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

Multi-modality Imagery Database for Plant Phenotyping

First Author · Second Author

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.

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

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

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 on plant imagery is very limited, with the only exception of [Scharr et al., 2014] and [Haug & Ostermann, 2014]. Both databases have the limitations in that they only capture RGB images, and are suitable for a small set of tasks.

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 a candidate dataset for a diverse set of plant image analysis tasks, including leaf segementation, leaf counting, leaf alignment, leaf tracking, leaf growth prediction, etc.

2 Prior Work

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 $940nm$ 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 $940nm$ 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 $940nm$ light for artifact subtraction.

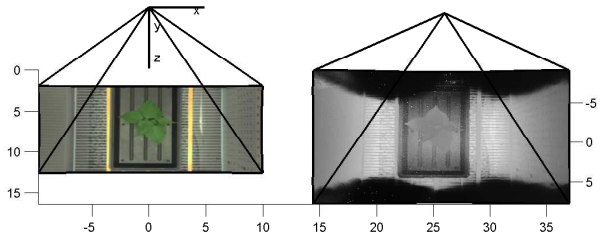

Fig. 1 A plot of the three cameras showing their relative configuration and fields of view as obtained through calibration. Units are in mm. In the center is the color camera, and its optical center defines the world coordinate system. On the right is the depth camera. Its points are projected into the world coordinate system. On the left is the fluorescent and IR camera.

[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).

3.3 Image Accusation

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 CITE. 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

Fig. 2 Noise and Bias model

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 CITE 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 [How many?] 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). The results are summarized in Fig. 2. [Need more summary here.]

In the recorded 3D data we subtracted our estimated bias, and averaged five depth images for each record. Hence the actual depth variance for our data is 1/5 of the the variances 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 a bias shift of [how much?]. We removed this bias shift 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

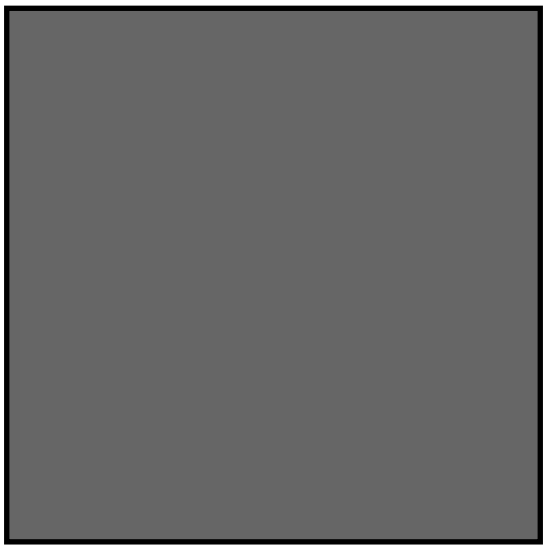

Fig. 3 Please write your figure caption here

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 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 marked as high standard deviation. [Got to check this].

3.6 Image Accusation

3.7 Image Calibration

4 Image Labeling

5 Baseline Performance

6 Conclusion

Paragraph headings Use paragraph headings as needed.

$$a^2 + b^2 = c^2 \tag{1}$$

References

[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

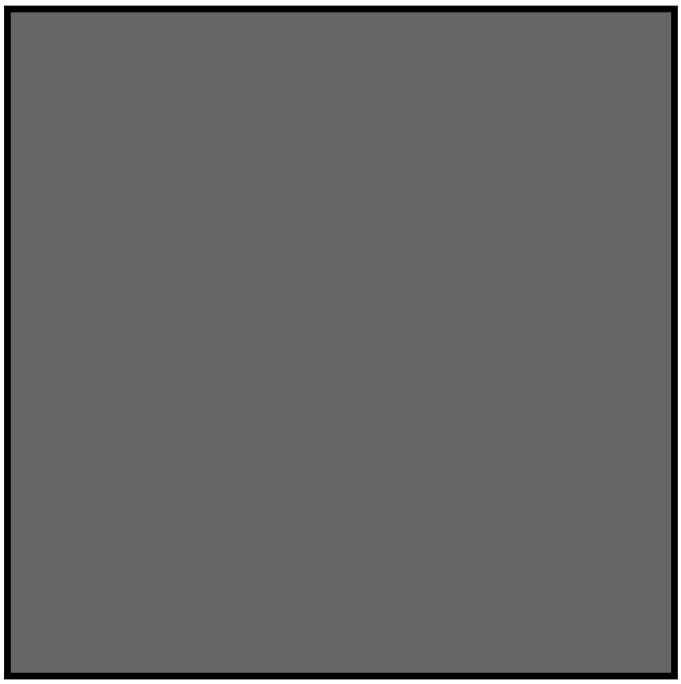

Fig. 4 Please write your figure caption here

Table 1 Please write your table caption here

first	second	third
number	number	number
number	number	number

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 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.

[Phillips et al., 2000] Phillips, P. J., H. Moon, P. J. Rauss, & S. Rizvi 2000. The FERET evaluation methodology for face recognition algorithms. IEEE T-PAMI, 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.

[Yin et al., 2014a] Yin, Xi, Xiaoming Liu, Jin Chen, & David M Kramer 2014a. Multi-leaf Alignment from Fluorescence Plant Images. In WACV.

[Yin et al., 2014b] Yin, Xi, Xiaoming Liu, Jin Chen, & David M Kramer 2014b. MULTI-LEAF TRACKING FROM FLUORESCENCE PLANT VIDEOS. In ICIP.