

Response letter on MVAP-D-15-00159, "Multi-modality Imagery Database for Plant Phenotyping"

July 14, 2015

We thank the Guest Editors for the efforts of handling our submission and two reviewers for providing the constructive and valuable feedback. The detailed comments help us further improve the quality of our manuscript. In the following, we list our detailed responses to all the comments. For your convenience, we add a separate reference list in the end that is used in this response letter.

REVIEWER 1

Comment 1: *The authors should clarify what is intended by "leaf alignment" and why it is important, since this feature is probably less common in plant phenotyping applications than the others that are mentioned.*

Response: Leaf alignment aims to estimate the structure of a leaf. In the case where the leaf structure is simple (e.g., ellipsoid-shaped leaf), leaf alignment amounts to estimating the inner and outer tips of the leaf. For more complex leaf structure (e.g., maple leaf), leaf alignment will estimate additional landmarks on the leaf contour that will consistently appear in all leaves. Leaf alignment is important for two reasons: First, it goes beyond leaf segmentation and provide structural information of a leaf - a discriminative signature for a leaf. For example, [1] uses a polygonal leaf model for leaf segmentation and plant identification. Second, plant biologists often hypothesize that different local parts of a leaf may have different photosynthetic efficiency. This leaf structure information can be used to quantitatively study this hypothesis and its implication to the overall photosynthetic efficiency of a plant.

Comment 2: *Figure 1 should be augmented (or a separate figure should be created) with zoom in details showing how the same plant parts (or scene portions) appear in the different modalities. This would help the reader get a better idea of the contrast and the type of information conveyed by each of the modalities adopted by the authors.*

Response: We have augmented Fig.1 by adding one zoom-in view for each image.

Comment 3: *Each of the databases mentioned in Table 1 should be accompanied by a reference.*

Response: We have the corresponding references in the prior work section, which follows the order as they appear in Table 2. To make it complete, we also added the reference to Table 2.

Comment 4: *The second paragraph of Sec. 2 begins with "we summarize all existing publicly available databases that are related to plant imagery". This is a strong statement, particularly because other plant related image datasets exist (e.g., ImageCLEF, ICL leaf database) which are omitted from the review. The authors should adjust language and/or expand the literature review.*

Response: We adjust language by the statement of “we summarize existing publicly available databases that are most related to our work”.

Comment 5: *In the second paragraph in Sec. 2, I believe the observation on single leaves imaged in a constrained environment does not apply to the dataset by (Haug and Ostermann, 2014). Please reformulate.*

Response: It is correct that the dataset (Haug and Ostermann, 2014) is not about single leaves imaged in a constrained environment. We still category it in the first type because it does not include leaf segmentation. All segmentation tasks are based on plant level. We have reformulated this in the second paragraph in Sec. 2.

Comment 6: *At the end of the first paragraph of Sec. 4.4, p. 8, the authors envision a use of the multiple modalities in their dataset in which algorithms are developed to handle missing modalities. Briefly presenting a use case scenario or citing relevant works may help clarify the importance of this point.*

Response: In this use case scenario, the training part of leaf segmentation algorithms may have both the RGB and depth modalities, while during the testing, we may only have the RGB modality available. This scenario is useful because it leverages additional information during the training, while does not increase the hardware/sensor cost during the testing. Specifically, we may use two types of approaches to implement this scenario. One is called learning with side information, which is the depth modality in this example (see [?]). The other is to use transfer learning that handles missing modality in the target domain (see [?]).

Comment 7: *In Sec. 5 no baseline method or result is reported for bean images. Could the authors comment on this?*

Response: The previous developed leaf alignment, segmentation, and tracking algorithm is designed particularly for rosette plants like Arabidopsis and tobacco. The rosette plant structure is used as a prior in the development of the algorithm. However, bean plant does not belong to rosette plants. Therefore, we did not apply the baseline method to bean plants. We have added this statement in the beginning of Sec 5.1.

Comment 8: *The conclusions (Sec. 6) should be expanded.*

Response: will expand after everything finished

Comment 9: *The acronym “MSU-PID” should be clearly defined when it first appears in text (p. 2, line 51)*

Response: Except in the Abstract, the “MSU-PID” first appears in fourth paragraph in Sec 1. We have clearly defined it as short for Michigan State University Plant Image Database.

Comment 10: *According to the SI (International System of Units), units of measure should be written in roman type, while italic type is reserved for variables. Besides, when they follow a number, a space should be included between numerical value and unit symbol. To improve clarity of the manuscript, the authors are therefore advised to adhere to the SI style conventions.*

Response:

Comment 11: *In Sec. 3.2.1, the "a" of "chlorophyll a" could be italicized to improve clarity.*

Response: We have changed "chlorophyll a" to "chlorophyll *a*" all over the paper.

Comment 12: *It is not clear on p. 8, line 15, if the size of the database is 380 megabytes (MB) or 380 megabits (Mb).*

Response: It is 380 MB. We have changed it in the text.

Comment 13: *In Eq. 3, p. 8, the authors should better define the symbols used. In particular, it is not clear if subscripts 1 and 2 refer to inner and outer leaf tips respectively, and in which order.*

Response: t_1 represents the outer leaf tip and t_2 represents the inner leaf tip. This representation is the same as the released leaf tip labels as illustrated in the end of the first paragraph in Sec. 4.3. The equation does not rely on the correspondence of inner or outer tips as long as the estimated tips correspond to the labeled tips.

Comment 14: *Sec. 5.1, p. 9, for completeness the authors should mention the approach they used to find a threshold for plant segmentation.*

Response: We use some training images with groundtruth plant segmentation masks to learn a threshold for binary segmentation. The learnt threshold is then applied to both training images and testing images to generate the image mask. We have added this statement to the paper.

REVIEWER 2

Comment 15: *The authors argue that the most closely related work is that of Scharr et al 2014. They identify as weaknesses that it i) uses only RGB images and ii) can be used for a few vision problems. However, upon reading the Scharr paper it is clear that the authors of that work, do mention that they collect additional data to be used in tracking context and for other applications. It is not clear if the authors here refer to what is available data from Scharr et al or what is described in the paper? Furthermore, the data in Scharr et al appear to originate from different mutants and under different treatments, and also imaged with different cameras. Since this work does use only wildtypes and single cultivars this distinction should be made. Furthermore, once leaf labeling is available several secondary annotations can be derived so in some sense Table 1 should be annotated accordingly. Nevertheless, I recommend the authors for arranging information as in Table 1.*

Response: We refer to Scharr et al 2014 that the Leaf Segmentation Challenge datasets only includes independent images without tracking context. The reason we do not include mutants is that we assume any computer vision algorithms that work on wildtypes should also work well on mutants as they have similar structure.

I do not understand what to be added to Table 1. TBD: Dr. Chen, please add some sentences about mutants in bean data. I recall we plan to capture additional bean data sets that have different mutants. Am I right?

Comment 16: *Your annotation process is completely interactive. Any way to "save interaction" by interactive segmentation approaches? I applaud your comment on the difficulty of merging super-pixel results, but I would add at least somewhere that easier ways to annotate will be beneficial.*

Response: Our annotation is not interactive. It is totally based on user input without any computer feedback except for visualization. Leaf label is implemented by clicking several points along the boundary of the leaf. Tip label is implemented by clicking the outer and inner tip points respectively. We did explore other options for labeling and found this simple naive way worked the best.

Comment 17: *It is unclear if you release this annotation tool? If not, please do... it will be extremely useful for the community.*

Response: We will release the labeling tool together with the dataset.

Comment 18: *Does the annotation _hour_YY_label.png contain a label for each pixel or only for the boundaries? If yes, then can you obtain single boundary definition in between overlapping leaves?*

Response: The annotated image is an image the same size as the original image. And each pixel is annotated with a number. The same number represents the same leaf and 0 represents the background. For overlapping leaves, we can visualize the relative location and annotate accordingly.

Comment 19: *Based on your annotations can you find which leaf occludes which?*

Response: During annotation, we can observe which leaf occludes which and annotate the overlapping area to the leaf on top. However, it is hard to refer which leaf occludes which from the annotations.

Comment 20: *The authors test an algorithm (developed by them) on the fluorescence part of the data. Granted this algorithm is presented elsewhere on the same data. I was wondering if the authors can either a) use the same algorithm on one of the other modalities, and how it would perform (assuming ground truth labeling obtained via propagation)? OR b) if they can apply another benchmark method even from the broad CV literature on one of the CV problems considered in another modality. I think doing one of the two, time permitting, will greatly assist the paper.*

Response: to be done: apply to RGB modality if label propagation works fine.

Comment 21: *Couple of typos:*

1. page 6, line 27 2nd col, please add "are" before present
2. page 7, lines 56 – 69, left col, the last two sentences should be together and not be separated by a period, but by a comma
3. page 8, line 30, left col, consider using performance instead of performances

Response: We have corrected those typos.

References

- [1] Guillaume Cerutti, Laure Tougne, Julien Mille, Antoine Vacavant, and Didier Coquin, "Understanding leaves in natural images—a model-based approach for tree species identification," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1482–1501, 2013.