Deep Learning to Estimate Homography Matrices for Maize Stand Counts from UAV Imagery

Dewi Endah Kharismawati, Hadi Ali Akbarpour, Rumana Aktar, Filiz Bunyak, Kannappan Palaniappan, and Toni Kazic

Dept. of Electrical Engineering and Computer Science, MU Plant Science Foundry, Missouri Maize Center, MU Informatics Institute
University of Missouri, Columbia

Abstract

Knowing how many plants have germinated in a farmer's or researcher's field is central to crop management and research. Counting a field's stand is commonly accomplished by walking the field and counting the plants in each row. This technique is labor intensive, slow, and error-prone. In this study, we want to develop a system that enables automated counting of early growth stages with a robust and accurate technology. Our approach is an image-based method using video from an Unmanned Aerial Vehicle (UAV) on maize fields. We address the problem in three parts: homography estimation with deep learning, mosaicking, and segmentation. Image mosaicking combines a large sequence of images to provide a wide field of view of the object of interest. Estimating the projective transformation between a consecutive pair of frames is a fundamental task in image mosaicking. We implemented an unsupervised deep learning algorithm that estimates the sequence of planar homography matrices of our corn field from imagery flown using a variety of trajectories and camera views. The algorithm does not use any vehicle or camera telemetry for matrix estimation. The algorithm performed faster than and with comparable accuracy as ASIFT, resulting in a faster mosaicking process. The mosaic of the whole field will next be registered with a Voronoi cell tiling to segment each plant from the soil for stand counting and to monitor plant growth.

Workflow

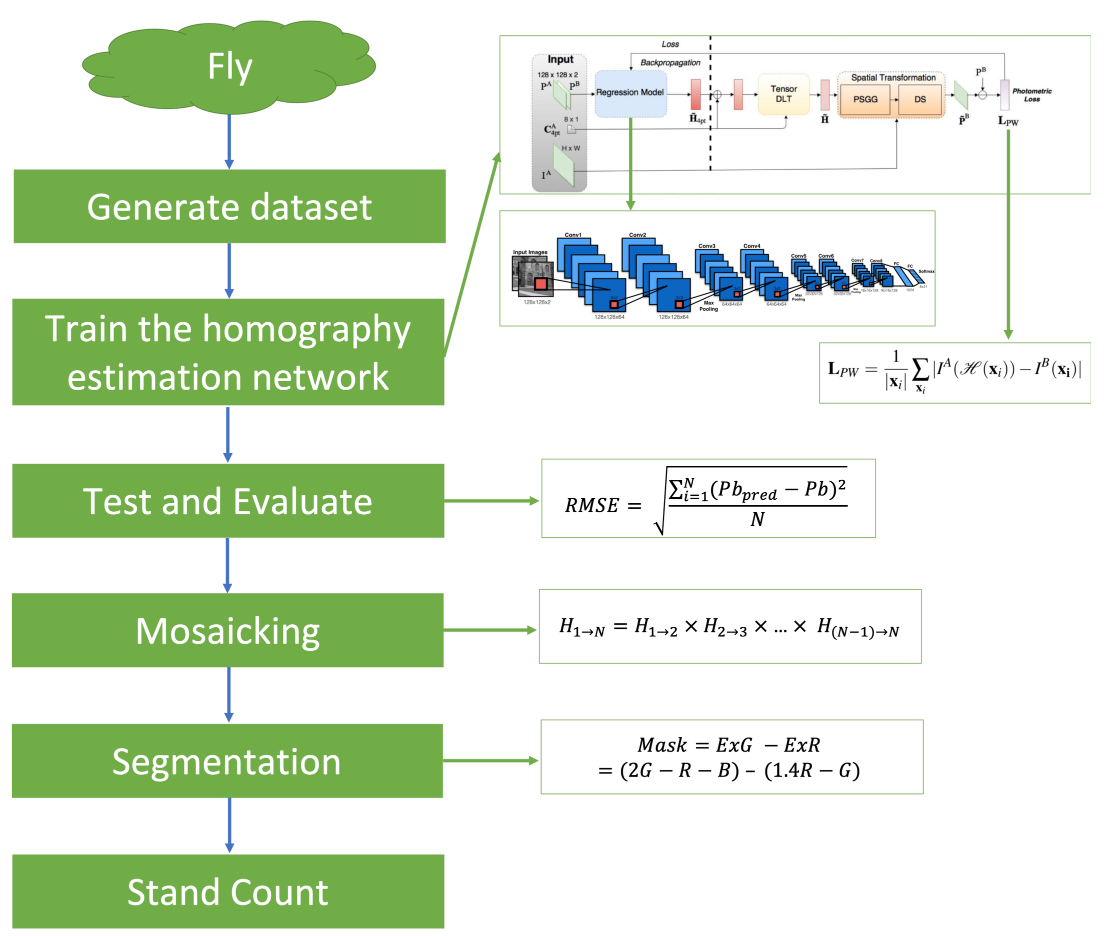

Fig. 1: Field operations are in clouds and code-supported process are in rectangles.

Model Training

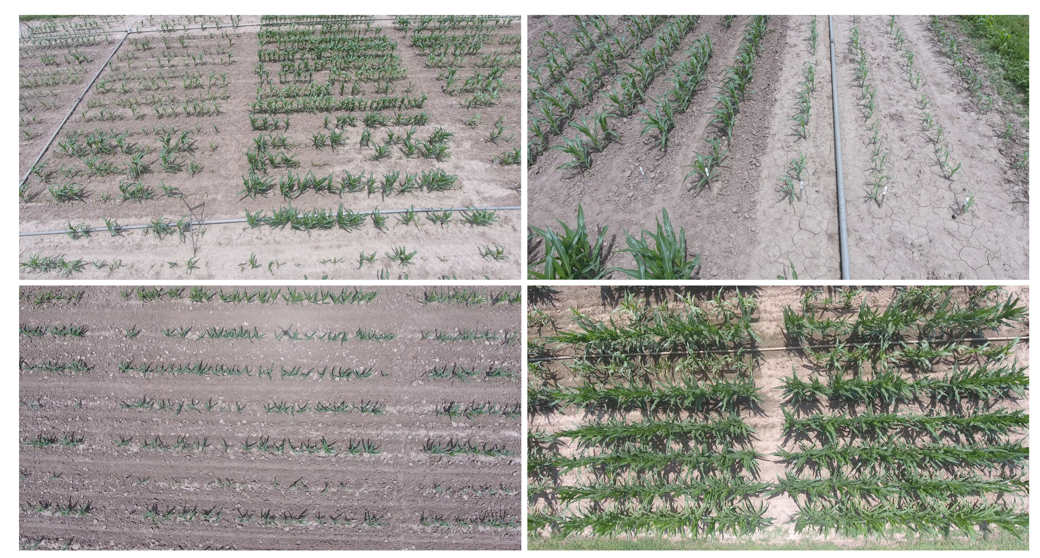

Fig. 2: Dataset Sample

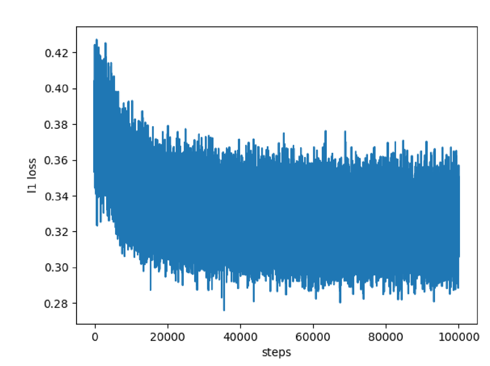

Fig. 3: Model training on nadir view dataset.

Homography Estimation

Fig. 4: Results on oblique view model, RMSE_left = 0.0807 RMSE_right = 38.29

Fig. 5: Good results on nadir view model, RMSE=2.385

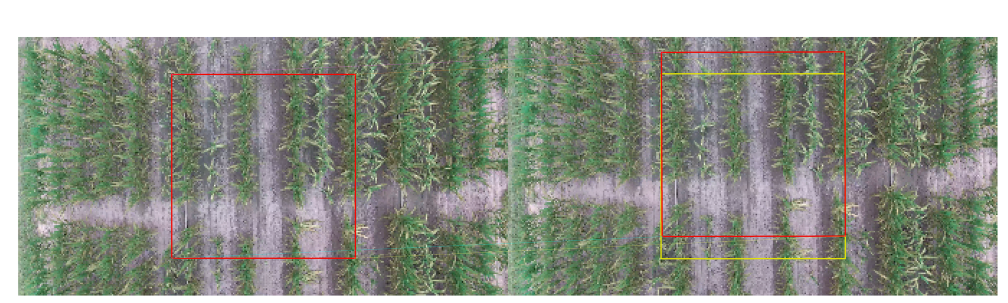

Fig. 6: Bad results on nadir view, RMSE = 130.1

Benchmarking

CNN	ASIFT
156.387 s	722.24 s

Fig. 7: Performance comparison between Deep Learning method and ASIFT feture based method in 500 test data.

Mosaicking

- Unsupervised deep homography algorithm estimates homography for a pair of concecutive frames.
- To mosaic all frames together, we need homography matrix for each frame with respect to frame 0.
- To get homography matrix of $H_0 \to n$, we need H_n! This requires high precision of matrix multiplication to minimize error. This method is prone to error accumulation.

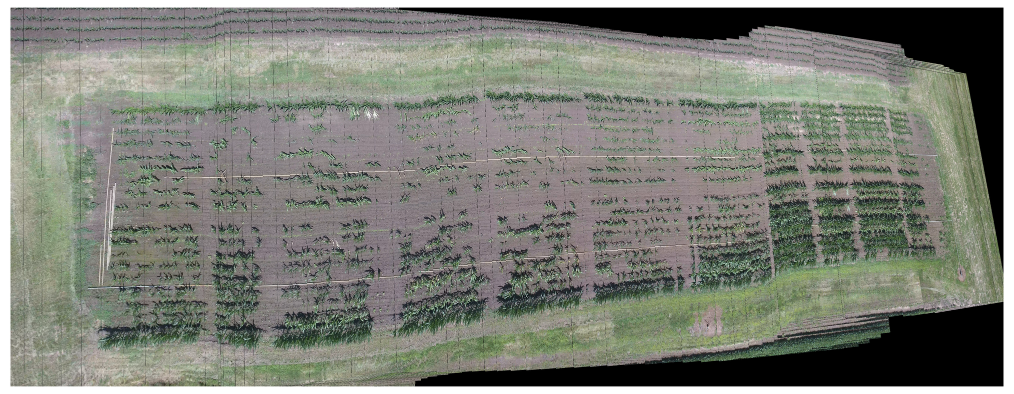

Fig. 8: Mosaic generated from aerial imagery on our maize field from 100 frames using deep learning homography matrices

Segmentation and Stand Counts

- To segment each plant from the soil, we use Excess Green vegetation - Excess Red [2].
- Extract foreground and background from segmented mask to get the Voronoi cell.
- Compare to manual stand count of plants in each row.

Challenges Encountered

- Validating the accuracy of homography matrices and mosaicking is hard to do because there are no good ground truth available. To validate the test data, we rely on our visual and ASIFT estimation as the ground truth.
- Maize imagery is more complex due to its homogenous appearance from above. Especially in the early growth stages, the network has difficulties extracting features. But this does not happen on more fully grown plants since they are more distinctive.
- We have trained different models for oblique and nadir view videos. On oblique videos, there is perpective on near and far distance view. The current network does not work well when the UAV travels parallel to the row, though its performance is improving as we accumulate more training data with different trajectories.
- The quality of training dataset influence the robustness of the network. The training data that was used to train models on nadir view has balanced trajectories in forward directions, but not backwards. This produces good results only for forward trajectories and high errors on other directions. Augmentating the training dataset by rotating pairs of frames 90, 180, and 270 degrees is insufficient because the orientation of one frame with respect to the next does not change. Instead, we are collecting training data with more trajectories and views.

Future Work

- Finishing building the training dataset that has balanced trajectories and directions of travel in different growth stages to increase the robustness of homography estimation.
- Improve precision of matrix multiplication to reduce error in mosaicking.
- Divide the video sequence into multiple groups and generate mini-mosaic from each group. These mini mosaics next are transformed to produce final mosaic [3]. This method reduces error accumulation from frame-to-frame registration.
- Register the Voronoi cells and stand counts for each row.
- Using deep learning to segment plantss.

Acknowledgments

We are grateful to the members of the Missouri maize community and thank Vinny for helpful discussions. This work was supported by NSF Midwest Big Data Hub Digital Agriculture Spoke, the Missouri Corn Growers Association, and an anonymous gift.