Reconstructing Watertight Meshes from Sparse Viewpoints of Crops

Harry Freeman (hfreeman), Gerard Maggiolino (gmaggiol), and David Russell (davidrus)
Carnegie Mellon University

April 3, 2022

1 Intro and Motivation

3D models of plants can be used to address a variety of problems in agriculture. Plant breeding programs are quickly adopting high-throughput phenotyping as an important step in informing the selection of the next generation [1]. This replaces laborious manual measurements which often prevent the whole plot from being surveyed, forcing decisions under greater uncertainty. Some factors, such as the leaf color, are relatively easy to assess from a single viewpoint. Others, such as the angle of a leaf to a stem or the surface area of a leaf, require explicit 3D reasoning. As described in [2], watertight meshes are important for many applications such as characterizing leaf area or plant volume or integration into existing agricultural simulation platforms. To the best of our knowledge, all prior approaches to this problem rely either on images which surround the plant in a controlled environment or active sensing such as Lidar to obtain high-fidelity results.

In this work, we seek to produce an accurate, watertight, and structurally-plausible mesh from few uncalibrated images of a single crop plant. We plan to do this by using neural surface reconstruction to reconstruct a mesh from sparse 2D images. In addition, we attempt to make our method generalizable by utilizing state of the art meta-learning techniques in order to create meshes for other crops with a small number of training images. Lastly, we aim to improve the quality of the mesh by applying a discriminator with the objective to distinguish rendered from real images. This would ultimately allow the ability to generate high quality meshes in unstructured fields where highly controlled data collection is impractical. It would also assist high-throughput phenotyping which can improve data-driven crop breeding outcomes for agricultural management.

We hope that additively combining approaches in each domain of sparse view reconstruction [3], meta-learning [4] [5] [6], and mesh refinement via discriminators [7] will qualitatively improve results for the challenging, non-regular meshes of plants. We can track the relative improvement of our combined approach by ablation of each component. While all of these approaches are established, attempting to combine them into one framework is novel.

2 Related Work

2.1 Geometry-only approaches to plant reconstruction

The work of Zhu et. al. [8] produces qualitatively good results from 60 calibrated images surrounding a single, isolated plant. They use SURF feature matching to generate a point cloud and then Poisson surface reconstruction to generate the mesh. Other work assumes a dense point cloud input to produce meshes [9, 2] or a structural model [10]. The work of Choudhury et. al. [11] uses segmented images to perform space carving and produce a voxelized representation. While this approach generates plausible results with relatively few images, it is coarse and produces artifacts such as detached leaves.

2.2 Reconstruction from limited viewpoints

Reconstructing a scene from a limited number of images is inherently challenging or ill-posed problem. There have been a number of approaches which seek to infer spatial structure from limited views using data-driven priors. The work of Ye et. al. [7] uses a GAN descriminator to enforce that novel views rendered from a 3D model a plausible.

Learned surface representations are explored in [12] on in the wild settings. While they demonstrate the effectiveness of surface rendering on sparse viewpoints, viewpoints surrounding the entire object are used.

2.3 Meta-learning

A model-agnostic meta-learning approach is presented by [4] that allows any gradient descent based model to be trained on new task with a small number of training instances. We will attempt to apply this work into our domain. Metalearning concepts have been applied specifically to NeRF models [6] to improve the speed of convergence and prediction quality for limited viewpoints. A task-informed meta-learning approach is presented by [5] which utilizes task-specific metadata to inform the transfer learning process for crop classification and yield estimation. This may be something we would like to explore time-permitting.

3 Methods

3.1 Data

We plan to primarily focus on the UNL-3DPPD [11] dataset, which contains 20 maize plants and 20 sorghum plants imaged from 10 viewpoints each. Each trial was replicated daily on four consecutive days. After completing baseline results with all available data, we will try to produce good results using a subset of frames and compare the performance to the pseudo-ground truth from all viewpoints. We will also conduct qualitative evaluation on data from unstructured environments, such as that collected in field trials by the Kantor Lab.

3.2 Algorithms

As a foundational approach, we plan to use NeuS [3] which produces an SDF directly from a set of posed images. This approach claims to deal well with thin surfraces which is imporant for accurately reconstituting leaves. To compute the pose, we will use a traditional structure from motion software such as ColMap [13]. We will initially run NeuS on all available viewpoints to compute a pseudo-groundtruth and then move onto the more challenging task of reconstrucing the plants from a subset of views from one direction.

Since the few-view reconstruction problem is very challenging and ill-posed for occluded regions, we consider two methods for improving it: the similarity between different instances of the same plant species and the fact that novel rendered views of object should look realistic. For both methods we make the realistic assumption that at test time we will have access to several images each of many different plants. We plan to leverage MAML to train a set of weights that can achieve good performance after a few training iterations on each of the individual tasks. We will then train a GAN on real images of the plant in question and use the descriminator of this network to enforce realism. This task may require images beyond our small 3D training set, which is reasonable since no additional labeling or metadata is required for these images.

Integrating the GAN into the system will be treated as a stretch goal. As a baseline, we propose to run experiments on novel rendered views and determine whether the discriminator's loss agrees with our qualitative assessment of novel-view realism. As an a stretch goal, we will extract a colored mesh from the SDF representation and use the approach described in [7] to further refine it using the realism loss from the GAN discriminator. Finally, as an even more ambitious stretch goal, we will seek to use the realism loss directly as an additional term in the NeuS optimization.

3.3 Evaluation

We will use the best results we obtained on all the viewpoints as a pseudo-groundtruth to evaluate the quality of our reconstruction. We will use Chamfer loss to assess the quality of the reconstructions and provide visualizations representing displaying the error per region. We may also consider task-specific metrics such as the thickness of the leaf or whether the result is a single connected mesh. Finally, due to the small size of the dataset, we will rely heavily on qualitative assessment.

References

[1] Eric Rodene et al. “A UAV-based high-throughput phenotyping approach to assess time-series nitrogen responses and identify trait-associated genetic components in maize”. In: *The Plant Phenome Journal* 5.1 (2022), pp. 1–12. ISSN: 2578-2703. DOI: 10.1002/ppj2.20030.

[2] Anjana Deva Prasad. “Deep implicit surface reconstruction of 3D plant geometry from point cloud”. In: (2022).

[3] Peng Wang et al. “NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction”. In: NeurIPS (2021). arXiv: 2106.10689. URL: http://arxiv.org/abs/2106.10689.

[4] Chelsea Finn, Pieter Abbeel, and Sergey Levine. “Model-agnostic meta-learning for fast adaptation of deep networks”. In: *34th International Conference on Machine Learning, ICML 2017* 3 (2017), pp. 1856–1868. arXiv: 1703.03400.

[5] Gabriel Tseng, Hannah Kerner, and David Rolnick. *TIML: Task-Informed Meta-Learning for Agriculture*. 2022. eprint: arXiv:2202.02124.

[6] Matthew Tancik et al. “Learned Initializations for Optimizing Coordinate-Based Neural Representations”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2021), pp. 2845–2854. ISSN: 10636919. DOI: 10.1109/CVPR46437.2021.00287. arXiv: 2012.02189.

[7] Yufei Ye, Shubham Tulsiani, and Abhinav Gupta. “Shelf-Supervised Mesh Prediction in the Wild”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2021), pp. 8839–8848. ISSN: 10636919. DOI: 10.1109/CVPR46437.2021.00873. arXiv: 2102.06195.

[8] Rongsheng Zhu et al. “Analysing the phenotype development of soybean plants using low-cost 3D reconstruction”. In: *Scientific Reports* 10.1 (2020), pp. 1–17. ISSN: 20452322. DOI: 10.1038/s41598-020-63720-2. URL: http://dx.doi.org/10.1038/s41598-020-63720-2.

[9] Riley M. Whebell et al. “Implicit reconstructions of thin leaf surfaces from large, noisy point clouds”. In: *Applied Mathematical Modelling* 98 (2021), pp. 416–434. ISSN: 0307904X. DOI: 10.1016/j.apm.2021.05.014. arXiv: 2009.10286.

[10] Paloma Sodhi et al. “Robust Plant Phenotyping via Model-Based Optimization”. In: *IEEE International Conference on Intelligent Robots and Systems* 1 (2018), pp. 7689–7696. ISSN: 21530866. DOI: 10.1109/IROS.2018.8594245.

[11] Sruti Das Choudhury et al. “Leveraging Image Analysis to Compute 3D Plant Phenotypes Based on Voxel-Grid Plant Reconstruction”. In: *Frontiers in Plant Science* 11.December (2020), pp. 1–18. ISSN: 1664462X. DOI: 10.3389/fpls.2020.521431.

[12] Jason Y. Zhang et al. “NeRS: Neural Reflectance Surfaces for Sparse-view 3D Reconstruction in the Wild”. In: NeurIPS (2021), pp. 1–18. arXiv: 2110.07604. URL: http://arxiv.org/abs/2110.07604.

[13] Johannes Lutz Schönberger and Jan-Michael Frahm. “Structure-from-Motion Revisited”. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.