Reviewer 2

- I expect the authors to rewrite the algorithm section. Firstly, the overview should be more neat and concise. Each notation, function, and equation should be clearly defined, and including a notation table would be helpful. Currently, the algorithm is presented in separate sections, with the basic concepts in Section 3 and the key technical contributions in Section 4, making it hard to follow.

  A detail of the algorithm explaining the steps with their symbolic notation has been included in the supplementary file primarily due to space limitations in the main paper.

- The algorithm stage is poorly organized. The writing needs to be improved.

  The algorithm is reorganized in way that the contribution section is highlighted and rest of the explanations that describes the steps of the pipeline is moved to the supplementary file.

Reviewer 3

- The title of the paper does not accurately describe the content. The title describes plant phenotyping, while the dataset used in the evaluation section of the paper only contains maize and tomato, which are crops, and other plant types such as trees are not validated in the evaluation.

  Maize and tomato dataset is used for benchmarking plant phenotyping in many existing works. Our work also depends on such dataset for determining the quality of the registration. Tomato is considered as a type of plant.

- The abstract of the paper states that the use of iterative beam search for correspondence matching is advantageous for handling large graphs. However, in the latter part of the paper, neither a definition of "large graphs" nor an experimental comparison and analysis of the paper's method for graphs of different scales is provided.

  A description of the relative meaning of the large graph is added in Section 3. The large graph signifies the amount of nodes that can be handled during the generation of transition and emission cost. Viterbi algorithm tends to be computationally expensive for a graph with nodes around 50 whereas using the beam search, we can handle graph with nodes numbering 100-120.

- The reason for adopting the L1-medial skeletonization algorithm stated in paper 3.1 is that no significant advantages in adopting methods that introduce additional overhead and require segmentation approach.

However, the paper lacks a detailed explanation of the analytical process underlying this argument.

The process of skeletonization adopted by Chebrolu et al. involves multiple steps including segmentation using a machine learning algorithm, clustering to identify individual leaves or stems as distinct entities. Subsequently, they determined keypoints or skeleton nodes for each organ using self-organizing maps. The computational cost and impact on performance of such an involved procedure is obvious (and evidenced by our measurements in section 5) and (we think) needs no further elaboration. On the other hand, as we could demonstrate in section 5, employing much simpler and faster L1-medial skeleton extraction had no detrimental effect on registration quality at all.

We rephrased the part of section 3 where this is first introduced to make it clearer that this modification was a kind of ablation rather than a focused analysis.

- The paper lacks a clear justification for the use of the Schunck et al. study as a dataset.

    The dataset has been widely used for benchmarking and the same is quoted in the section 4.1 of the main paper. The choice of the dataset is attributed mainly to its widespread usage across different research papers focusing on plant phenotyping

- In paper 3.4, there is no explanation as to why the unknown parameters per node parameter for 3D affine transformation is 12.

    An explanation is added to the supplementary file elaborating that the 12 parameters include 3X3 rotation matrix and 3X1 translation vector.

- In addition to the comparison with the method of Chebrolu et al., it would be beneficial to compare the paper with the method of the last two years to enhance the persuasive effect of the article.

    We were unfortunately unable to include additional comparisons in time for the camera-ready version. More specifically, Daviet et al. (2022, *PhenoTrack3D*) could likely not be tested on the tomato plants of the dataset without modification because of their reliance on maize-specific traits. No source code seems to be publically available for Zhang et al. (2023, *Spatio-temporal registration of plants non-rigid 3-d structure*) and reimplementation was not feasible in the available time.

- The paper's depiction of the results is relatively clear, although it lacks a visual representation of the overall flow of the method.

An algorithmic pseudocode is added to the supplementary file (Section 1) to space constraints in the main paper. The pseudocode clearly emphasizes on the major steps included in the algorithm with respective input and corresponding output.

Reviewer 4

- The proposed method, while effective, appears to be computationally intensive and may require significant expertise to implement and fine-tune. This could limit its accessibility and practicality for broader applications without further simplification or optimization.

    The runtime comparison shows that our method consumes little time while performing different task in the overall pipeline. The phenotyping is performed generally in a greenhouse environment therefore such timings can still be used without being a limitation.

- The experimental validation, although comprehensive, is limited to maize and tomato plants. It would be beneficial to evaluate the method on a more diverse set of plant species to demonstrate its generalizability and robustness.

    The dataset used for comparison in our method is widely adopted across multiple research publications on phenotyping. In fact, we are not aware of any published method comparable to ours that evaluated on plant species other than maize or tomato, likely because they do demonstrate quite complex growth suitable for validating the result in a representative manner. On the other hand, we do agree that testing on more plant species would be beneficial for sure, but we are not aware of suitable datasets that offer ready-to-use 3D point cloud time series. Reconstructing them from monocular camera images like Daviet et al. (2022, *PhenoTrack3D*) is beyond the scope of our contribution. It is noteworthy though that even they evaluated on maize only despite being free of that limitation.

- While the paper compares the proposed method with a few existing techniques, a more extensive comparison with a broader range of state-of-the-art methods would strengthen the validation.

    [Reproduced from the answer to Reviewer 3]: We were unfortunately unable to include additional comparisons in time for the camera-ready version. More specifically, Daviet et al. (2022, *PhenoTrack3D*) could likely not be tested on the

tomato plants of the dataset without modification because of their reliance on maize-specific traits. No source code seems to be publically available for Zhang et al. (2023, *Spatio-temporal registration of plants non-rigid 3-d structure*) and reimplementation was not feasible in the available time.

- Explore potential optimizations to reduce the computational complexity and make the method more accessible for practical applications.

  The runtime comparison shows that our method consumes little time while performing different task in the overall pipeline. The phenotyping is performed generally in a greenhouse environment therefore such timings can still be used without being a limitation.