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

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, where the understanding of the genetic mechanisms to control plant growth is a key enabler [Döös, 2002]. For this purpose, plant scientists make various genetic mutant strains of plants, 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 discover plant morphological or physiological patterns that tightly associate with key growth factors [Houle et al., 2010].

* 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

While many factors can be assessed quantitatively, which is essential for high-throughput study, one of the bottleneck in this research line is plant visual phenotyping [Walter et al., 2015].

Plant visual phenotyping aims to analyze and categorize the visual appearance of plants []. In old days, this was conducted via manual visual observation [Erblichkeit, 1903]. Today, with the increasing lower cost of imaging sensors and advances of computer vision technologies, image-based automatic plant visual phenotyping is growing into a desirable and viable solution [Cruz et al., 2015]. In this interdisciplinary field, scientists employ various imaging sensors to capture plants and design advanced algorithms to automatically analyze the captured plant imagery, with the goal of raising testable biological hypotheses to solve the aforementioned problems.

Due to diverse variations of leaf appearance, layout, growth and movement, plant image analysis is a non-trivial computer vision task []. 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 still 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 is only suitable for a small set of applications.

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, termed "MSU-PID". The MSU-PID includes the imagery of two types of plants (Arabidopsis and bean), both are widely used in plant research, captured by four types of sensors, i.e. Fluorescence, IR, RGB color, and depth. All four sensors are synchronized and are programmed to periodically capture data 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. 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 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 fusion in plant image analysis.
- Our imaging setup and the variety of manual label make MSU-PID an ideal candidate for evaluating a diverse set of plant image analysis tasks,

including leaf segmentation, leaf counting, leaf alignment, leaf tracking, leaf growth prediction.

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 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 considerably 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 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 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 capture plant images without interfering the growth of plants. Thus, non-destructive approaches are taken and the 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 night and day temperatures of $18°C$ and $24°C$, and daylight intensity set at 200 $\mu mol\ photons\ m^{-2}s^{-1}$. Note that the bean plants were watered with half-strength Hoaglands solution 3 times per week.

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

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

In this section, we introduce the hardware used for capturing Fluorescence, IR, RGB color, and depth imagery data.

3.2.1 Fluorescence and IR images

Chlorophyll a fluorescence images were captured once per hour during the daylight period in a growth chamber [Cruz 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 ($< 400msec$) 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.

3.2.2 RGB color and depth images

The depth and RGB color images were collected using a Creative Senz3D sensor [Nguyen et al., 2015]. 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.

The imagery data were collected once per hour. 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 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.

3.3 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

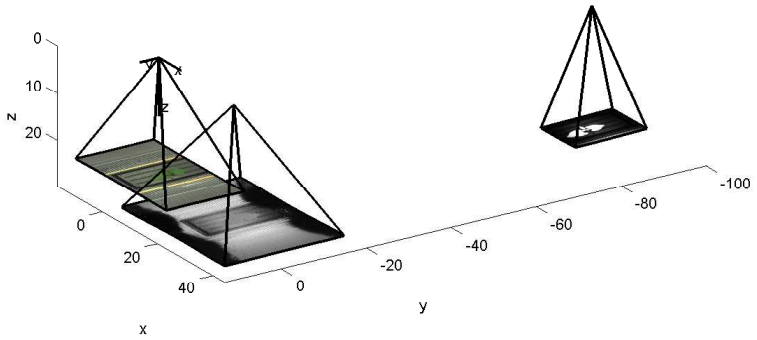


Fig. 1 A plot of the three cameras showing their relative configuration and fields of view as obtained through calibration. Units are in mm. The optical center of the color camera (on the left) defines the world coordinate system. Close to it is the depth camera. Its points are projected into the world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

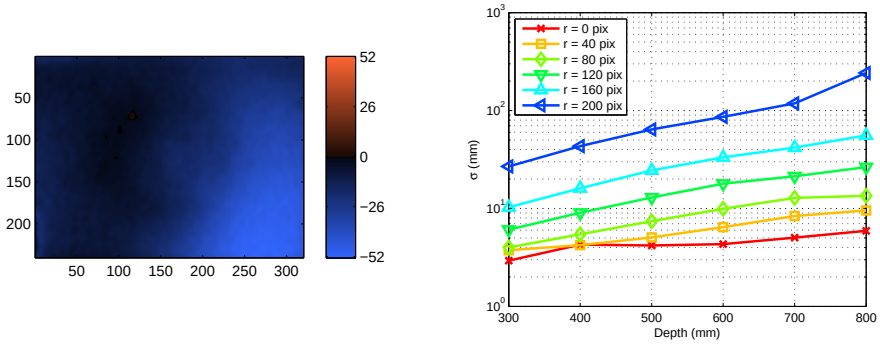
parameters including a 2-parameter radial distortion of each camera, as well as a calculation of their relative poses. The optical center of the color camera is used to define the world coordinates of our data. The depth values of the depth camera are projected along their pixel rays and then rotated 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.

3.3.1 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.



(*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.

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.

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 a TXT file for each frame. Leaf segmentation results are stored in a PNG image for each frame with one color for each leaf. The same color is used to represent the same leaf over a sequence of frames.

We use Fluorescence images for labeling because of clear background. 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. 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.

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

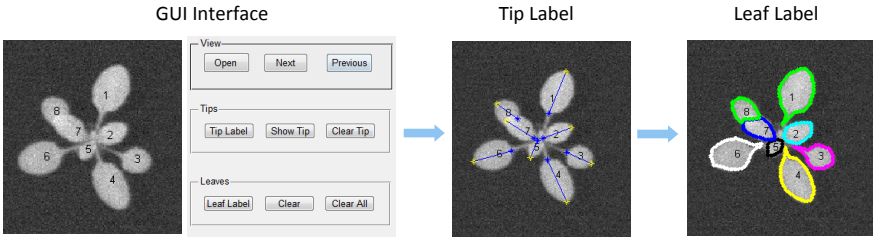


Fig. 3 Leaf labeling process, including tip labels and leaf annotation.

deleted by clicking the right button of the mouse near the labeled point and relabeled by clicking the left button again.

Leaf label is implemented by clicking the boundary of one leaf at each time. The leaf boundary is overlaid on the image for better visualization to guide the next action. Incorrect label can be deleted right after the labeling. This process continues until all leaves have been annotated. After the labeling of one plant, we visually go through the results and correct inaccurate labels. One example of leaf label result for one plant is shown in Fig. 4, where one color is used to represent one leaf. As we can see, there is leaf showing up and cover up the leaf underneath, which will not be annotated later (transition between day 4 and day 5).

Fig. 4 Leaf label results for one Arabidopsis plant over nine days with one image per day.

5 Baseline Performance

5.1 Our Framework

We apply our automatic leaf segmentation, alignment, and tracking framework [Yin et al., 2014a, Yin et al., 2014b] to the Arabidopsis imagery to provide a baseline for future research. Leaf segmentation aims at finding the correct number of leaves and the corresponding boundary of each leaf in each image. Leaf alignment is to align the two tips of each leaf. Leaf tracking is designed to track each leaf over time.

As shown in Fig 5, the input of this framework is a plant video and a set of predefined templates with various shapes, scales, and orientations. We also generate the two tips for each template for finding the corresponding tip

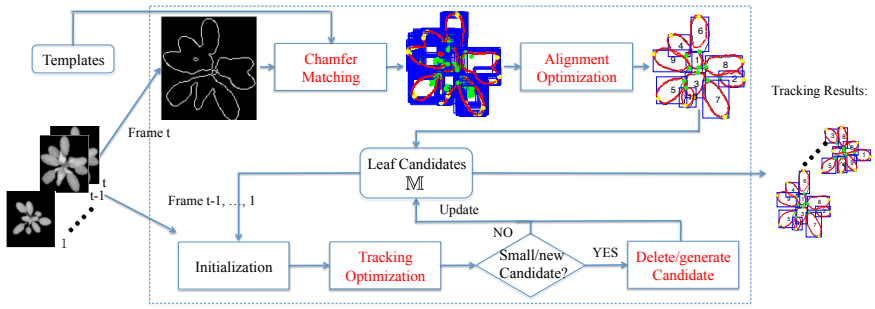


Fig. 5 Overview of the baseline method.

points of each leaf via Chamfer Matching First, 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 boundaries of each leaf. Second, 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 develop a procedure to generate new leaves and delete small leaves. For each frame of the video, we can generate a label image with each leaf being labeled with one color and the tip locations for each piece of leaf. The label for each leaf in the video maintain the same during tracking process.

5.2 Performance Evaluation

To evaluate the performance of leaf segmentation, alignment, and tracking, we use four criteria and provide the Matlab implementations. Three of them are based on tip-based error, which is defined as 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:

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

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 []; We define a threshold τ to operate on the corresponding tip-based errors. By varying τ, we compute the first three criteria as:

- *Unmatched Leaf Rate(ULR)*, the percentage of unmatched leaves w.r.t. the total number of labeled leaves. This can attribute to two sources. First, miss detections and false alarms. Second, matched leaves with tip-based errors larger than τ.
- *Landmark Error(LE)*, the average tip-based errors smaller than τ of all frame-to-frame correspondent leaves.

– *Tracking Consistency(TC)*, the average tip-based errors smaller than τ of all video-to-video correspondent leaves.

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 result and ground truth label:

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

The results of our algorithm by varying τ from 0 to 1 is shown in Fig.XXX. And the SBD score is XXX by averaging over all plants.

6 Conclusion

References

[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.

[Cruz et al., 2015] Cruz, J A, L J Savage, R Zegarac, W Kovac, C Hall, J Chen, & D M Kramer 2015. Dynamic Environmental Photosynthetic Imaging (DEPI) Reveals Emergent Phenotypes Related to the Environmental Responses of Photosynthesis. Nature Biotechnology, in revision.

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

[Erblichkeit, 1903] Erblichkeit, Johannsen W. 1903. Populationen und reinen Linien. Gustav Fischer Verlag.

[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.

[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.

[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.

[Nguyen et al., 2015] Nguyen, VD, MT Chew, & S Demidenko 2015. Vietnamese sign language reader using Intel Creative Senz3D. In Automation, Robotics and Applications (ICARA), 2015 6th International Conference on, pages 77–82. IEEE.

[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.

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

[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.