Morris Responses

July 15, 2015

Comment 1: *The authors do a nice work of calibrating the cameras and measuring noise in the depth camera. However, in page 2 introduction say that this calibration allows for the explicit correspondence between pixels of any modality. The authors rely on this to annotate data in one modality and propagate labels in the other (at least this is what I understand from later on description of methodology). However, this is a VERY strong assumption and depends completely on the distance between the cameras, the angles, the distance between object and sensors and the actual object arrangement. From our experience even when imaging co-planar plants (such as young arabidopsis) at a distance of 70cm even when the camera sensors are really close to each other (less than 5cm) some differences in view are there and occlusions are present. The authors should comment on this and should show as supportive evidence examples of plants at different growth stages in all 4 modalities in raw and annotated form to show how close this correspondence is matching and how the propagated labels. Furthermore, additional supporting evidence could be obtained either by arranging for two external and blind annotators to label some data (different age, different placement in the tray to show the effect of angle) in another modality (e.g., optical) and then measure inter-observer variability. Then they can propagate annotations from fluorescence to images of that modality and measure agreement. If this agreement is better than the in between rater variability then you could argue that propagating annotations is ok to do and actually beneficial. Nevertheless, you should definitely mention this limitation of your work.*

Response: write response

Comment 2: *Fig 3 (a) ... From Fig 2 it appears the distance from plant to sensor to be greater than 60mm (6cm) !!! ie., to me it looks close to 60cm, but either Fig 3 has wrong axis range or something else is going on. Can you please explain/update? Also same figure (3), shouldn't the images overlap? why are shown translated?*

Response: Yes, the distance to the plant is roughly 620mm. The image planes are plotted not at the plant location, but at a distance proportional to the focal length of each camera. We added dashed lines to the figure to make it clearer that these planes do not correspond to the plant depth, and updated the caption to explain it better.

Comment 3: *On manual annotation process: Did you use an extra annotator or a supervisor?*

Response: What does this question mean – what is a supervisor?

Comment 4: *Again going back to the problem of lack of exact 1-1 correspondence between modalities, how can you guarantee that labels on one are good on the other given such high differences in resolutions among modalities? Please include at least some visual examples.*

Response: response similar to above

Comment 5: *I do not subscribe to the term dense for such low resolution depth cameras. In my book dense refers to high res depth maps. Maybe the authors can reconsider the use of the term.*

Response: We agree and removed the term dense from the depth map image.