues are evident also in the examples of Fig. 8c: compared to the ground truth, the approach of Pape and Klukas [48] tends to over-segment (when moss is present, e.g., the top image), separate petioles, and merge leaves. Both images present AbsDiffFGLabels = 2, so count is off but for different reasons. In order to distinguish between foreground-background segmentation errors and leaf-wise segmentation errors, ground truth foreground-background masks were provided for LSC 2015. While the average SymmetricBestDice improved to 71.3%, this performance is still quite low, which indicates that for leaf-wise segmentation reliable and highly accurate tools are still not available.

Table 2
Average performance among all participants compared to the leading approaches on the Leaf Segmentation Challenge (LSC) of CVPPP 2014 and CVPPP 2015. In parentheses standard deviations among participants for the first column, and among the three different datasets for the second column. *Note that ground truth foreground-background masks were made available in LSC 2015.

	Overall LSC 2014	Leader LSC 2014	Leader LSC 2015
SymmetricBestDice [%]	48.7 (12.8)	62.6 (19.0)	71.3 (15.1)
DSC [%]	83.1 (10.9)	95.3 (10.1)	–*
AbsDiffFGLabels	6.3 (4.36)	2.4 (2.1)	1.1 (1.2)
DiffFGLabels	1.3 (4.4)	−1.9 (2.7)	−0.3 (1.6)

A comparative study of improved versions of the algorithms devised and tested in the context of the LSC 2014 can be found in [53]. Within an in-depth analysis and discussion of the results are offered.

We also explored learning directly leaf count density based on leaf center annotations, following the approach by Lempitsky and Zisserman [31]. We used a subset of 84 images from the Ara2012 dataset, with images depicting leaf occlusions, some being out of focus, and some showing moss in the pot. The overall number of leaves per plant varied from 12 to 19. Based on leaf center annotations (Section 3.3.5), the goal was to learn via appropriate losses, density functions, the integration of which provides object counts. We extracted from the green color channel dense SIFT descriptors [33] in 20 of the 84 images, with 7 SIFT bins and fixed orientations. Subsequently, we quantized the SIFT space using *k*-means clustering to create a codebook of size 1500. Using this codebook, we learned a linear transformation of the feature representation on the codebook approximating the density function at each pixel on 64 training images [31]. Testing on 33 images of Ara2012 of the LSC dataset, obtained an average

Table 3
Performance of the leading approaches on the Leaf Counting Challenge (LCC) of CVPPP 2015. In parentheses standard deviations among the three different datasets.

LCC 2015	[19]	[49]
AbsDiffFGLabels	1.43 (1.51)	1.1 (1.2)
DiffFGLabels	−0.51 (2.02)	−0.3 (1.6)

AbsDiffFGLabels = 2.36(2.9) (cf. Section 3.3.5). Among those, in 67% of images we observed either no counting error or the count was off by 1 or 2 leaves at most. In comparison, referring to findings of Pape and Klukas [48] on the same testing data, counting via segmentation performed with AbsDiffFGLabels = 2.2(1.3). For the example of Fig. 8, by learning, the count was off only by 1 leaf for both test images. While these results are preliminary they demonstrate the promise of learning-based leaf count estimation.

These early findings illustrate the complexity of the problems at hand. In fact, the leaf counting problem motivated a new challenge, the leaf counting challenge (LCC) for the 2015 edition of CVPPP. The leading methods are presented in [19,49]. The approaches, while they do rely on global regression, they are different as to the features used. Pape and Klukas [49] extract features out of the foreground mask (e.g., perimeter, area, etc.), which after feature selection are used in a regression framework to learn leaf count per image. Several regression functions are evaluated such as support vector regression, random forest regression, and others. Different features were used for each dataset. On the other hand, the solution proposed by Giuffrida et al. [19] learns features directly from the data in an unsupervised fashion to build a codebook. From the images, image patches are extracted and projected on the codebook. Responses are max pooled, leading to a global per image descriptor. Leaf count is estimated from this descriptor using support vector regression. Summary results are shown in Table 3. We see that results clearly improved with respect to the best LSC 2014 results (Table 2).

5. Discussion and conclusions

This paper describes the first ever collection of image datasets of growing Arabidopsis and tobacco plants in a plant phenotyping context, together with annotations for a series of computer vision and learning tasks at different levels of granularity. Plant phenotyping is central to the understanding of plant function and is a tool that can enable us to meet agricultural demands of the future. Computer vision and machine learning are ideally suited to help, and this collection of datasets is intended to promote the exploration of plant phenotyping problems.

As a benefit to the scientific community, we will be continuously releasing in the public domain specially formatted datasets and annotations, accompanied by appropriate functions implementing evaluation criteria. We will be accepting contributions and corrections from the scientific community in an effort to keep these data curated. This paper will serve as the reference document on the structure and importance of these datasets. In order to facilitate future annotation by us and the broad community, following ideas from other domains [18], we have already developed annotation tools [39]. Implementing in the future such tools in web-based environments would permit the crowd-sourcing of annotations [21,30]. We will be also investigating additional evaluation criteria as we obtain feedback from the community.

We do hope that our publicly available datasets and future augmented versions and recent multi-modal plant image databases [10] (this work appeared at the time this article went in press) will be adopted by the broad computer vision community as well (as with the PASCAL [15], or the biologically focused Broad Bioimage Benchmark Collection [32]). Our datasets can be used to learn suitable image statistics [27], adapt and test counting algorithms with [17] and without temporal information [1,31], segmentation algorithms [6], multi-label segmentation [25,45] or detection [3] approaches, and others. Additional depth information as can be computed from a pair of images with different focus [16] of the Arabidopsis dataset or stereo images [5] of the Tobacco dataset, to be released in the future, may further facilitate segmentation [12]. Using the presented datasets, image-based plant phenotyping will evolve in parallel to (and benefit from) advances in computer vision, by tracking the performance of approaches referencing these data. More importantly, it will also introduce this societally important application to a wider audience.

Acknowledgments

Tobacco research has received funding from EU's 7th Framework Programme (FP7/2007-2013) under Grant no. 247947 (GARNICS). Part of this work was performed within the German-Plant-Phenotyping Network, which is funded by the German Federal Ministry of Education and Research (project identification no. 031A053). Arabidopsis research was partially supported by a Marie Curie Action: "Reintegration Grant" (Grant no. 256534) of the EU's 7th Framework Programme.

The authors would like to thank Prof. Pierdomenico Perata and his group from Scuola Superiore Sant'Anna, Pisa, Italy, for providing plant samples and instructions on growth conditions of Arabidopsis. Finally, they also thank several annotators who have contributed to this project: Fabiana Zollo, Ines Dedovic, Mario Valerio Giuffrida, and Vasileios Sevetlidis.

References

[1] C. Arteta, V. Lempitsky, J.A. Noble, A. Zisserman, Learning to detect partially overlapping instances, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, IEEE, 2013, pp. 3230–3237.
[2] M. Augustin, Y. Haxhimusa, W. Busch, W.G. Kropatsch, Image-based phenotyping of the mature Arabidopsis shoot system, in: Computer Vision - ECCV 2014 Workshops, Springer, 2015, pp. 231–246.
[3] O. Barinova, V.S. Lempitsky, P. Kohli, On detection of multiple object instances using hough transforms, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1773–1784.
[4] S. Bergsträsser, D. Fanourakis, S. Schmittgen, M. Cendrero-Mateo, M. Jansen, H. Scharr, U. Rascher, HyperART: non-invasive quantification of leaf traits using hyperspectral absorption–reflectance–transmittance imaging, Plant Methods 11 (1) (2015) 1–17.
[5] B. Biskup, H. Scharr, U. Schurr, U. Rascher, A stereo imaging system for measuring structural parameters of plant canopies, Plant Cell Environ. 30 (2007) 1299–1308.
[6] T. Chan, L. Vese, Active contours without edges, IEEE Trans. Image Process. 10 (2) (2001) 266–277.
[7] A. Chavarria-Krauser, K. Nagel, K. Palme, U. Schurr, A. Walter, H. Scharr, Spatio-temporal quantification of differential growth processes in root growth zones based on a novel combination of image sequence processing and refined concepts describing curvature production, New Phytol. 177 (2008) 811–821.
[8] P. Clauw, F. Coppens, K. De Beuf, S. Dhondt, T. Van Daele, K. Maleux, V. Storme, L. Clement, N. Gonzalez, D. Inze, Leaf responses to mild drought stress in natural variants of Arabidopsis thaliana, Plant Physiol. (2015) (http://www.plantphysiol.org/content/early/2015/01/20/pp.114.254284.1).
[9] CORESTA, A scale for coding growth stages in tobacco crops, 2009, http://www.coresta.org/Guides/Guide-No07-Growth-Stages_Feb09.pdf.
[10] J.A. Cruz, X. Yin, X. Liu, S.M. Imran, D.D. Morris, D.M. Kramer, J. Chen, Multi-modality imagery database for plant phenotyping, Mach. Vis. Appl. (2015). In press.
[11] J. De Vylder, F.J. Vandenbussche, Y. Hu, W. Philips, D. Van Der Straeten, Rosette tracker: an open source image analysis tool for automatic quantification of genotype effects, Plant Physiol. 160 (3) (2012) 1149–1159.
[12] B. Dellen, G. Alenyà, S. Foix, C. Torras, Segmenting color images into surface patches by exploiting sparse depth data, in: Proceedings of the IEEE Workshop on Applications of Computer Vision WACV, 2011, pp. 591–598.
[13] B. Dellen, H. Scharr, C. Torras, Growth signatures of rosette plants from time-lapse video, IEEE/ACM Trans. Comput. Biol. Bioinform. PP (99) (2015) 1–11.
[14] M.P. Dubuisson, A.K. Jain, A modified Hausdorff distance for object matching, in: Proceedings of the Twelfth IAPR International Conference on Pattern Recognition, 1994, pp. 566–568.
[15] M. Everingham, L. Van Gool, C.K.I. Williams, J. Winn, A. Zisserman, The PASCAL visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[16] P. Favaro, S. Soatto, 3-D Shape Estimation and Image Restoration: Exploiting Defocus and Motion-Blur, Springer-Verlag, 2007.

[17] L. Fiaschi, K. Gregor, B. Afonso, M. Zlatic, F.A. Hamprecht, Keeping count: leveraging temporal context to count heavily overlapping objects, in: Proceedings of the 2013 International Symposium on Biomedical Imaging (ISBI), 2013, pp. 656–659.
[18] D. Giordano, I. Kavasidis, S. Palazzo, C. Spampinato, Nonparametric label propagation using mutual local similarity in nearest neighbors, Comput. Vis. Image Underst. 131 (2015) 116–127.
[19] M.V. Giuffrida, M. Minervini, S.A. Tsaftaris, Learning to count leaves in rosette plants, in: Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP) Workshop, BMVA Press, 2015, pp. 1.1–1.13.
[20] H. Goëau, P. Bonnet, A. Joly, I. Yahiaoui, D. Barthélémy, B. Nozha, J.-F. Molino, The ImageCLEF 2012 plant identification task, in: Proceedings of the Conference and Labs of the Evaluation Forum CLEF, 2012, pp. 1–25.
[21] S.A. Goff, et al., The iPlant collaborative: Cyberinfrastructure for plant biology, Front. Plant Sci. 2 (34) (2011) 1–16.
[22] C. Granier, L. Aguirrezabal, K. Chenu, S.J. Cookson, M. Dauzat, P. Hamard, J.-J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, F. Tardieu, PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in Arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit, New Phytol. 169 (3) (2006) 623–635.
[23] A. Hartmann, T. Czauderna, R. Hoffmann, N. Stein, F. Schreiber, HTPheno: An image analysis pipeline for high-throughput plant phenotyping, BMC Bioinform. 12 (1) (2011) 148.
[24] S. Haug, A. Michaels, P. Biber, J. Ostermann, Plant classification system for crop/weed discrimination without segmentation, in: Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2014, pp. 1142–1149.
[25] X. He, S. Gould, An exemplar-based CRF for multi-instance object segmentation, in: Proceedings of the Twenty-Seventh IEEE Conference on Computer Vision and Pattern Recognition CVPR, IEEE, 2014, pp. 296–303.
[26] G. van der Heijden, Y. Song, G. Horgan, G. Polder, A. Dieleman, M. Bink, A. Palloix, F. van Eeuwijk, C. Glasbey, SPICY: towards automated phenotyping of large pepper plants in the greenhouse, Funct. Plant Biol. 39 (11) (2012) 870–877.
[27] M. Heiler, C. Schnörr, Natural image statistics for natural image segmentation, Int. J. Comput. Vis. 63 (2005) 5–19.
[28] E. van Henten, D. Goense, C. Lokhorst (Eds.), Precision Agriculture '09, Wageningen Academic, 2009.
[29] M. Jansen, F. Gilmer, B. Biskup, K. Nagel, U. Rascher, A. Fischbach, S. Briem, G. Dreissen, S. Tittmann, S. Braun, I.D. Jaeger, M. Metzlaff, U. Schurr, H. Scharr, A. Walter, Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in Arabidopsis thaliana and other rosette plants, Funct. Plant Biol. 36 (10/11) (2009) 902–914.
[30] I. Kavasidis, S. Palazzo, R. Di Salvo, D. Giordano, C. Spampinato, An innovative web-based collaborative platform for video annotation, Multimed. Tools Appl. 70 (1) (2014) 413–432.
[31] V.S. Lempitsky, A. Zisserman, Learning to count objects in images, in: Proceedings of the Twenty-Third Annual Conference on the Advances in Neural Information Processing Systems, 2010, pp. 1324–1332.
[32] V. Ljosa, K.L. Sokolnicki, A.E. Carpenter, Annotated high-throughput microscopy image sets for validation, Nat. Methods 9 (7) (2012) 637–638.
[33] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[34] D. Martin, C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549.
[35] U. Meier, Growth stages of mono-and dicotyledonous plants, Blackwell Wissenschafts-Verlag, 1997.
[36] B. Méndez-Vigo, M.T. de Andrés, M. Ramiro, J.M. Martínez-Zapater, C. Alonso-Blanco, Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in Arabidopsis thaliana, J. Exp. Bot. 61 (6) (2010) 1611–1623.
[37] V. Mezaris, I. Kompatsiaris, M. Strintzis, Still image objective segmentation evaluation using ground truth, in: Proceedings of the Fifth COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication, 2003, pp. 9–14.
[38] M. Minervini, M.M. Abdelsamea, S.A. Tsaftaris, Image-based plant phenotyping with incremental learning and active contours, Ecol. Infor. 23 (2014) 35–48.
[39] M. Minervini, M.V. Giuffrida, S.A. Tsaftaris, An interactive tool for semi-automated leaf annotation, in: Proceedings of the Computer Vision Problems in Plant Phenotyping Workshop, BMVA Press, 2015, pp. 6.1–6.13.
[40] M. Minervini, C. Rusu, S.A. Tsaftaris, Learning computationally efficient approximations of complex image segmentation metrics, in: Proceedings of the International Symposium on Image and Signal Processing and Analysis, 2013, pp. 60–65.
[41] M. Minervini, H. Scharr, S.A. Tsaftaris, Image analysis: the new bottleneck in plant phenotyping, IEEE Signal Process. Mag. 32 (4) (2015) 126–131.
[42] M. Müller-Linow, F. Pinto-Espinosa, H. Scharr, U. Rascher, The leaf angle distribution of natural plant populations: assessing the canopy with a novel software tool, Plant Methods 11 (1) (2015) 1–16.
[43] K. Nagel, A. Putz, F. Gilmer, K. Heinz, A. Fischbach, J. Pfeifer, M. Faget, S. Blossfeld, M. Ernst, C. Dimaki, B. Kastenholz, A.-K. Kleinert, A. Galinski, H. Scharr, F. Fiorani, U. Schurr, GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons, Funct Plant Biol. 39 (2012) 891–904.
[44] T.H. Nawaz, A. Cavallaro, A protocol for evaluating video trackers under real-world conditions, IEEE Trans. Image Process. 22 (4) (2013) 1354–1361.
[45] C. Nieuwenhuis, E. Toeppe, D. Cremers, A survey and comparison of discrete and continuous multi-label optimization approaches for the Pott's model, Int. J. Comput. Vis. 104 (3) (2013) 223–240.
[46] M.-E. Nilsback, A. Zisserman, Delving deeper into the whorl of flower segmentation, Image Vis. Comput. 28 (6) (2010) 1049–1062.
[47] R.C. O'Malley, J.R. Ecker, Linking genotype to phenotype using the Arabidopsis unimutant collection, Plant J. 61 (6) (2010) 928–940.
[48] J.-M. Pape, C. Klukas, 3-D histogram-based segmentation and leaf detection for rosette plants, in: Computer Vision - ECCV 2014 Workshops, Springer, 2014, pp. 61–74.
[49] J.-M. Pape, C. Klukas, Utilizing machine learning approaches to improve prediction of leaf counts and individual leaf segmentation of rosette plants, in: Proceedings of the Computer Vision Problems in Plant Phenotyping Workshop, BMVA Press, 2015, pp. 3.1–3.12.
[50] A. Perina, M. Cristani, V. Murino, 2LDA: Segmentation for recognition, in: Proceedings of the International Conference on Pattern Recognition, 2010, pp. 995–998.
[51] F.J. Richards, A flexible growth function for empirical use, J. Exp. Bot. 10 (2) (1959) 290–301.
[52] H. Scharr, M. Minervini, A. Fischbach, S.A. Tsaftaris, Annotated Image Datasets of Rosette Plants, Technical Report FZJ-2014-03837, Forschungszentrum Jülich, 2014.
[53] H. Scharr, M. Minervini, A.P. French, C. Klukas, D.M. Kramer, X. Liu, I.L. Muntión, J.-M. Pape, G. Polder, D. Vukadinovic, X. Yin, S.A. Tsaftaris, Leaf segmentation in plant phenotyping: a collation study, Mach. Vis. Appl. (2015). In press.
[54] P.F. Silva, A.R. Marçal, R.M.A. da Silva, Evaluation of features for leaf discrimination, in: Image Analysis and Recognition, Springer, 2013, pp. 197–204.
[55] T.K. To, J.-M. Kim, Epigenetic regulation of gene responsiveness in Arabidopsis, Front. Plant Sci. 4 (548) (2014) 1–6.
[56] S.A. Tsaftaris, C. Noutsos, Plant phenotyping with low cost digital cameras and image analytics, in: Information Technologies in Environmental Engineering, in: Environmental Science and Engineering, Springer, 2009, pp. 238–251.
[57] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 929–944.
[58] A. Walter, H. Scharr, F. Gilmer, R. Zierer, K.A. Nagel, M. Ernst, A. Wiese, O. Virnich, M.M. Christ, B. Uhlig, S. Jünger, U. Schurr, Dynamics of seedling growth acclimation towards altered light conditions can be quantified via GROWSCREEN: a setup and procedure designed for rapid optical phenotyping of different plant species, New Phytol. 174 (2) (2007) 447–455.
[59] A. Walter, U. Schurr, The modular character of growth in Nicotiana tabacum plants under steady-state nutrition, J. Exp. Bot. 50 (336) (1999) 1169–1177.
[60] P. Wohlhart, M. Donoser, P.M. Roth, H. Bischof, Detecting partially occluded objects with an implicit shape model random field, in: Proceedings of the Asian Conference on Computer Vision (ACCV), 2013, pp. 302–315.
[61] S. Wu, F. Bao, E. Xu, Y.-X. Wang, Y.-F. Chang, Q.-L. Xiang, A leaf recognition algorithm for plant classification using probabilistic neural network, in: Proceedings of the IEEE International Symposium Signal Processing and Information Technology (ISSPIT), 2007, pp. 11–16.
[62] X. Yin, X. Liu, J. Chen, D.M. Kramer, Multi-leaf tracking from fluorescence plant videos, in: Proceedings of the 2014 IEEE International Conference on Image Processing, 2014, pp. 408–412.