Noname manuscript No.
(will be inserted by the editor)

Multi-modality Imagery Database for Plant Phenotyping

Jeffrey A Cruz* · Xi Yin* · Xiaoming Liu · Saif M Imran · Daniel D Morris · David M Kramer · Jin Chen

Received: date / Accepted: date

Abstract Among many applications of machine vision, plant image analysis has recently began to gain more attention due to its potential impact on plant visual phenotyping, particularly in understanding plant growth, assessing the quality/performance of crop plants, and improving crop yield. Despite its importance, the lack of publicly available research databases containing plant imagery has substantially hindered the advancement of plant image analysis. To alleviate this issue, this paper presents a new multi-modality plant imagery database named "MSU-PID", with two distinct properties. First, MSU-PID is captured using four types of imaging sensors, fluorescence, infrared (IR), RGB color, and depth. Second, the imaging setup and the variety of manual labels allow MSU-PID to be suitable for a diverse set of plant image analysis applications, such as leaf segmentation, leaf counting, leaf alignment, and leaf tracking. We provide detailed information on the plants, imaging sensors, calibration, labeling, and baseline performances of this new database.

Keywords Plant Phenotyping · Computer Vision · Plant image · Leaf segmentation · Leaf tracking · Multiple sensors · Arabidopsis · Bean

1 Introduction

With the rapid growth of world population and the loss of arable land, there is an increasing desire to improve the yield and quality of crops. Key to increasing yields is gaining understanding of the genetic mechanisms that influence plant growth [Döös, 2002]. A classic genetic approach is to produce a diverse population of mutant lines, grow them either in growth chambers with simulated environmental conditions or directly in the field, visually observe the plants during the growth period, and finally identify plant morphological or physiological patterns that tightly associate with key growth factors [Houle et al., 2010]. While many factors can be assessed quantitatively, which is essential for high-throughput study, one of the bottleneck in this research pipeline is plant visual phenotyping [Walter et al., 2015].

The complex interaction between genotype and environment determines how plants will develop physiologically (from the biochemical level to structural morphology), which ultimately influences field performance, for example resource acquisition, resource partitioning and seed yield. In order to understand the genetic basis for these economically important parameters, it is essential to quantitatively assess plant phenotypes and then identify the latent relationships between genotype and environmental factors. Plant visual phenotyping has been performed by farmers and breeders

* denotes equal contribution by the authors. This research was supported by Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (award number DE-FG02-91ER20021), and by Center for Advanced Algal and Plant Phenotyping (CAAPP), Michigan State University.

Jeffrey A Cruz · Jin Chen · David M Kramer
Department of Energy Plant Research Laboratory, Michigan State University, East Lansing, MI 48824, USA
E-mail: {cruzj,jinchen,kramerd8}@msu.edu

Xi Yin · Xiaoming Liu (✉)
Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA
E-mail: {yinxi1,liuxm}@cse.msu.edu

Saif Imran · Daniel D Morris
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA
E-mail: {imransai,dmorris}@msu.edu

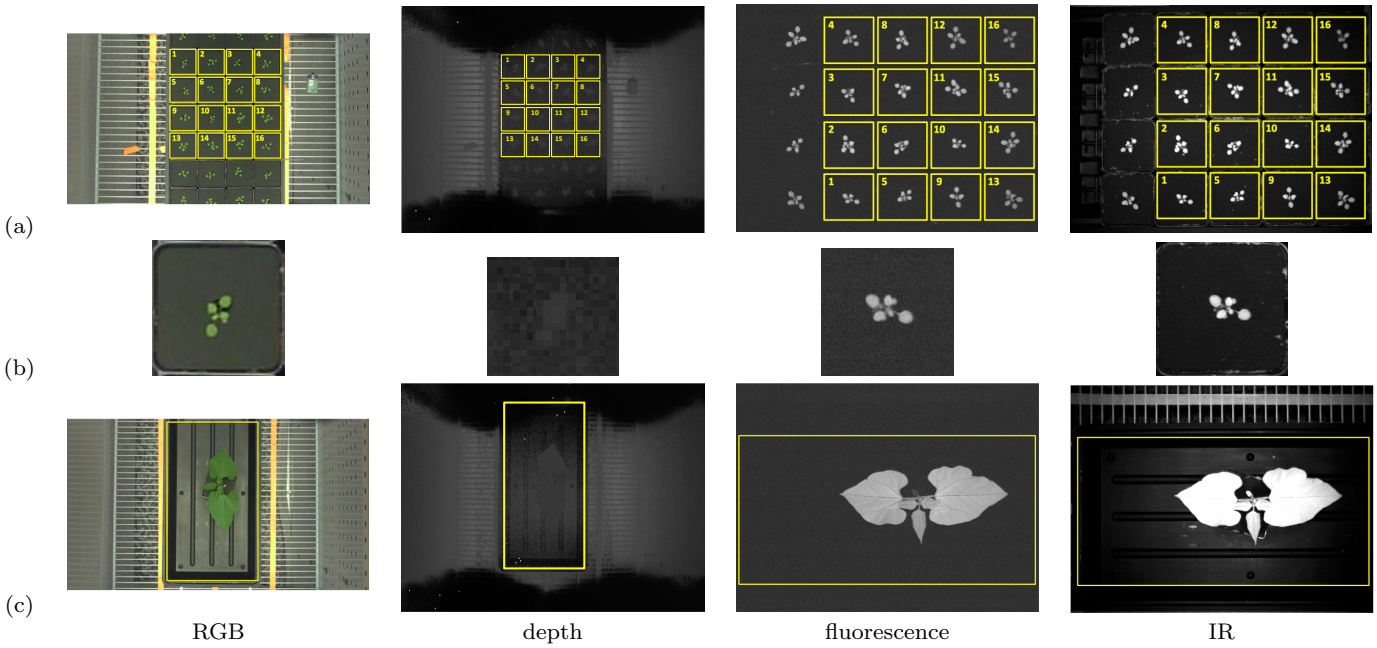


Fig. 1 The multi-modality plant imagery database of MSU-PID. (a) four modalities of Arabidopsis; (b) zoom in view of Arabidopsis plant 1; (c) four modalities of bean.

for more than 5,000 years to develop plant lines with desirable traits. This traditional method is based on experience and intuition and is both labor intense and time-consuming [Johannsen, 1903]. Recent progress in imaging sensor, robotics and automation technologies has recently led to the rapid development of highly-automated, non-destructive, high-throughput methods for visual phenotyping [Furbank & Tester, 2011, Kramer et al., 2015]. The intent of modern plant visual phenotyping is to accelerate the collection, analysis and categorization visual (e.g. morphology, size) and spectral (e.g. chlorophyll fluorescence, chlorophyll content, water content, leaf temperature) data sets in order to identify and quantify specific plant traits, especially those related to plant performance. This interdisciplinary field requires scientists and engineers to not only develop or employ various image sensors platforms to collect under relevant conditions, but also design advance algorithms for the automated processing and analysis of image data.

One goal of plant visual phenotyping is to accurately determine genetic components related to photosynthetic performance, growth, yield, and resilience to environmental stress. For this approach to be of effective, plants must be monitored continuously, over developmental time scales and under conditions that approximate the native environment (dynamic 'field' conditions) [Fahlgren et al., 2015, Walter et al., 2015]. The capacity to identify and track individual leaves over time is important to clearly define phenotypic behaviors related to leaf-specific factors such as developmental age [Schöttler et al., 2015] and/or following the application of stress. For example, decrease in the average photosynthetic efficiency (Φ_{II}, quantum yield of photochemistry at photosystem II) of an Arabidopsis plant may reflect a heterogeneous distribution of Φ_{II} values across its entire photosynthetic surface area [Oxborough, 2004]. In this case, mapping distributions to individual leaves can distinguish between a systemic (whole plant) effect and localization of the effect to subset leaves. In addition, using more detailed information, such as leaf angles, height, and position as well as canopy density and size, it is possible to improve the estimation of total photosynthetic capacity over time (particularly for plants with more complex canopies, e.g., common bean or tobacco), since these factors influence absorption and availability of the light energy used to drive photosynthesis.

Due to diverse variations of leaf shape, appearance, layout, growth and movement, plant image analysis is a non-trivial computer vision task [Minervini et al., 2015]. In order to develop advanced computer vision algorithms, image databases that are well representative of this application domain is highly important. In fact, computer vision research lives on and advances with databases, as evidenced by the successful databases in the field (e.g., FERET [Phillips et al., 2000] and LFW [Huang et al., 2007]). However, the publicly available databases for plant phenotyping are

Table 1 Plant image databases.

Database	Modality	Applications[a]	Plant Type	Subject/ Classe #	Total Image #	Labeled Image #
Swedish leaf [Söderkvist, 2001]	Scaned leaf	LC	Swedish trees	15	1,125	1,125
Flavia [Wu et al., 2007]	RGB	LC	Leaves	32	2,120	2,120
Leafsnap [Kumar et al., 2012]	RGB	LC	USA trees	184	29,107	29,107
Crop/weed [Haug & Ostermann, 2014]	RGB	Weed detection	Crop/weed	2	60	60
LSC [Scharr et al., 2014]	RGB	LS, LO, LT[b]	Arabidopsis	43	6,287	201
			Tobacco	80	165,120	83
MSU-PID	fluorescence, IR, RGB, depth	LS, LO LA, LT	Arabidopsis	16	$2,160 \times 4$	576×4
			Bean	5	325×4	175×2

[a] The abbreviation in "Applications" column is defined as Leaf Classification (LC), Leaf Segmentation (LS), Leaf Counting (LO), Leaf Alignment (LA), and Leaf Tracking (LT).
[b] The data collected in [Scharr et al., 2014] does have temporal images which will allow leaf tracking. However, the dataset that they release does not include any temporal information.

still very limited, with the only exception of LSC database [Scharr et al., 2014], which, nevertheless, has its own limitations on the type of images (RGB only) and not having released data related to tracking (although the authors claim in [Scharr et al., 2014] that such data is available.)

To facilitate future research on plant image analysis, as well as remedy the limitation of existing databases in the field, this paper presents a newly collected multi-modality Plant Imagery Database through an interdisciplinary effort at Michigan State University (MSU), which is termed "MSU-PID". As illustrated in Figure 1, the MSU-PID database includes the imagery of two types of plants (Arabidopsis and bean, both are widely used in plant research) captured by four types of imaging sensors, i.e., fluorescence, infrared (IR), RGB color, and depth. All four sensors are synchronized and are programmed to periodically capture imagery data for multiple consecutive days. Checkerboard-based camera calibration is performed on each modality providing intrinsic camera parameters and poses. In addition explicit correspondence between the pixels of *any* two modalities is obtained using an homographic warping of a plane through the Arabidopsis plants.

The type and amount of manual labels on a database is a critical enabler to the potential applications of the database. For a subset of the MSU-PID database, we manually label the ground truth regarding the leaf identification number, locations of leaf tips, and leaf segments. As a result, MSU-PID is suitable for a number of applications, including 1) *leaf segmentation* that aims at estimating the correct segmentation mask of each leaf in an image, 2) *leaf counting* that estimates the correct number of leaves within a plant, 3) *leaf alignment* that aligns the two tips of each leaf – the cornerstone of the leaf structure, and 4) *leaf tracking* that is designed to track each leaf over time. Finally, to provide a performance baseline for future research and comparison, we apply our automatic leaf segmentation framework [Yin et al., 2014a, Yin et al., 2014b] to the Arabidopsis imagery and demonstrate the unique challenge of image analysis on this database.

In summary, this paper and our database have made the following main contributions.

- MSU-PID is the first *multi-modality* plant image database. This allows researchers to study the strength and weakness of individual modality, as well as their various combinations in plant image analysis.
- Our unique imaging setup and the variety of manual labels make MSU-PID an ideal candidate for evaluating a diverse set of plant image analysis applications including leaf segmentation, leaf counting, leaf alignment, leaf tracking, and potentially leaf growth prediction and 3D leaf reconstruction.

2 Prior Work

Databases drive computer vision research. Hence, it is always important to develop and promote properly captured databases in the vision community. While there is a clear desire to apply computer vision to plant image analysis, the lack of publicly available plant image databases has been an obstacle for further study and development.

We summarize existing publicly available databases that are most related to our work in Table 1. In terms of potential applications of these databases, they can be categorized into two types. The first

type is for the general purpose of recognizing a particular species of tree or plant. The Swedish leaf database [Söderkvist, 2001] is probably the first leaf database even though the images are captured by scanners. The Flavia database [Wu et al., 2007] is considerably larger and a neural network is utilized to train a leaf classifier. The most recent Leafsnap project [Kumar et al., 2012] is an impressive effort that includes a very large dataset of leaves for 184 tree types. A mobile phone application is also developed to make the leaf classification system portable. Finally, the crop/weed image database [Haug & Ostermann, 2014] is captured by a robot in the real field, and used for the classification of crop vs. weed. Note that except for [Haug & Ostermann, 2014] where images are captured in the wild for a large area of plants, other databases in this type normally capture only a single leaf in a relatively constrained imaging environment. Therefore, the challenging problem of leaf segmentation has been bypassed.

The second type of databases is for plant phenotyping, where it is important to capture plant images without interfering the growth of plants. Thus, non-destructive imaging approaches are taken and the entire plant is imaged. The LSC database [Scharr et al., 2014] is the most relevant one to ours. It captures a large set of RGB images for Arabidopsis and tobacco plants. The provided manual labels allow the evaluation of leaf segmentation and leaf counting. Although it is claimed that they have collected data for leaf tracking, it has not been included in the current LSC dataset. In comparison, our MSU-PID database utilizes four sensing modalities in the data capturing, each providing different aspects of plant visual appearance. Our diverse manual labels also enable us to develop algorithms for additional applications such as leaf tracking and leaf alignment.

The four sensing modalities in MSU-PID provide unique opportunities to comprehensively characterize plant morphological and physiological phenotypes. The use of chlorophyll fluorescence at 730 nm to 750 nm as a tool for evaluating photosynthetic performance is well established [Baker, 2008], and against non-fluorescent background, it clearly defines the chlorophyll-containing (photosynthetically active) leaf area. The depth measurement has been a component of a number of recent non-plant RGB-D databases designed for object recognition [Lai et al., 2011], scene segmentation [Silberman & Fergus, 2011], human analysis [Sung et al., 2011, Barbosa et al., 2012], and mapping [Sturm et al., 2012]. By including depth map for a plant database, we anticipate enabling development of new 3D plant canopy analysis algorithms and thus probing the total energy intake and storage. Near infrared (IR) reflectance at $\sim$ 940 nm has been used by others to detect water content in leaves [Chen et al., 2014]. However, it may also be useful in determining leaf angle and curvature [Woolley, 1971]. Furthermore, since imaging occurs at a wavelength that is effectively non-absorbing for photosynthesis or known light receptors [Butler et al., 1964, Eskins, 1992] in plants, it can be used for imaging during the night cycle to observe night time leaf expansion or in some cases circadian movements [McClung, 2006].

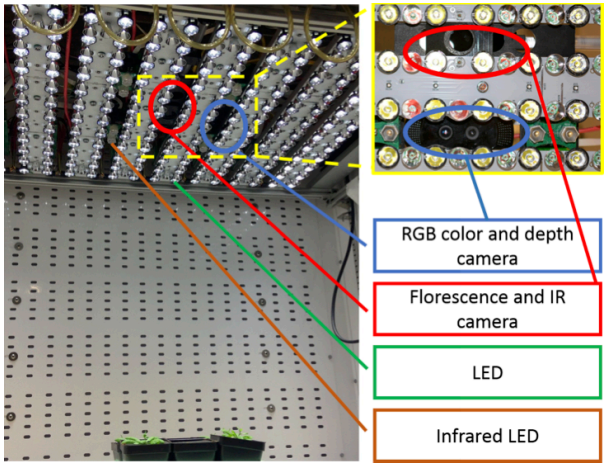


Fig. 2 The hardware setup for our data collection.

3 Data Collection

3.1 Plants and Growth Conditions

Arabidopsis thaliana (ecotype Col-0) plants were grown at 20°C, under a 16 hr : 8 hr day-night cycle with a daylight intensity set at 100 $\mu mol\ photons\ m^{-2}s^{-1}$. Black bean plants (*Phaseolus vulgaris L.*) of the cultivar Jaguar, were grown under a 14 hr : 10 hr day-night cycle with day and night temperatures of 24°C and 18°C respectively, and a daylight intensity set at 200 $\mu mol\ photons\ m^{-2}s^{-1}$. Note that the bean plants were watered with half-strength Hoaglands solution three times per week.

In all cases, seeds were planted in soil covered with a black foam mask in order to minimize the fluorescence background from algal growth. Two-week-old plants (*Arabidopsis* and bean) were transferred to imaging chambers and allowed to acclimate for 24 hours to the LED lighting before starting the data collection. Aforementioned growth conditions were maintained for each set of plants for the duration of image collection.

3.2 Hardware Setup

In this section, we introduce the hardware used for capturing fluorescence, IR, RGB color, and depth imagery data for both plants. Figure 2 illustrates the hardware and imaging setup used in our data collection.

3.2.1 Fluorescence and IR Images

Chlorophyll *a* fluorescence images were captured once every hour during the daylight period in a growth chamber [Kramer et al., 2015]. A set of 5 images were captured using a Hitachi KP-F145GV CCD camera (Hitachi Kokusai Electric America Inc., Woodbury, NY) outfitted with an infrared long pass filter (Schott Glass RG-9, Thorlabs, Newton, NJ), during a short period ($< 400\ ms$) of intense light saturating to photosynthesis ($> 10,000\ \mu mol\ photons\ m^{-2}s^{-1}$) provided by an array of white Cree LEDs (XMLAWT, 5700K color temperature, Digi-Key, Thief River Falls, MN) collimated using a 20 mm Carclo Lens (10003, LED Supply, Lakewood, CO). Chlorophyll *a* fluorescence was excited using monochromatic red LEDs (Everlight 625 nm, ELSH-F51R1-0LPNM-AR5R6, Digi-Key), collimated (minimizing the dispersion of light) using a Ledil reflector optic ($C11347\_REGINA$, Mouser Electronics, Mansfield, TX) and pulsed for 50 μs during a brief window when the white LEDs were electronically shuttered. In addition, a series of 5 images were also collected in the absence of the excitation light for artifact subtraction.

Infrared images were collected once every hour with the same camera and filter used for chlorophyll *a* fluorescence. Pulses of 940 nm light were provided by an array of OSRAM LEDs (SFH 4239, Digi-Key), collimated using a Polymer Optics lens (Part no. 170, Polymer Optics Ltd., Berkshire, England). Since 940 nm light does not influence plant development or drive photosynthesis, images were also collected during the night period [Eskins, 1992]. Note that the other modalities were captured only at daytime, so that they will not interfere with the circadian cycle of the plant.

Sets of 15 images were collected for averaging, in the absence of saturating illumination. As with chlorophyll *a* fluorescence, images were captured in the absence of 940 nm light for artifact subtraction.

3.2.2 RGB Color and Depth Images

The RGB color and depth images were collected using a Creative Senz3D sensor [Nguyen et al., 2015]. The sensor contains both a $1,280 \times 720$ color camera directed parallel to, and separated by roughly 25 mm from, a depth camera which has a resolution of 320×240 pixels. The color images are JPEG-compressed coming out of the sensor and so have block artifacts. The depth sensor uses a flash near IR illuminator and measures the time-of-flight of the beam at each pixel to obtain depth estimates at each pixel [Hansard et al., 2013].

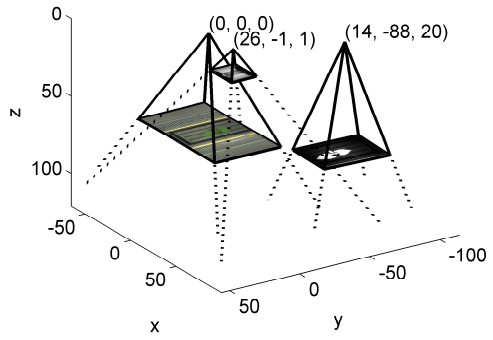
The depth sensor is limited largely by the strength of the signal returned. Low-reflectivity surfaces, such as the foam under the plant leaves, are accurate only at close range (on the order of 20 or 30 cm). On the other hand, specular and highly reflective objects, such metal and plastic surfaces in the growth chamber, can saturate the sensor pixels leading to unreliable depths for those pixels. Fortunately the primary goal of the depth sensor is to obtain depths of leaf pixels, and plants provide good, roughly Lambertian, reflections of IR [Chelle, 2006]. As a result the most useful and reliable depth pixels are those that fall on plant foliage, and we usually ignore non-foliage pixel depths.

3.2.3 Image collection

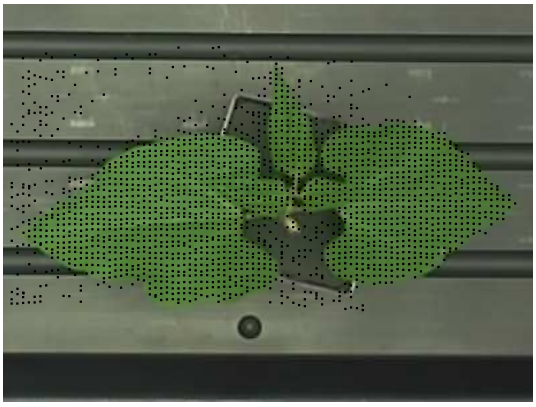
The imagery data, including fluorescence, IR reflectance, RGB color, and the 3D depth images, is collected once every hour. Five minutes before the end of each hour, 3D depth images and the color image were captured using the Creative Senz3D sensor (the depth points were transformed into the world coordinates and expressed in the unit of mm), followed by fluorescence and IR reflectance images collected sequentially at 2-minute interval by the IR filtered CCD camera. No substantial movements or growth were observed within the about 4-minute period required for image capture using all four modalities, which minimize the potential problems in image calibration.

3.3 Sensor Calibration

A planar checkerboard pattern was used to calibrate all three cameras to obtain both intrinsic and extrinsic parameters. While the grid pattern is not visible in the depth image, it is nevertheless observed in the reflected IR image whose pixels correspond to the depth pixels. This enables the use of Zhang's method [Zhang, 2000] to calculate the intrinsic parameters including a 2-parameter radial distortion of each camera. In this process the poses of all three cameras are also calculated relative to the checkerboard. The intrinsic and extrinsic parameters are stored as text files, and a Matlab function is provided that reads the parameters and can plot the camera poses as shown in Figure 3.



(*a*)



(*b*)

Fig. 3 (*a*)A plot of the three cameras showing their relative configuration and fields of view as obtained through calibration. Units are in *mm*. The distance to the target bean plant is roughly 620 *mm* in this example. For clarity the image planes are plotted not at actual target depth, where they would overlap, but at depths proportional to their focal lengths. The optical center of the color camera (left-most) defines the world coordinate system. Close to it is the depth camera having much lower resolution. The right-most camera is the combined fluorescent and IR camera. (*b*) 3D points from the depth camera are projected along their rays into the world coordinates and then projected into the color and fluorescent camera images. This shows the projection into the color image (with 90° rotation for economical use of space). Only points in a rectangular region around the plant in the depth image are selected, and it is further filtered by eliminating points with high standard deviation. An algorithm requiring only 3D points on the plant could select only those that fall on the leaf pixels.

3.3.1 Noise Characterization

The time of flight depth measurements can have significant noise, and it is useful to both model it and quantify it. Doing so can lead to strategies to reduce noise as well as providing guidance to algorithms that use the depth measurements. Our goal in this section is to provide a simple noise model that can predict the empirically

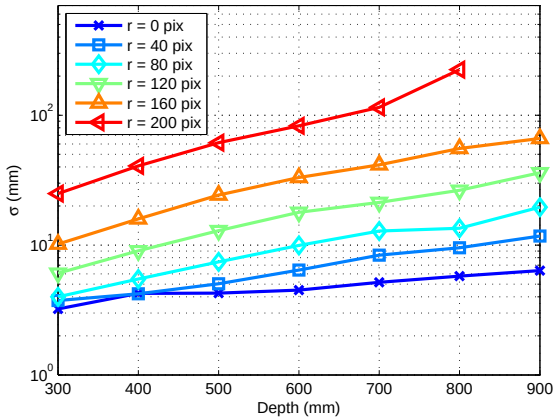


Fig. 4 Noise analysis for a depth camera obtained by imaging a flat surface at various depths. We found that the standard deviation, $\sigma_I(d, r)$ from Eq. (2) of the pixel depth measurements had a large dependency on both the target depth, d, and the pixel radius, r, from the image center, and these are plotted. A radius of 0 pixels is the image center, and of 200 pixels corresponds to the corners of the depth camera, which can be seen to have far larger standard deviation than at the image center for the same depth.

observed depth noise on smooth, Lambertian surfaces, such as plant leaves.

The depth noise ε is modeled as the sum of an image dependent term ε_I and a sensor dependent term ε_S:

$$\varepsilon = \varepsilon_I + \varepsilon_S. \tag{1}$$

The term ε_I is a random variable for each pixel with a value that varies between subsequent images taken from a fixed pose of a static scene. On the other hand, ε_S is a random variable for each pixel that models its depth offset, and its value only changes when the scene changes. The variance of ε_I is estimated for each pixel of a fixed scene observed over multiple images. In our experiments we observed a flat, uniform albedo (the ratio of reflected light to incident light) surface perpendicular to the camera at a sequence of depths, and for each depth acquired 300 images. Object depth, d, has a large impact on ε_I. For constant depth, we observed that the primary factor affecting the variance is the pixel radius r, from the image center. Physically we expect this dependence is due to a circularly symmetric illuminating beam closely aligned with the optical axis. Based on these observations we model $\sigma_I(d, r)$, the standard deviation of ε_I, as a function of depth and pixel radius. From our experiments we build up a lookup table for this as plotted in Figure 4.

Averaging over repeated images of a scene will not remove all the depth error as there are pixel depth offsets that are constant for images of the same scene. We model these with ε_S. To estimate its standard deviation σ_S, we first average over many depth measurement im-

Table 2 Summary of Arabidopsis and Bean databases. The "$\times n$" represents the number of modalities.

Plants	Subjects	Days	Images/Day	Total Images	Annotated Images
Arabidopsis	16	9	15	$2,160 \times 4$	576×4
Bean	5	5	13	325×4	175×2

Table 3 Plant image resolution of Arabidopsis and Bean databases, computed based on the yellow ROIs in Figure 1.

Plants	Fluorescence	IR	RGB	Depth
Arabidopsis	$\sim 240 \times 240$	$\sim 240 \times 240$	$\sim 120 \times 120$	$\sim 25 \times 25$
Bean	$1,000 \times 640$	$1,000 \times 640$	380×720	90×190

ages, in our case 300, to obtain pixel depth estimations that approximately eliminate the effect of ε_I. Then by calculating true ray depths on a known surface, in our case the observed plane, and further assuming that ε_S has the same standard deviation for all pixels, σ_S is obtained as the standard deviation of the error between averaged depths and known depths. In our experiments we obtained $\sigma_S = 6.5\ mm$ and found that it was insensitive to changes in depth.

The recorded depth images in the data collection are the result of averaging $N = 5$ subsequent depth images. Assuming independence of ε_I and ε_S, the variance of pixel depth measurements is given by:

$$\sigma^2(d, r) = \frac{\sigma_I^2(d, r)}{N} + \sigma_S^2. \quad (2)$$

There are additional sources of noise that are not modeled by this. Object albedo has an impact although this is fairly weak for strong signal reflections. Factors with large impact on signal noise include: object specularities, sharp variations in object albedo, mixed-depth pixels on object edges, and cases of very-low signal reflection, all of which can lead to very large variances. One of the utilities of having a model for variance is that it can be compared with the measured variance, and the difference used as a cue for portions of the scene that violate our modeling assumptions.

In addition, we noticed that the chamber light shades blocked some of the depth camera field of view, and in doing so reflected some of the IR illumination. This resulted in a small constant depth shift for the pixels. We measured this shift for each chamber experiment and provide it as an optional correction to the depth images.

4 Annotation, Files and Protocol

4.1 Data Statistics

MSU-PID includes two subsets, one for each plant type: Arabidopsis and bean. The statistic information of these two subsets are summarized in Table 2. The images were acquired every hour. As there is no light at night hours, plants cannot be imaged by the fluorescence and RGB color sensors while IR and depth cameras can still perform the capturing during the night. In order to make sure that all four modalities are present at the same time, we release the part of images captured only in the day hours, which are 15 images per day for Arabidopsis and 13 for bean for all four modalities.

The two subsets differ in image resolutions. As shown in Figure 1, we grow and image a single bean plant while a whole tray of Arabidopsis are grown at the same time. Therefore, the resolution of a Arabidopsis image is much lower than that of a bean image. We manually crop 16 Arabidopsis plants, which have been captured by all three cameras simultaneously. Table 3 summaries the image resolution of each plant in all four modalities.

The trade-off between image resolution and sample throughput is a common conundrum with designing visual phenotyping experiments. We chose to image a whole tray of Arabidopsis at lower resolution rather than an individual at higher resolution, because it better reflects high-throughput protocols, which in practice allow direct comparison of multiple genotypes in the same experiment. In contrast, we chose to image a single bean plant because we anticipate enabling development of 3D plant canopy reconstruction algorithms for bushy plants.

4.2 Manual Annotations

Part of the database is manually annotated to provide ground truth tip locations, leaf segmentation results, and leaf consistency overtime. We use the fluorescence images as the input for labeling because of their clean and uniform background.

For Arabidopsis images, we labeled 4 frames per day. While for bean images, we labeled 7 frames per day due to the plant's fast and spontaneous leaf movements. A Matlab GUI interface was developed for leaf labeling,

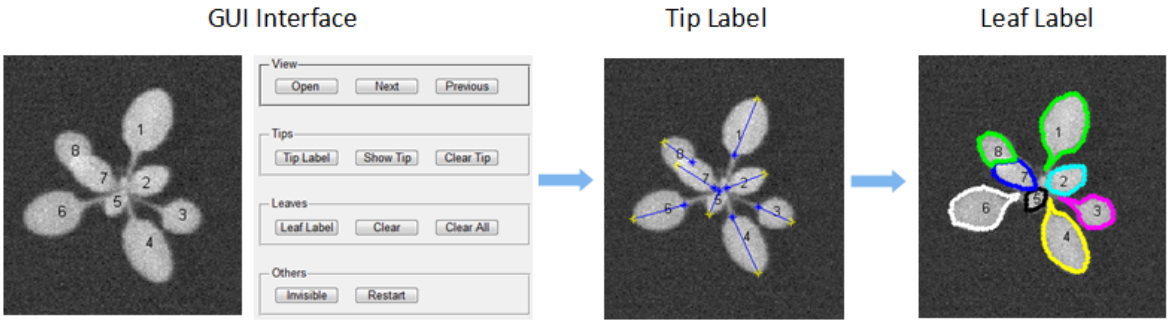


Fig. 5 Leaf annotation process, including leaf tip and leaf segment labels.

as shown in Figure 5, which will be released to the public. A user can open a plant image ("Open") to label the two tips and annotate each leaf segment. The results will be automatically saved once a user moves to a previous or next image ("Previous", and "Next"). For consistent annotation of the same leaf over time, we show a number on the center of each leaf indicating the order of labeling from the previous frame. The users should follow the order to label leaves.

The labeling of the leaf segment ("Leaf Label") is implemented by clicking the boundary of one leaf at each time. In order to provide more accurate labeling, a fairly high point density ($\sim$ 20 points on average) is typically used to define the boundary for each leaf. . The labeled leaf boundary is overlaid on the image for better visualization to guide the next action. An incorrect label can be deleted right after the labeling by clicking "Clear", or all leaf labels can be deleted by clicking "Clear All". This process continues until all leaf segments have been annotated. Once a leaf is invisible due to occlusion, the label can be skipped by clicking "Invisible".

The labeling of leaf tips ("Tip Label") is implemented by clicking pairs of points on the image. The outer tip is always clicked first before the inner tip. For visualization, a line connecting each pair of tips will be shown immediately after clicking the inner tip. Inaccurate labels can be deleted by clicking the right button of the mouse near the labeled point and relabeled by clicking the left button again immediately after deletion. All tips can be deleted by clicking "Clear Tip" and relabel again. The "Show Tip" option is to select whether to show the tips or not. For relabeling of both leaf segments and leaf tips on the current image, click "Restart".

After the labeling, we visually go through the results and correct inaccurate labels. One example of the labeling results for one plant is shown in Figure 9 (*b*), where one color is used to represent each specific leaf. As we can see during the transition between day 5 and day 6, there is one leaf showing up and covering up the leaf underneath, which disappears and will not be annotated later.

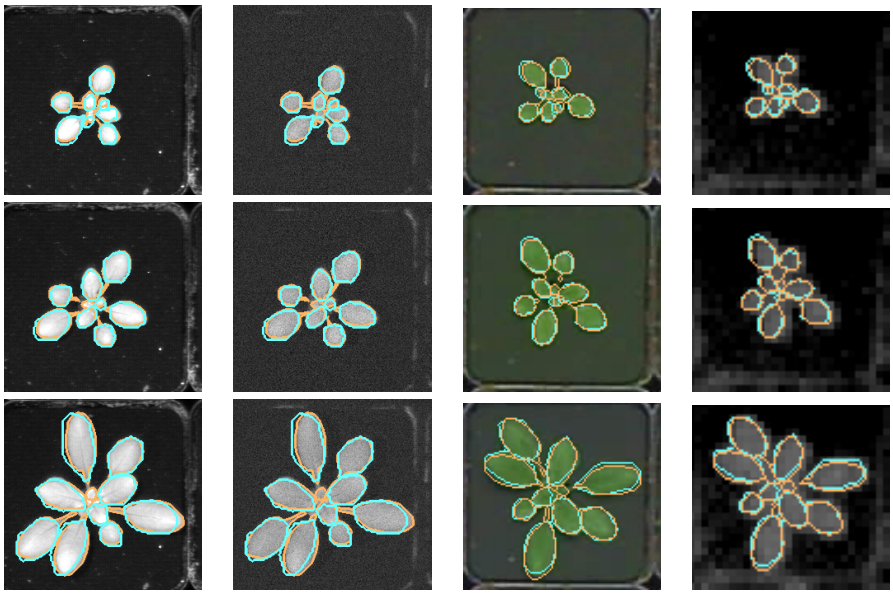
Note that one alternative approach for labeling leaf segments is to directly label the membership of superpixels instead of drawing a polygon along the boundary. Our experience is that since a noticeable percentage of extracted superpixels cover pixels of two neighboring leaves; the extra effort of breaking a super pixel into two makes it a less efficient alternative.

4.3 Multimodal Annotations

The leaf labeling can be propagated from the fluorescence images to each of the other modalities for the Arabidopsis sequences. To do this, we approximate a plant image with a plane, and estimate homographies that transform the fluorescence images into the images in other modalities. This provides a direct mapping of the pixel labels between each modality.

To quantify the precision of the label propagation, we manually label 3 images for each of 3 Arabidopsis plants (9 images in total) on fluorescence and RGB modalities. The label in each modality is propagated on other modalities. The result of one plant is illustrated in Figure 6. This mapping will introduce errors due to depth-based parallax. We use *SBD* score, which will be introduced in Section 4.5, to evaluate the similarity between the manual labels and the propagated labels. In order to compute the inter-annotators variability, we ask two different annotators to label the same 9 images and compute the *SBD* score. The results are shown in Table 4.

We can make two conclusions from Table 4. First, the similarity for both manual annotation and label propagation increases as the plant grows. This is expected because *SBD* depends on the overlap ratio between two leaves, which inherently favors large leaves. Second, the performance of label propagation is only



(a) (b) (c) (d)

Fig. 6 Label propagation between all four modalities (*a*) IR, (*b*) fluorescence, (*c*) RGB color, and (*d*) depth, of a sample plant in the Arabidopsis collection for day 3 (top row), day 5 (middle row), and day 9 (bottom row). In the dataset the manual segmentation is performed on the fluorescence images and is outlined here as orange lines propagated to all modalities (i.e., the orange lines in (b) are manual labels while these in other three modalities are propagated labels). The IR images are taken by the same camera and so will have exact pixel correspondence. To assess the segmented pixel propagation to other modalities, we also manually labeled a subset of the color plant images (shown as cyan in (c)) and propagated this label to other three modalities (i.e., the cyan in (a,b,d) are propagated labels). Comparing these boundaries in each modality gives a measure of the propagation errors due to parallax, and we provide quantitative analysis in Table 4. Note that the color and depth images are rotated 90 degrees.

Table 4 Human label performance vs. label propagation performance. L_1 and L_2 are the manual label results from two annotators, and L_t is the propagated results. The *SBD* score is averaged over 3 plants.

	fluorescence		RGB	
	L_1 vs. L_2	L_1 vs. L_t	L_1 vs. L_2	L_1 vs. L_t
day_3	0.808	0.804	0.827	0.802
day_5	0.830	0.836	0.871	0.837
day_9	0.903	0.886	0.877	0.789
average	0.847	0.842	0.858	0.809

slightly worse than the performance of human annotations. Therefore, we use label propagation to provide the labels for all four modalities for Arabidopsis sequences. There are two benefits: 1) it decreases the need for manual label; 2) labeling is consistent for all four modalities.

In the case of the bean imagery, the pixel association between modalities is more difficult as the within-plant depth variations are large. We found that a homograph-based mapping performed poorly, and so the manual annotations we supply apply to just the infrared and fluorescence images that use the same camera.

4.4 Name Conventions and File Types

We release training and testing sets in two separate folders. In each folder, there are two subfolders named Arabidopsis and Bean. The files in each subfolder have the following form:

- plant_ID_day_X_hour_YY_modality.png: the original images in each modality separately;
- plant_ID_day_X_hour_YY_label_modality.png: the labeled images of each modality if available;
- plant_ID_day_X_hour_YY_tips_modality.txt: the labeled tip locations of each modality if available;
- plant_ID_day_X_hour_YY_depthSigma.png: the depth standard deviation images;

where ID indicates the plant subject ID number (1 to 16 for Arabidopsis, 1 to 5 for bean), X is an integer indicating the date ($1-9$ for Arabidopsis, 1 to 5 for bean), and YY represents the hour index within a day ($9-23$ for Arabidopsis, $9-21$ for bean). For each combination of day and hour, we provide the original images in all four modalities (_rgb, _fmp, _ir, _depth) in PNG files. For annotated modalities, we have two additional files (_label, _tips) saving the annotation results. Leaf segmentation results are encoded as indexed PNG files,

where each leaf is assigned a unique and consistent leaf ID over time. Leaf ID starts from 1 and continuously increases till the total number of leaves. And the background is encoded as 0. Tips locations are saved in TXT files where each line has the following format:

– leaf ID tip1_x tip1_y tip2_x tip2_y

where leaf ID is an integer number that is consistent with the segmentation label in the _label file. tip1_x and tip1_y represent the coordinates of the outer tip point. tip2_x and tip2_y represent the coordinates of the inner tip point. Any "nan" value in the file indicates an invisible leaf.

The total storage of our database is around $380MB$, which is convenient for downloading via Internet.

4.5 Experimental Protocols

As shown in Table 1, MSU-PID can be used for applications such as leaf segmentation, leaf counting, leaf alignment, and leaf tracking. To facilitate future research, we separate the database into training set and testing set. 40% of the data is used for training and 60% for testing. Specifically, 6 plants of Arabidopsis and 2 plants of bean are selected for training. For fair comparison, both supervised learning and unsupervised learning methods should evaluate their performance on the training and testing sets separately.

The user may decide to utilize one or multiple modalities of the plant imagery for training and testing respectively. The availability of multiple modalities allows user to design novel experimental setup. For example, using RGB and depth modalities for training and RGB for testing can take advantage of additional information during the learning without incurring extra sensor cost during the testing, which can be implemented via either learning with side information [Chen et al., 2013], or transferring learning with missing modality [Ding et al., 2014, Chen & Liu, 2013].

Performance Metric To evaluate the performance of leaf segmentation, alignment, tracking, and counting, we use four performance metrics, whose Matlab implementations will be provided along with the data. Three of them are based on the tip-based error, which is defined as the average distance of a pair of estimated leaf tips $\hat{\boldsymbol{t}}_{1,2}$ with a pair of labeled leaf tips $\boldsymbol{t}_{1,2}$ normalized by the labeled leaf length:

$$e_{la}(\hat{\boldsymbol{t}}_{1,2}, \boldsymbol{t}_{1,2}) = \frac{||\hat{\boldsymbol{t}}_1 - \boldsymbol{t}_1||_2 + ||\hat{\boldsymbol{t}}_2 - \boldsymbol{t}_2||_2}{2||\boldsymbol{t}_1 - \boldsymbol{t}_2||_2}. \quad (3)$$

We build the frame-to-frame and video-to-video correspondence respectively and generate two sets of tip-based errors. More details can be find in [Yin et al., 2015]. We define a threshold τ to operate on the corresponding tip-based errors. By varying τ, we compute the first three metrics as follows:

– *Unmatched Leaf Rate (ULR)*, the percentage of unmatched leaves with respect to the total number of labeled leaves. This can attribute to two sources. The first is miss detections and false alarms. The second is matched leaves with tip-based errors larger than τ. When τ is large enough, this value is equal to the leaf counting error.
– *Landmark Error (LE)*, the average tip-based errors smaller than τ of all frame-to-frame correspondent leaves. This is used to measure the leaf tip alignment error.
– *Tracking Consistency (TC)*, the percentage of video-to-video correspondent leaves whose tip-based errors are smaller than τ. This is used to measure leaf-tracking accuracy.

In order to evaluate the leaf segmentation accuracy, we adopt an additional metric [Scharr et al., 2014] based on the Dice score of estimated segmentation results and ground truth labels:

– *Symmetric Best Dice (SBD)*, the average symmetric best Dice among all labeled leaves.

The Matlab function for computing *SBD* is provided by [Scharr et al., 2014].

5 Baseline Method and Performance

To facilitate future research on this database, we apply our automatic multi-leaf segmentation, alignment, and tracking framework [Yin et al., 2014a, Yin et al., 2014b] to the testing set of Arabidopsis imagery to provide a baseline. Specifically, we apply it on fluorescence and RGB modalities. Our work is motivated by Chamfer Matching (CM) technique [Barrow et al., 1977], which is used to align two edge maps. We extend it to simultaneously align multiple overlapping objects.

Note that [Yin et al., 2014a, Yin et al., 2014b] is designed for rosette plants like Arabidopsis. Therefore, it will not be applied to bean imagery, as it does not belong to rosette plant.

5.1 Multi-leaf Segmentation and Tracking Framework

As shown in Figure 7, the input of this framework is a plant video and a set of predefined templates with various shapes, scales, and orientations. We treat all

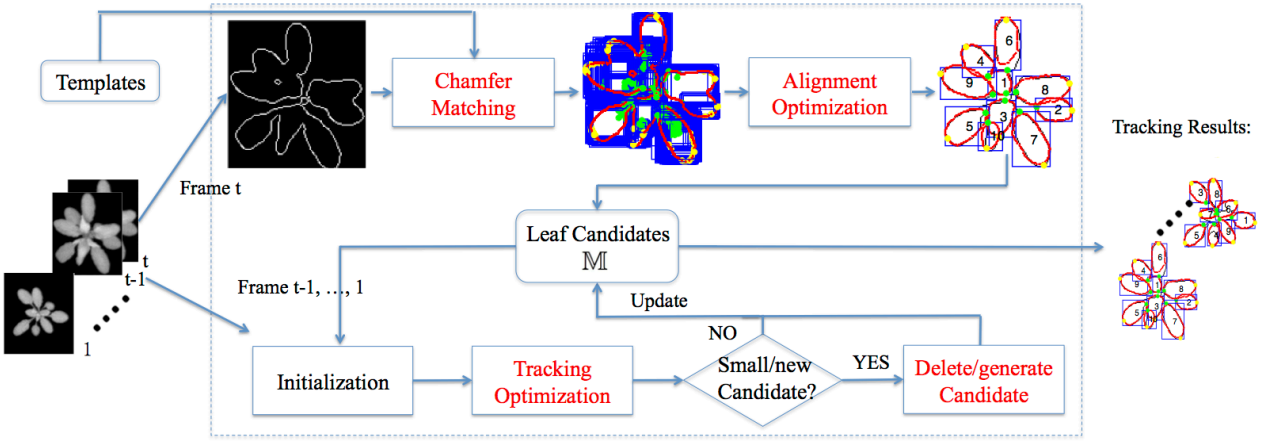


Fig. 7 Overview of the baseline method.

images in 9 days as a video from first image on the first day to the last image on the last day. To generate the template set, we first select 12 templates with different aspect ratios from the labeled images in the training set together with the corresponding tip locations. For each template, we scale it to 10 different sizes in order to cover the entire range of leaf sizes in the database. The scales for fluorescence images and RGB images are different due to the different image resolutions. For each scale template, we rotate every 15° to generate 24 templates at different orientations. Tip locations are scaled and rotated accordingly. Finally, we generate 2,880 leaf templates for each modality.

For plant segmentation, we use simple thresholding process and edge detection to generate an image mask and edge map. The best threshold is learnt from the training set, which is done by tuning the threshold in a certain range and find the best one by evaluating the overlap of the segmented masks with the ground truth label masks. The edge map and mask are used in the alignment and tracking optimization.

First, we find the best location of each template in the edge map that has the minimal CM distance, which will result in an over-completed set of leaf candidates from all templates. Second, we apply multi-leaf alignment [Yin et al., 2014a] approach to find an optimal set of leaf candidates on the last frame of the video, which will provide the information of the number of leaves, tip locations, and boundary of each leaf. Third, we apply multi-leaf tracking [Yin et al., 2014b] approach, which is based on leaf template transformation, to track leaves between continuous two frames.

In the tracking process, we delete a leaf when it becomes too small. A new leaf is generated when there is a relatively large portion of the mask that has not been covered by current leaf candidates. For each frame of the video, we can generate the tip locations for each estimated leaf and a label image with each leaf being labeled with one color. The labeled color for each leaf in the video remains the same during the tracking process.

5.2 Performance and Analysis

We apply our algorithm to all frames of each video and evaluate the performance on labeled frames. Figure 8 shows some examples of leaf alignment results on the last frame of each video. Our framework works very well on segmenting large leaves with no overlap to neighbor leaves. For overlapping leaves, it becomes more challenging as the edges in the overlapping area are more difficult to be detected. However, when the overlapping leaves are further away from the center, they will have a higher chance to be detected as shown in Figure 8 (c). When the overlapping leaves are close to the center, smaller leaves will be covered by larger leaves as shown in Figure 8 $(a), (d), (e)$.

Figure 9 shows the leaf tracking result of one Arabidopsis video in both fluorescence and RGB modalities. We can see the high quality leaf label propagation results on this video. The leaf template transformation works well for most of the leaves. As plant grows, younger leaves may grow faster than older leaves and occlude the older leaves. As shown in Figure 9 (b),(e), purple leaf replaces the red leaf at day 6. The two leaves are still being considered as one leaf (ID 8) in day 4 and day 3 (Figure 9 (c)). Leaf 8 in day 1 and 2 is a leaf ID switch w.r.t. the purple ground truth leaf and will not be considered as a consistently tracked leaf (Figure 9 (c)). However, they are still evaluated as well aligned and segmented leaves.

For quantitative evaluation, we vary τ from 0 to 1 and generate the first three evaluation metrics, as

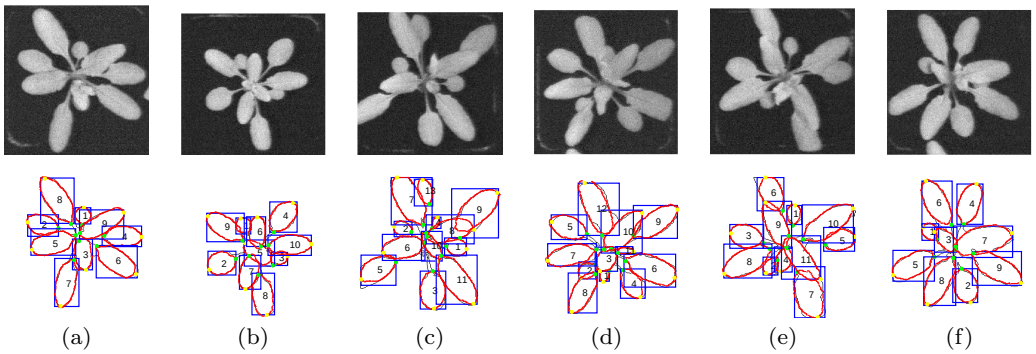


Fig. 8 Leaf alignment results on the last frame of 6 fluorescence Arabidopsis videos. First row shows the original images. Second row shows the alignment results with red points denoting the boundaries of the leaf templates in the blue bounding boxes. The numbers on the leaves are the leaf IDs representing the order of the leaf being selected and will be consistent during tracking.

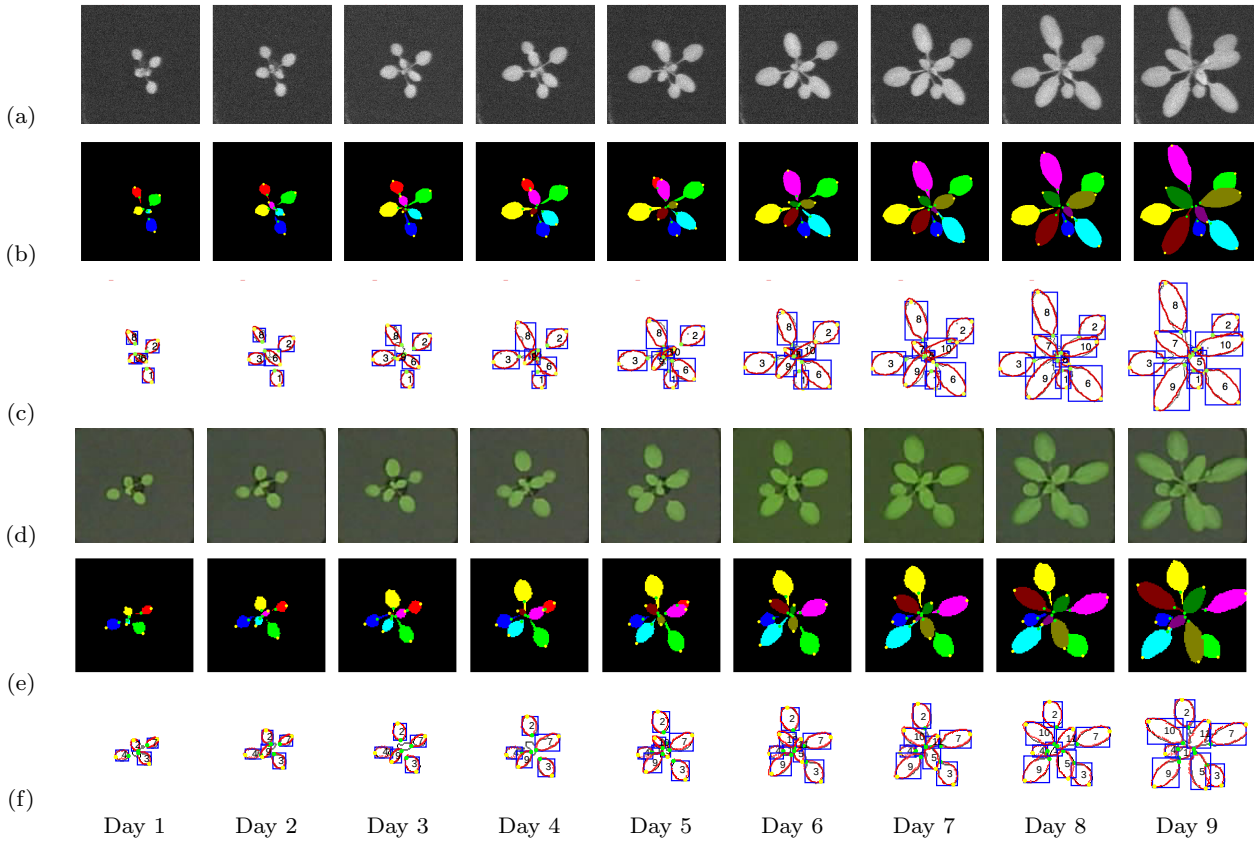


Fig. 9 Tracking result for plant 16 with a first frame for each day. (*a*) Example frames in fluorescence modality. (*b*) Manual leaf label results overlaid with tip locations. (*c*) Leaf tracking results on fluorescence images. (*d*) Example frames in RGB modality. (*e*) Label propagation from fluorescence images to RGB images. (*f*) Leaf tracking results on RGB images.

shown in Figure 10. For both RGB and fluorescence modalities, *ULR* decreases as τ increases as more leaves are being considered as matched leaves. As τ keeps increasing, *ULR* approaches a constant value, which is the different number in leaf counting that results from both miss detection and false alarms. *LE* increases as τ increases as it includes leaves with larger tip-based errors for averaging. *TC* increases as τ increases as more leaves are being considered as correctly tracked leaves. Note that *TC* is influenced by the length of frames evaluated for each video. As the longer frames we evaluate, the higher chance tracking will fail. Our method can detect 87% and track 50% of all labeled leaves with less than 20% average tip-based errors on fluorescence images.

We generate a label image for each frame based on the leaf segmentation results and compute the *SBD*

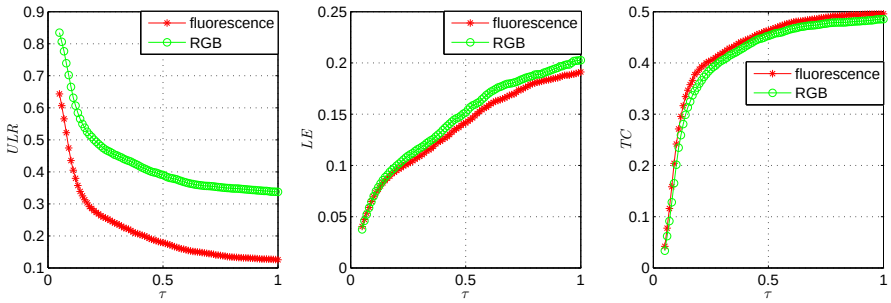


Fig. 10 Performance of the baseline method on the testing set of the fluorescence and RGB modalities of Arabidopsis plant.

score for each labeled image. The average *SBD* is 0.61 for fluorescence images and 0.57 for RGB images.

Overall, the performance on fluorescence images is better than the performance on RGB images. This is due to two reasons. First, the resolution of RGB images is much lower than that of fluorescence images, which will results in more miss detections of small leaves, and more severe leaf occlusions. Second, the label propagation causes some error, which is inevitable with the label homography-based mapping.

6 Conclusion and Discussion

This paper presents a newly collected multi-modality plant imagery dataset "MSU-PID". It has two subsets for Arabidopsis and bean plants respectively. Compared to existing databases in the field, MSU-PID uses multiple calibrated modalities including fluorescence, infrared, RGB color, and depth. Detailed image capture process and camera calibration are studied. We provide our manual labels about leaf tip locations, leaf segments, and leaf consistency over time on fluorescence modality. The labels are propagated to other modalities using homograph mapping for Arabidopsis imagery. Our annotations enable a wide variety of plant image analysis applications.

It should be obvious that all plants in MSU-PID belong to the same genotype and that no treatments were applied to induce phenotypic differences. This is because we have chosen to focus on a fundamental issue with visual phenotyping, i.e. accurate and automated identification and tracking of individual leaves over developmental time scales (days to weeks). The inherent challenge is that as leaves emerge and grow they change in size, position and shape and they may overlap or be overlapped by other leaves. Yet, as emphasized in the introduction, these factors may be important determinants for defining phenotypic differences between/among groups. In addition, development of these methods would invaluable for providing data to refine models of photosynthesis in plants with more complex canopies and in canopy systems.

For others to use our dataset, we have designed an experimental protocol with various evaluation metrics for different applications. To facilitate future research, we apply our automatic multi-leaf segmentation, alignment, and tracking algorithm on the fluorescence and RGB modalities of Arabidopsis imagery, where the labels for RGB modality is provided via label propagation. The experimental results on fluorescence and RGB images indicate that our dataset is very challenging.

We believe this new database will be beneficial to the research community in terms of algorithm development, performance evaluation, and identifying new research problems in plant image analysis. Furthermore, We are also open to suggestions and comments from the users of this database to further enhance our imaging setup and capturing protocol, so that we can develop new databases in the future.

References

[Baker, 2008] Baker, Neil R 2008. Chlorophyll fluorescence: a probe of photosynthesis in vivo. Annu. Rev. Plant Biol., 59:89–113.

[Barbosa et al., 2012] Barbosa, Igor Barros, Marco Cristani, Alessio Del Bue, Loris Bazzani, & Vittorio Murino 2012. Re-identification with RGB-D sensors. In First International Workshop on Re-Identification, pages 433–442.

[Barrow et al., 1977] Barrow, Harry G., Jay M. Tenenbaum, Robert C. Bolles, & Helen C. Wolf 1977. Parametric correspondence and Chamfer matching: Two new techniques for image matching. Technical report, DTIC Document.

[Butler et al., 1964] Butler, WL, SB Hendricks, & HW Siegelman 1964. Actton Spectra Of Phytochrome In Vitro. Photochemistry and Photobiology, 3(4):521–528.

[Chelle, 2006] Chelle, Michael 2006. Could plant leaves be treated as Lambertian surfaces in dense crop canopies to estimate light absorption? Ecological Modelling, 198(1):219 – 228.

[Chen et al., 2014] Chen, Dijun, Kerstin Neumann, Swetlana Friedel, Benjamin Kilian, Ming Chen, Thomas Altmann, & Christian Klukas 2014. Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. The Plant Cell, 26(12):4636–4655.

[Chen & Liu, 2013] Chen, Jixu, & Xiaoming Liu 2013. Transfer Learning with One-Class Data. Pattern Recognition Letters, 37:32–40.

[Chen et al., 2013] Chen, Jixu, Xiaoming Liu, & Siwei Lyu 2013. Boosting with side information. In Proc. Asian Conf. Computer Vision (ACCV), pages 563–577. Springer.

[Ding et al., 2014] Ding, Zhengming, Shao Ming, & Yun Fu 2014. Latent low-rank transfer subspace learning for missing modality recognition. In Proc. of the AAAI Conference on Artificial Intelligence (AAAI).

[Döös, 2002] Döös, Bo R 2002. Population growth and loss of arable land. Global Environmental Change, 12(4):303–311.

[Eskins, 1992] Eskins, Kenneth 1992. Light-quality effects on Arabidopsis development. Red, blue and far-red regulation of flowering and morphology. Physiologia Plantarum, 86(3):439–444.

[Fahlgren et al., 2015] Fahlgren, Noah, Malia A Gehan, & Ivan Baxter 2015. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. Current opinion in plant biology, 24:93–99.

[Furbank & Tester, 2011] Furbank, Robert T, & Mark Tester 2011. Phenomics–technologies to relieve the phenotyping bottleneck. Trends in plant science, 16(12):635–644.

[Hansard et al., 2013] Hansard, Miles, Seungkyu Lee, Ouk Choi, & Radu Horaud 2013. Time-of-Flight Cameras: Principles, Methods and Applications. Springer, New York, NY.

[Haug & Ostermann, 2014] Haug, Sebastian, & Jörn Ostermann 2014. A Crop/Weed Field Image Dataset for the Evaluation of Computer Vision Based Precision Agriculture Tasks. In Proc. European Conf. Computer Vision Workshops (ECCVW), pages 105–116. Springer.

[Houle et al., 2010] Houle, D, DR Govindaraju, & S Omholt 2010. Phenomics: the next challenge. Nature Review Genetics, 11(12):855–866.

[Huang et al., 2007] Huang, Gary B., Manu Ramesh, Tamara Berg, & Erik Learned-Miller 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, University of Massachusetts, Amherst.

[Johannsen, 1903] Johannsen, Wilhelm Ludwig 1903. Erblichkeit in Populationen und in reinen Linien. Gustav Fischer Verlag.

[Kramer et al., 2015] Kramer, David, Jeffrey Cruz, Christopher Hall, William Kent Kovac, & Robert Zegarac 2015. PLANT PHENOMETRICS SYSTEMS AND METHODS AND DEVICES RELATED THERETO. US Patent 20,150,204,787.

[Kumar et al., 2012] Kumar, Neeraj, Peter N. Belhumeur, Arijit Biswas, David W. Jacobs, W. John Kress, Ida C. Lopez, & João VB. Soares 2012. Leafsnap: A computer vision system for automatic plant species identification. In Proc. European Conf. Computer Vision (ECCV), pages 502–516. Springer.

[Lai et al., 2011] Lai, Kevin, Liefeng Bo, Xiaofeng Ren, & Dieter Fox 2011. A large-scale hierarchical multi-view rgb-d object dataset. In IEEE International Conference on Robotics and Automation (ICRA), pages 1817–1824.

[McClung, 2006] McClung, C Robertson 2006. Plant circadian rhythms. The Plant Cell, 18(4):792–803.

[Minervini et al., 2015] Minervini, Massimo, Hanno Scharr, & Sotirios A. Tsaftaris 2015. Image analysis: the new bottleneck in plant phenotyping. IEEE Signal Processing Mag., 32(4):126–131.

[Nguyen et al., 2015] Nguyen, VD, MT Chew, & S Demidenko 2015. Vietnamese sign language reader using Intel Creative Senz3D. In IEEE International Conference on Automation, Robotics and Applications (ICARA), pages 77–82.

[Oxborough, 2004] Oxborough, Kevin 2004. Imaging of chlorophyll a fluorescence: theoretical and practical aspects of an emerging technique for the monitoring of photosynthetic performance. Journal of Experimental Botany, 55(400):1195–1205.

[Phillips et al., 2000] Phillips, P. J., H. Moon, P. J. Rauss, & S. Rizvi 2000. The FERET evaluation methodology for face recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell., 22(10):1090–1104.

[Scharr et al., 2014] Scharr, Hanno, Massimo Minervini, Andreas Fischbach, & Sotirios A Tsaftaris 2014. Annotated image datasets of rosette plants. Technical Report FZJ-2014-03837.

[Schöttler et al., 2015] Schöttler, Mark Aurel, Szilvia Z Tóth, Alix Boulouis, & Sabine Kahlau 2015. Photosynthetic complex stoichiometry dynamics in higher plants: biogenesis, function, and turnover of ATP synthase and the cytochrome b 6 f complex. Journal of experimental botany, 66(9):2373–2400.

[Silberman & Fergus, 2011] Silberman, Nathan, & Rob Fergus 2011. Indoor scene segmentation using a structured light sensor. In IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pages 601–608.

[Söderkvist, 2001] Söderkvist, Oskar 2001. Computer vision classification of leaves from swedish trees. Master thesis, Linköping University.

[Sturm et al., 2012] Sturm, Jürgen, Nikolas Engelhard, Felix Endres, Wolfram Burgard, & Daniel Cremers 2012. A Benchmark for the Evaluation of RGB-D SLAM Systems. In Proceedings of the International Conference on Intelligent Robot Systems (IROS), pages 573–580.

[Sung et al., 2011] Sung, Jaeyong, Colin Ponce, Bart Selman, & Ashutosh Saxena 2011. Human Activity Detection from RGBD Images. CoRR, 64.

[Walter et al., 2015] Walter, Achim, Frank Liebisch, & Andreas Hund 2015. Plant phenotyping: from bean weighing to image analysis. Plant methods, 11(1):14.

[Woolley, 1971] Woolley, Joseph T 1971. Reflectance and transmittance of light by leaves. Plant physiology, 47(5):656–662.

[Wu et al., 2007] Wu, Stephen Gang, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang, & Qiao-Liang Xiang 2007. A leaf recognition algorithm for plant classification using probabilistic neural network. In IEEE International Symposium on Signal Processing and Information Technology, pages 11–16.

[Yin et al., 2014a] Yin, Xi, Xiaoming Liu, Jin Chen, & David M Kramer 2014a. Multi-Leaf Alignment from Fluorescence Plant Images. In IEEE Winter Conf. on Applications of Computer Vision (WACV), Steamboat Springs CO.

[Yin et al., 2014b] Yin, Xi, Xiaoming Liu, Jin Chen, & David M Kramer 2014b. Multi-Leaf Tracking from Fluorescence Plant Videos. In Proc. Int. Conf. Image Processing (ICIP), Paris, France.

[Yin et al., 2015] Yin, Xi, Xiaoming Liu, Jin Chen, & David M Kramer 2015. Joint Multi-Leaf Segmentation, Alignment, and Tracking from Fluorescence Plant Videos. arXiv:1505.00353.

[Zhang, 2000] Zhang, Zhengyou 2000. A flexible new technique for camera calibration. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(11):1330–1334.