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 Insert your abstract here. Include keywords, PACS and mathematical subject classification numbers as needed.

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

1 Introduction

As the growing of population in the world and the reduction of arable land, there is an increasing desire to improve the yield and quality of various crops, where the undersanding of biology and mechanism of plant growth is a key enabler. For this purpose, plant biologists create different mutations of the plant, grow them in either indoor chamber or outdoor field, visually observe them during the growth perid, and finally discover the patterns that associate the key factors (e.g., genre, environments) with the outcome (e.g., visual phenotype, yield) []. While many factors in this research procedure can be assessed quantitatively, which is necessary in order to perform large-scale study, one of the bottleneck is the automatic visual phenotyping.

* Equal contributor

** Contact author. E-mail: liuxm@cse.msu.edu

Jeffrey A Cruz, Jin Chen, David M Kramer
Department of Energy Plant Research Laboratory, Michigan State University
East Lansing, MI 48824, USA

Xi Yin, Xiaoming Liu
Department of Computer Science and Engineering, Michigan State University
East Lansing, MI 48824, USA

Saif Imran, Daniel Morris
Department of Electrical and Computer Engineering, Michigan State University
East Lansing, MI 48824, USA

Plant phenotyping aims to analyze and categorize the visual appearance of plants []. This is typically conducted via manual visual observation. With the increasing lower cost of imaging sensors and advances of Computer Vision technologies, image-based automatic plant phenotyping is growing into a desirable and viable solution []. In this interdisciplinary solution, computer vision scientists employ various imaging sensors to capture the plants and design advanced algorithms to automatically analyze the plant imagery, with the goal of answering the questions posed by plant biologists.

Plant image analysis is a non-trivial computer vision task, due to diverse variations of leaf appearance, overlapping and dynamics. In order to develop advanced 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], LFW [Huang et al., 2007], Caltech101 [Fei-Fei et al., 2004]). However, the publicly available database for plant phenotyping is very limited, with the only exception of LSC database [Scharr et al., 2014]. The LSC database has the limitations in that it only captures RGB images, and are suitable for a small set of applications.

To facilite 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, termed "MSU-PID". The MSU-PID includes the imagery of two types of plants (Arabidopsis and bean) captured by four types of sensors, Fluorescence, IR, RGB color, and depth. All four sensors are synchonized to perform the data capturing periodically for multiple days. Checkerboard-based camera calibration is perform between the multiple sensors, which results in the explicit correspondence between the pixels of any two modalities. For a subset of the database, we manually label the ground truth on the leaf identification number, leaf tip locations and leaf segments, using our labeling tool. To provide a performance baseline for future comparison, we apply our automatic leaf segmentation approach [Yin et al., 2014a, Yin et al., 2014b] to the Arabidopsis imagery and demonstrate the challenge of image analysis on this database.

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

- 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 fusion in plant image analysis.
- Our imaging setup and the variety of manu labele make MSU-PID an ideal candidate for evaluating a diverse set of plant image analysis tasks, including leaf segementation, leaf counting, leaf alignment, leaf tracking, leaf growth prediction, etc.

Table 1 Plant image databases, where the abbreviation in "Applications" column is defined as Leaf Classification (LC), Leaf Segmentation (LS), Leaf Counting (LO), Leaf Alignment (LA), and Leaf Tracking (LT).

Database	Modality	Appli-cations	Plant Type	Subject/ Classe #	Total Image #	Labeled Image #
Swedish leaf	Scaned leaf	LC	Swedish trees	15	1,125	1,125
Flavia	RGB	LC	Leaves	32	2,120	2,120
Leafsnap	RGB	LC	USA trees	184	29,107	29,107
Crop/weed	RGB	Weed det.	Crop/Weed	2	60	60
LSC	RGB	LS, LO	Arabidopsis	43	6287	201
			Tobacco	80	165,120	83
MSU-PID	Fluorescence, IR, RGB, Depth	LS, LO, LA, LT	Arabidopsis	20	*XXX*	*XXX*
			Bean	5	*XXX*	*XXX*

2 Prior Work

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

We summarize all existing publicly available databases that are related to plant imagery in Table 1. In terms of potential applications of the 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 consideraly larger and a neural network is used to train a leaf classifier. The most recent leafsnap project is an impressive effort that includes a very large dataset of 184 tree types [Kumar et al., 2012]. A mobilephone app 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 classifying crop vs. weed. Note that in this type of databases, normally only a single leaf is imaged and as a result, the challenging problem of leaf segmentation has been bypassed.

The second type of databases is for plant phenotyping where it is important to keep the plant live while imaged. Thus, no destructive action is taken and an entire plant is imaged. The LSC database [Scharr et al., 2014] is the most similar one to our database. It captures a large set of RGB images for the Arabidopsis and Tobacco plants. The provided manual labels allow the evaluation of leaf segmentation and leaf counting. In comparison, our MSU-PID database utilizes four sensing modalities in the data capturing, each providing different aspects of plant appearance. Our manual label also enable us to develop algorithm on additional applications such as leaf tracking and alignment.

3 Data Collection

3.1 Plants

Arabidopsis thaliana (ecotype Col-0) plants were grown at $20°C$, under a 16 hr:8 hr day night cycle with 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 daylight intensity set at $200 \mu mol\ photons\ m^{-2}s^{-1}$ and night and day temperatures of $18°C$ and $24°C$. Bean plants were watered with half-strength Hoaglands solution 3 times per week. In all cases, seeds were planted in soil covered with a black foam mask to minimize fluorescence background from algal growth. Two-week-old plants (*Arabidopsis* or bean) were transferred to imaging chambers and allowed to acclimate for 24 hours to the LED lighting before the start of the experiments. Growth conditions as described above were maintained for each set of plants for the duration of image collection.

3.2 Hardware Setup

Chlorophyll a fluorescence images were captured once per hour during the daylight period in a DEPI chamber [CITE]. A set of 5 images (for averaging) were captured using a Hitachi KP-F145GV CCD (Hitachi Kokusai Electric America Inc., Woodbury, NY) camera outfitted with an infrared long pass filter (Schott Glass RG-9, Thorlabs, Newton, NJ), during a short period ($< 400 msec$) 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 20mm Carclo Lens (10003, LED Supply, Lakewood, CO). Chlorophyll a fluorescence was excited using monochromatic red LEDs (Everlight 625nm, ELSH-F51R1-0LPNM-AR5R6, Digi-Key), collimated using a Ledil reflector optic ($C11347\_REGINA$, Mouser Electronics, Mansfield, TX) and pulsed for $50 \mu s$ during a brief window when the white LEDs were electronically shuttered. A series of 5 images were also collected in the absence of excitation light for artifact subtraction.

Infrared images were collected once per hour with the same camera and filter used for chlorophyll 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. 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.

[cite]Cruz, J.A. et al. Dynamic Environmental Photosynthetic Imaging (DEPI) Reveals Emergent Phenotypes Related to the Environmental Repsonses of Photosynthesis. Nature Biotechnology Submitted (2015).

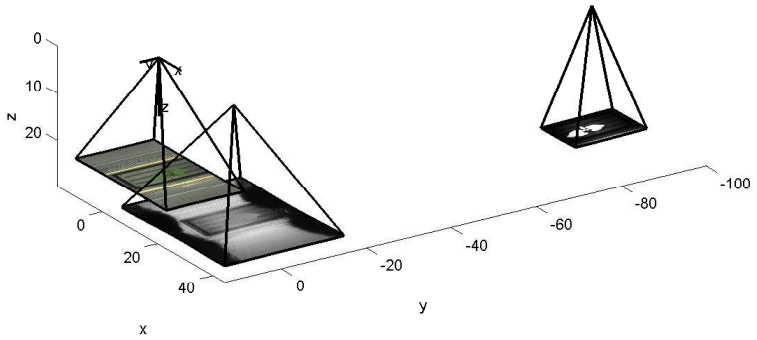


Fig. 1 A plot of the three cameras showing their relative configuration and fields of view as obtained through calibration. Units are in mm. The optical center of the color camera (on the left) defines the world coordinate system. Close to it is the depth camera. Its points are projected into the world coordinate system. On the right is the combined fluorescent and IR camera.

3.3 Image Acquisition

3.4 Image Calibration

3.5 Creative Senz3D

The depth and color images were collected using a Creative Senz3D sensor. This section describes characterizes the data from this sensor, particularly the depth data. The sensor contains both a 1280×720 color camera directed parallel to, and separated by roughly 25 mm from, a depth camera which has resolution of 320×240 pixels. The depth sensor uses a flash near IR illuminator and measures the time-of-flight of the beam at each pixel to obtain dense depth estimates along with an IR reflectance at each pixel.

There are a number of limitations to the depth sensor including sources of depth error. The primary measurement limit on the range-to-target is the strength of the reflected beam. Hence dark, matt surfaces are measured reliably only at close range on the order of 20 or 30cm. Highly reflective surfaces also pose problems with direct reflections leading to saturation and highly unreliable depths. In addition reflective surfaces at grazing angles are less reliably measured as little signal is reflected. Hence in our data portions of the chamber floor visible in Fig. 1 are highly reflective and have incorrect depths. Fortunately the primary goal of the depth measurements are to obtain leaf depths, and plants provide good, roughly Lambertian reflections of IR [Chelle, 2006]. Thus For these reasons the non-leaf depth pixels in the 3D data are unreliable and should be ignored. Another limitation is that the IR illuminator is slightly offset on the left of the sensor and this results in shadows to the right of some objects, as well as mixed pixels on depth discontinuities. Both of these can be readily detected as large standard deviations in the depth image.

3.5.1 Sensor Calibration

A checkerboard pattern was used to calibrate all three cameras to obtain both intrinsic and extrinsic parameters. While the checkerboard pattern is not visible as variations in depth, it is nevertheless observed as variations 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, as well as a calculation of their relative poses. The optical center of the color camera is used to define the world coordinates of our data. The depth values of the depth camera are projected along their pixel rays and then rotated and translated by the pose of the depth camera, and thus recorded as 3D points in the world coordinate system. Hence it is straight forward to project these points onto any of the three camera images.

3.5.2 Depth Bias and Noise Characterization

We characterized both the bias and the variance of the depth cameras as follows. A flat printed checkerboard with a surrounding white board was positioned at a large number of poses in front of the sensor. The pose of the checkerboard is calculated in each case using the color and IR reflectance images. This defines a plane relative to the depth camera which we use to calculate the ground truth depths for each pixel in the depth camera. At each pose we collect multiple depth images; this provides both a bias and variance measurement for each pixel at multiple depths.

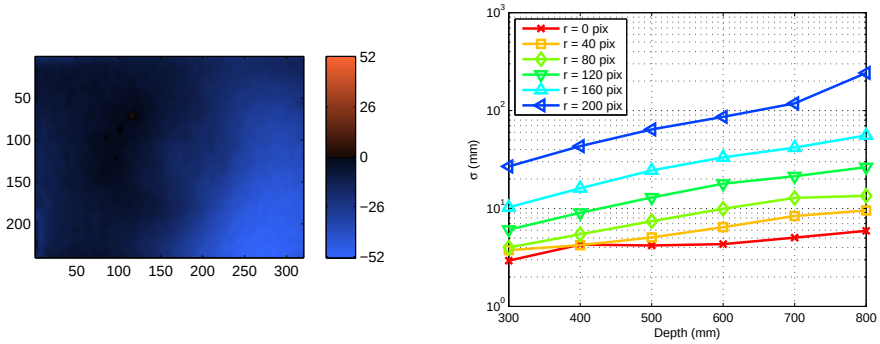
Next we sought to model the depth bias as a linear function of depth. Two parameters were fit for each pixel (a linear coefficient and an offset). We found that the bias was close to constant as a function of depth, although it varied across the depth image as shown in Fig. 2. The standard deviation of the pixel depth varied as a function of depth and roughly with the distance of the pixel from the optical center. Estimates of the noise is shown in Fig. 2.

In the recorded 3D data we subtracted our estimated bias, and averaged five depth images for each record. Hence the actual depth standard deviations for our data are $1/\sqrt{5}$ of the the standard deviations shown in Fig. 2 and the bias is zero.

Now 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 an additional bias shift which we measured and removed from the depth data.

3.5.3 Data Description

The data consist of 5 images per hour taken within 10 minutes of each other. These include the fluorescent image, the IR reflectance image with the same camera, the color image, the 3D-depth and a confidence image. The 3D-depth image is built from the depth sensor by transforming the points into world



(*a*) (*b*)

Fig. 2 Noise analysis for a depth camera. (*a*) Bias was fit and we found it to be roughly constant as a function of depth. This constant offset is shown for the camera pixels with red positive and blue negative offset in mm. Inside the chamber, the obstructions around the depth camera altered the bias somewhat and we re-calibrated the bias there. In all cases we subtracted the modeled constant bias from the depth images. (*b*) The standard deviation of the depth camera depends on depth as well as roughly on radius from the image optical center. A radius of 200 pixels corresponds to the corners of the depth camera which can be seen to have large noise.

coordinates and is expressed in mm. The confidence image is the standard deviation of the depths pixels. This is useful for identifying pixels at depth discontinuities which are unreliably detected and result in large standard deviations. In addition pixels with no response and saturated pixels are given the maximum standard deviation, and should be filtered out.

3.6 Image Accusation

3.7 Image Calibration

4 Image Labeling

Part of the dataset is manually annotated to provide ground truth tip locations, leaf segmentation results and leaf consistency overtime. Tip locations are saved in TXT file (TBD). Leaf segmentation results are stored in PNG images with one color for each leaf. The same color is used to represent the same leaf over a sequence of frames.

We use RGB images for labeling because of better visualization. All plants are cropped out from the background and saved as JPG files. For Arabidopsis images, we label four frames each day. While for bean images, we label seven frames each day because of faster leaf movement. We develop a Matlab GUI interface for leaf labeling, as shown in Fig 3. A user can open an image to label the two tips and annotate each leaf. The results will be automatically

saved once a user goes to label the next image. This GUI is used to annotate leaves in each frame. For consistent annotation of the same leaf over time, we use the labeled results of each frame to find the correspondence according to leaf centers. Incorrect correspondence will be manually corrected.

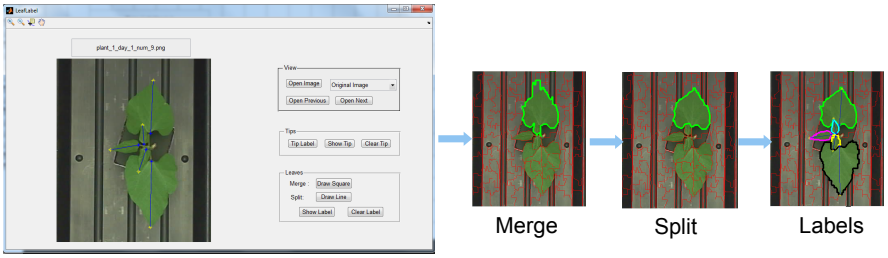


Fig. 3 Leaf labeling process, including tip labels and leaf annotation.

Tip label is implemented by clicking pairs of points on the image. The outer tip is always clicked before the inner tip. A line connecting each pair of tips will be shown immediately for visualization. Inaccurate labels can be deleted and relabeled.

For leaf annotation, we use the idea of merging and splitting super pixels. There are six different numbers of super pixels: 100, 200, 300, 500, 800, 1000. The user can specify which level to use depending on how well the super pixels can separate the leaves from the background. Because one leaf can be covered by several super pixels and one super pixel can across two leaves or the foreground and background. We allow merging and splitting super pixels. Merging is implemented by drawing a rectangle on the image. Every super pixel overlapping with this rectangle will be merged together. Splitting is implemented by drawing a line separating a leaf from the background or from another leaf.

As shown in Fig. 3, several super pixels that covers one leaf are first merged together to form a large super pixel. Since the top part of the super pixel covers some of the background and the bottom part covers another small leaf, two lines are drawn on this super pixel to split it. The leaf boundary is overlaid on the image for better visualization to guide the next action. This process continues until all leaves have been annotated.

5 Baseline Performance

6 Conclusion

References

[Chelle, 2006] Chelle, Michal 2006. Could plant leaves be treated as Lambertian surfaces in dense crop canopies to estimate light absorption? Ecological Modelling, 198(12):219 – 228.

[Fei-Fei et al., 2004] Fei-Fei, L., R. Fergus, & P. Perona 2004. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. In Proc. of the CVPR 2004 Workshop on Generative-Model Based Vision, pages 178–178.

[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.

[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.

[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.

[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.

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

[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 Signal Processing and Information Technology, 2007 IEEE International Symposium on, pages 11–16. IEEE.

[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.

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