Response letter on MVAP-D-15-00159R1,
"Multi-modality Imagery Database for Plant Phenotyping"

We thank the Guest Editors and two reviewers for providing feedback and the decision. Here we prepare the response to the comments and indicate the changes we have made in this revision.

Comment 1: *In Table* 1*, I believe LSC has wrong reference. It should be [Scharr et al.,* 2014*] and not [Haug and Osterman,* 2014*].*

Response: We have corrected the wrong reference.

Comment 2: *Also in Table* 1*, according to [Scharr et al.,* 2014*], the data they collected appear to handle at least also Leaf Tracking (maybe also leaf alignment). So the table should be corrected. While, from my understanding, the LSC dataset did not make those available (ie., where not released), from their paper it is clear that this was done. I believe it is best you are explicit on that rather than rely on reader's interpretation.*

Response: We have added "leaf tracking" to LSC dataset in Table 1 with a footnote indicating that the temporal data for leaf tracking is not available to the public. However, the dataset does not include labels which enables leaf alignment task.

Comment 3: *Having said the above, you could potentially edit the part on page* 3*, first col, lines* 20 *to* 24 *to read something along the lines "with the only exception of LSC database [Scharr et al.,* 2014*], which has the limitation of including a single modality (RGB only) and not having released with the LSC* 2014*, data related to tracking and alignment (although in [Scharr et al.,* 2014*], the authors claim that such data is available)"*

Response: We have changed the corresponding text to: "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)."

Comment 4: *Possibly also the text on page* 4*, first col, line* 25 *to* 34*, should be corrected to reflect the comment that [Scharr et al.,* 2014*] does in fact has data for other vision tasks but they were not included in the LSC. This should be clear to the reader, to reflect what is stated in the report.*

Response: We have change the corresponding text to "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"

Comment 5: *Thank you for your clarification on the Figure 3a, but i still believe that there is an error on the units of the z axis. It shows that cameras are at $60\ mm$, but it should $620\ mm$ according to your response and caption.*

Response: We have changed the caption slightly to clarify the confusion. "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."

Comment 6: *Based on how you made the annotations, I still do not think you can create a single pixel boundary between leaves (at best it will be two pixel thick).*

Response: We annotate each leaf separately. For those overlapping leaves, it is possible that we can still get a single pixel boundary if we happen to annotate the same points on the boundary. However, it is also likely that they will not share the boundary during annotation. So it will be two boundaries for each leaf and they can be connected or not. This is an approximation of the boundary and we think the error made by not a single pixel boundary between leaves can be ignored.

Comment 7: *Maybe my comment on interaction was not clear. Completely interactive in my book means totally with human interaction (not - automated). I was asking if you believe that some interactive segmentation algorithms can help.*

Response: Thanks for the clarification. We believe that there are some interactive segmentation algorithms, level set for example, can help on the segmentation. However, we did not explore the possible segmentation algorithms. Instead, we used totally human interaction because it was very accurate and end up without too much labor.