Learned Visual Navigation for Under-Canopy Agricultural Robots

Arun Narenthiran Sivakumar[1], Sahil Modi[2], Mateus Valverde Gasparino[1], Che Ellis[3], Andres Baquero Velasquez[1], Girish Chowdhary[1*], Saurabh Gupta[4*],

[1]Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign (UIUC)
[2]Department of Computer Science, University of Illinois at Urbana-Champaign (UIUC)
[4]Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign (UIUC)
[3]Earthsense Inc.

Abstract

This paper describes a system for visually guided autonomous navigation of under-canopy farm robots. Low-cost under-canopy robots can drive between crop rows under the plant canopy and accomplish tasks that are infeasible for over-the-canopy drones or larger agricultural equipment. However, autonomously navigating them under the canopy presents a number of challenges: unreliable GPS and LiDAR, high cost of sensing, challenging farm terrain, clutter due to leaves and weeds, and large variability in appearance over the season and across crop types. We address these challenges by building a modular system that leverages machine learning for robust and generalizable perception from monocular RGB images from low-cost cameras, and model predictive control for accurate control in challenging terrain. Our system, CropFollow, is able to autonomously drive 485 meters per intervention on average, outperforming a state-of-the-art LiDAR based system (286 meters per intervention) in extensive field testing spanning over 25 km.

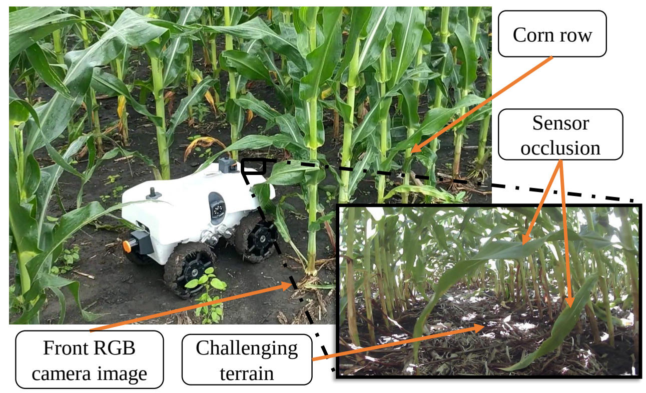

Figure 1: CropFollow is an autonomous navigation system for under-canopy agriculture robots. It uses RGB images from a front-facing camera to output steering commands to drive the robot in crop rows.

Introduction

This paper describes the design of a visually-guided navigation system for compact, low-cost, under-canopy agricultural robots for commodity row-crops (corn, soybean, sugarcane etc), such as that shown in Figure 1. Our system, called CropFollow, uses monocular RGB images from an on-board front-facing camera to steer the robot to autonomously traverse in between crop rows in harsh, visually cluttered, uneven, and variable real-world agricultural fields. Robust and reliable autonomous navigation of such under-canopy robots has the potential to enable a number of practical and scientific applications: High-throughput plant phenotyping (Mueller-Sim et al. 2017; Kayacan, Zhang, and Chowdhary 2018; Young, Kayacan, and Peschel 2019), ultra-precise pesticide treatments, mechanical weeding (McAllister et al. 2020), plant manipulation (Chowdhary et al. 2019; Uppalapati et al. 2020), and cover crop planting. Such applications are not possible with over-canopy larger tractors and UAVs, and are crucial for increasing agricultural sustainability (Shamshiri et al. 2018; Foley et al. 2011).

Autonomous row-following is a foundational capability for robots that need to navigate between crop rows in agricultural fields. Such robots cannot rely on RTK (Real-Time Kinematic)-GPS based methods which are used for over-the-canopy autonomy (Reid et al. 2000; Bak and Jakobsen 2004; Bakker et al. 2011) because of GPS signal attenuation and multi-path errors. LiDAR is known to work in under-canopy and orchard environments and can return geometric information (Barawid Jr et al. 2007; Velasquez et al. 2020; Higuti et al. 2019). However, LiDAR is costly, and it does not capture semantic information leading to low robustness of autonomy, as reported by low distance-between-interventions (Higuti et al. 2019). This motivates our use of richer sensing and lower-cost modalities in the form of RGB images. However, visual variability during the day and across the season as well as the clutter in under-canopy environment limits heuristic based crop-lane detection algorithms (Zhang, Reid, and Noguchi 1999; Ball et al. 2016; Xue, Zhang, and Grift 2012) and therefore necessitates the use of learning. However, the lack of large-scale datasets,

Project website: https://ansivakumar.github.io/learned-visual-navigation/. Full version of this paper was published in the proceedings of Robotics: Science and Systems 2021.

*Girish Chowdhary and Saurabh Gupta contributed equally and are listed alphabetically.

the difficulty of collecting field data, and the infeasibility of building a simulator for this task, makes it challenging to employ machine learning.

Our contribution in this paper is a field-validated modular vision based crop-row following system to overcome the above challenges. We term this system CropFollow, as it provides the foundational row-following capability to small, low-cost robots. Our system decouples perception and control. The perception system uses monocular RGB image from the on-board camera to estimate row-relative robot pose. It does so by directly estimating the robot's relative heading to the row (measured as the angle the robot makes with the row direction), and robot's placement in row (measured as the ratio of distance from the left row to inter-row separation). These data are fused with inertial measurements using a Bayesian sensor fusion system (Extended Kalman filter (EKF)), and utilized to generate row-following control in terms of desired angle and speed for staying in the center of the row using a nonlinear robust controller (Model Predictive Control (MPC)). The ability to directly predict relative heading and distance from monocular RGB images is one key novelty of our approach, and has key efficiency and robustness benefits: the approach avoids having to first detect the plants (which can be many) (Gu et al. 2020), or explicitly segmenting the ground from plants (which is highly challenging with more clutter in the environment) (Xue, Zhang, and Grift 2012). Our presented system is able to successfully traverse crop rows regardless of the crop's growth stage. In field trials of about 25 kilometers, our system required fewer interventions than a LiDAR based system (Velasquez et al. 2021)(485 meters per intervention 286 m), while at the same time cutting down sensing cost by 50×. These results clearly establish that our modular visual navigation system enables vision based autonomy for under-canopy field robots.

System Design

Figure 2 shows an overview of our presented system. Images from on-board RGB camera on the robot are processed through a convolutional network to predict robot heading ϕ, and relative placement d between crop rows. This relative placement is converted into the robot's distance from the left and the right crop rows by multiplying with the lane width. These heading and distance predictions are filtered using a Bayesian filter (we use the Extended Kalman Filter) that optionally also fuses them with high-frequency input from an inertial measurement unit. The filtered heading and distances are used to generate a course correcting reference path in the robot coordinate frame. A model predictive controller is used to compute angular velocity commands to achieve this reference path. A lower-level proportional–integral–derivative (PID) controller is used to track the commanded angular velocity.

In this section the CNN architecture, the Extended Kalman Filter, and the model predictive controller.

Perception Model. We choose a learning approach due to its superior generalizability compared to color-based segmentation navigation proposed by previous works. CropFollow's perception model takes in 320×240 RGB images and outputs the robot heading (in degrees) and its relative placement in the crop row. Heading ϕ is the angle of the robot relative to crop rows. The relative distance d is the ratio of the distance to the left of the row to the lane width, $d = \frac{d_L}{d_L + d_R}$, where d_L and d_R are the distances to left and right crop rows.

The perception model uses a ResNet-18 (He et al. 2016) backbone that has been pretrained on ImageNet (Deng et al. 2009). We truncate ResNet-18 right before the average-pooling layer, and add in an additional convolutional layer, a fully connected layer, dropout, and final prediction layer. The final prediction layer outputs the heading ϕ, and the distance ratio d. We found that independent networks to predict heading and distance ratio worked better than a single joint network.

IMU Fusion with Extended Kalman Filter. An Extended Kalman Filter was used to reduce the effect of uncertainties in distance and heading estimations by fusing the inertial data with the vision data. We used $s = \left(d_L \; d_R \; \phi \right)^T$ as the state. State s_k evolves over time as per the prediction function $f(s_{k-1}, u_{k-1})$ (derived using the robot's kinematics). Here s_{k-1} is the state at the previous time step, and u_{k-1} is the linear and angular velocity at the previous time step. Robot's linear speed v and angular speed ω are calculated from wheel encoders, and IMU respectively. We assume additive zero-mean Gaussian process and measurement noise. As we directly observe s, the measurement function is an identity function. Output from the CNN is used in the update step.

Model Predictive Controller. We used a non-linear Model Predictive Controller (MPC) to generate angular speed commands to the robot given the reference path to be followed, as shown in Figure 2. MPC uses the fused output states $s = \left(d_L \; d_R \; \phi \right)^T$ from the EKF, the Unicycle kinematic model of the robot and reference path, which is a straight line through the center of the lane, to solve a constrained optimization problem with the minimum and maximum curvature radius as the constraints. The output is a path defined in terms of the curvature ρ, which determines the angular velocity $\omega = \rho \; v$ where v is the linear velocity. The angular speed for the first point in the output path is applied and the optimization process is repeated. A PID controller is used to maintain the commanded angular speed, based on feedback from IMU's yaw angular speed.

Data Collection and Ground Truthing

Given lack of any under-canopy agriculture datasets, we collected a large dataset by driving the TerraSentia robot under the canopy. We manually operated the robot in 19 corn and 4 soybean fields across Illinois and Indiana, and collected time-series data from the front-facing RGB camera, LiDAR, and IMU. We collected 2.7 hours of corn data and 1.2 hours of soybean data, and made sure to collect data for different growth stages. We also included data where the robot was driven in a zigzag manner. This was done to expose the perception models to a broader distribution of data that may be experienced during autonomous runs.

Ground Truthing. To get the ground truth heading and distance ratio for training the networks, we designed an indirect annotation procedure. We asked humans to label the

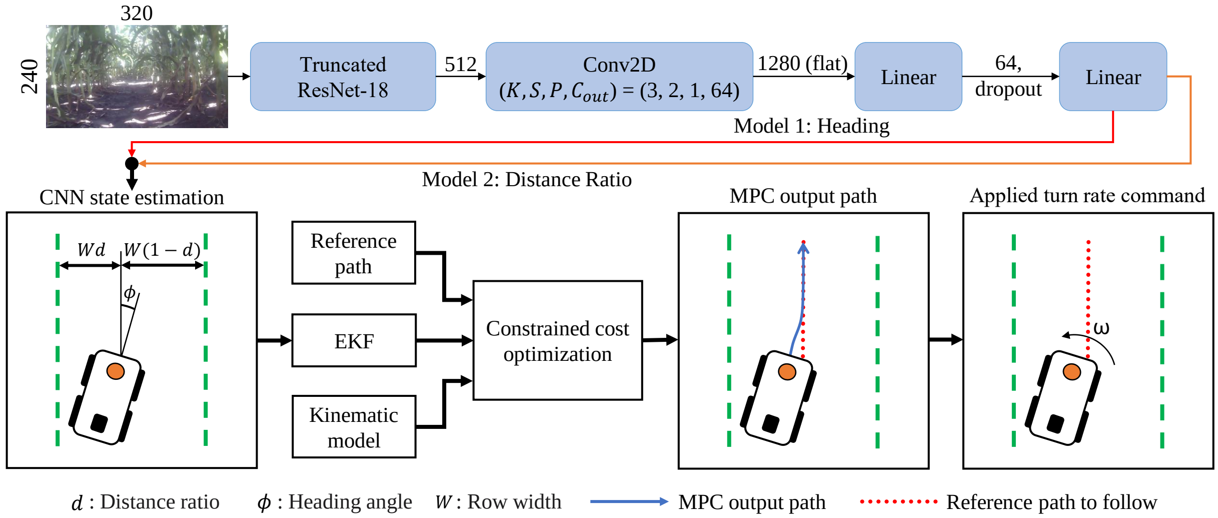

Figure 2: **CropFollow Overview**. We use a convolutional network to output robot heading and placement in row. This is used to compute the row center which is used as a reference trajectory. A model predictive controller converts reference trajectories to angular velocity commands.

Figure 3: **Annotations.** We annotate the horizon and crop rows for early season images (left). For late season images when the horizon is not visible, we annotate the vertical corn stalks (right).

Model	**Mean**		**Median**		**95%ile**	
	ϕ_{err}	d_{err}	ϕ_{err}	d_{err}	ϕ_{err}	d_{err}
Baseline	11.41	0.48	8.81	0.48	30.33	0.65
CropFollow	**1.99**	**0.04**	**1.21**	**0.03**	**4.71**	**0.10**

Table 1: **Perception Module Performance:** We report L1 error in heading (in °) and distance ratio prediction. The trivial baseline model always predicts median ϕ, d from the training set. CropFollow after training is significantly better than trivial baseline.

horizon and the vanishing lines corresponding to the crop row (Figure 3 (left)). This together with the camera calibration information allows us to recover the robot heading and placement in row using projective geometry. For the case where the horizon is not visible, we instead ask humans to mark out vertical crop stalks (Figure 3 (right)). This allows us to estimate the vanishing point for the vertical direction which readily provides the slope of the horizon. We annotated a total of $25,296$ corn images. 28% of these are from early growth stage, while 72% are from late growth stage. We split the dataset into a training and a validation set (83% training, 17% validation). We made sure that data from the same video is either entirely in the training set, or entirely in the validation set. Our main experiments use this corn data.

Experimental Results

Our experiments are designed to test the autonomous crop row traversal capability of our proposed system and effectiveness of the proposed modular policy. We evaluate these aspects through a combination of offline and online (field) experiments. Offline experiments are conducted on our collected dataset. They allow us to chose models for online experiments. Online experiments are conducted in the field, and allow us to study the interplay between perception and control systems. We also conduct end-to-end evaluation for the task of crop row traversal, and compare against an existing system based on LiDAR (Velasquez et al. 2021).

Offline Evaluation of Perception Model

Offline evaluation of the perception module is conducted on the collected dataset.

Metrics. We measure prediction performance using L1 error in heading and distance ratio predictions, ϕ and d.

Training. We used ResNet-18 (He et al. 2016) pretrained on ImageNet (Deng et al. 2009) to initialize our models. Models were trained to minimize the $L2$ loss with the Adam optimizer (Kingma and Ba 2014) for 50 epochs. We started with an initial learning rate of 10^{-4} and dropped it by a factor of 10 at 40^{th} and 45^{th} epochs. All layers of the network were optimized.

Results. Table 1 presents the performance of our CNN models.Our best model achieves an average L1 error of $1.99°$ for heading, and 0.04 for distance ratio. Inference speed for this model on the robot was around 20 FPS, which is fast enough for accurate control. Our main field experiments are conducted with this model.

Figure 4: Sample images from field trials. Bottom row consists of traditionally adverse conditions for vision-based navigation.

In Field End-to-End System Evaluation

We conducted end-to-end system evaluation with the model described above. We compared the performance of the following 2 systems, along with 2 variants each:

- **CropFollow (w/ IMU)**. This is our proposed system that uses the above CNN model for heading and distance ratio prediction, EKF for fusing IMU information, and MPC for executing control commands. We also compare with a variant that does not use IMU information (denoted by CropFollow (w/o IMU)).
- **LiDAR System (Velasquez et al. 2021) (w/ IMU)**. This system uses readings from the LiDAR mounted on top of the robot to estimate the robot heading and distance from the crop rows using line fitting. Other parts of the system are same as our system: Use of an EKF to fuse information from the IMU, and use of MPC for generating control commands. We also compare to a variant that does not use IMU information (denoted by LiDAR System (Velasquez et al. 2021) (w/o IMU)).

Evaluation Methodology. All 4 systems are tested on the same unique 4.85 km. These 4.85 km come from 15 different experiments that were done in different parts of the field, over different growth stages, different days, different time of the day, and weather conditions. While there is a lot of variability in these 4.85 km, we attempted to minimized the variability in conditions for the 4 systems to ensure result comparability. Runs for the different systems for each of the 15 experiments were done one after another over the same routes, and with the same constant linear robot velocity of 0.6 m/s. Run order for the different systems was randomized to prevent environmental bias. This experiment thus presents results pooled over field trials of 19.4 km. For each method, we measure the number of human interventions needed to complete the experiment. Human interventions were required when the robot crashed into the corn stalks. This metric measures autonomy effectiveness.

Results. Table 2 reports the number of interventions for the 4 systems that we evaluated. We separately report results for early and late season experiments. Note that LiDAR system from (Velasquez et al. 2021) can't operate in early season data since early season corn stalks are shorter than the robot, and not detected by the 2-D LiDAR. Our vision based systems works reasonably well. In late season when the LiDAR based system does work, we note that it had more interventions than our system, 72 vs 8 without IMU, and 13 vs 7 with IMU. Thus, our presented vision-based system outperforms the LiDAR based system, while also reducing sensing cost by $50\times$. The quality of our output is further shown by the fact that our system is closing the loop only at about 20Hz, 40Hz for the LiDAR system, but still achieves a better end performance.

Growth Stage	Length (in m)	LiDAR w/ IMU	LiDAR w/o IMU	CropFollow w/ IMU	CropFollow w/o IMU
Early	1120	-	-	3	4
Late	3726	13	72	7	8

Table 2: **Field Experiments:** We report the number of interventions for the different methods. LiDAR can't operate in early season as crops are too short. Our system can work under both conditions and requires interventions.

Different error modes. We characterized the different error modes in CropFollow and LiDAR navigation system. Large gaps in crop rows was the common cause of failure in CropFollow (our training data did not include such cases). Sensor occlusion and bumpy terrain were the other rare causes of failures. In contrast, failure due to gaps was rarely observed in LiDAR since it was specifically engineered to be robust to it. But because of its high sensitivity to noise, even minor sensor occlusion by leaves affects LiDAR performance and leads to interventions. CropFollow's performance in gaps could be improved with adding training data whereas LiDAR's occlusion problem is a sensor limitation.

Stress testing. To test the performance in challenging conditions, CropFollow (w/ and w/o IMU) was tested in a field with sharp curves, gaps and occlusion from weeds. 3 and 6 interventions w/ and w/o IMU respectively was observed in a test of 600m. Last row in Figure 4 shows the challenging condition in this field. Also, CropFollow's performance at higher speeds was tested. CropFollow showed same stable behavior at $1m/s$ but oscillations in trajectory due to latency was observed at $1.4m/s$ or more.

Conclusion

We presented a vision based autonomous under-canopy navigation system. Through a modular architecture and a learning-based approach we showed that machine vision can be applied for reliable and robust navigation in cluttered, changing, and harsh under-canopy environments. 25 km of real-world validation on an under-canopy robot demonstrated that our visual navigation approach is not only $50\times$ more cost-effective than LiDAR but also leads to fewer interventions. Our system forms a new benchmark for visual navigation under the canopy, and our openly accessible dataset (1030 labeled images and 24266 unlabeled images of our corn data) will enable further research.

References

Bak, T.; and Jakobsen, H. 2004. Agricultural robotic platform with four wheel steering for weed detection. *Biosystems Engineering*, 87(2): 125–136.

Bakker, T.; van Asselt, K.; Bontsema, J.; Müller, J.; and van Straten, G. 2011. Autonomous navigation using a robot platform in a sugar beet field. *Biosystems Engineering*, 109(4): 357–368.

Ball, D.; Upcroft, B.; Wyeth, G.; Corke, P.; English, A.; Ross, P.; Patten, T.; Fitch, R.; Sukkarieh, S.; and Bate, A. 2016. Vision-based obstacle detection and navigation for an agricultural robot. *Journal of field robotics*, 33(8): 1107–1130.

Barawid Jr, O. C.; Mizushima, A.; Ishii, K.; and Noguchi, N. 2007. Development of an autonomous navigation system using a two-dimensional laser scanner in an orchard application. *Biosystems Engineering*, 96(2): 139–149.

Chowdhary, G.; Gazzola, M.; Krishnan, G.; Soman, C.; and Lovell, S. 2019. Soft Robotics as an Enabling Technology for Agroforestry Practice and Research. *Sustainability*, 11(23): 6751.

Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; and Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. IEEE.

Foley, J. A.; Ramankutty, N.; Brauman, K. A.; Cassidy, E. S.; Gerber, J. S.; Johnston, M.; Mueller, N. D.; OaConnell, C.; Ray, D. K.; West, P. C.; Balzer, C.; Bennett, E. M.; Carpenter, S. R.; Hill, J.; Monfreda, C.; Polasky, S.; Rockstram, J.; Sheehan, J.; Siebert, S.; Tilman, D.; and Zaks, D. P. M. 2011. Solutions for a cultivated planet. *Nature*, 478(7369): 337–342.

Gu, Y.; Li, Z.; Zhang, Z.; Li, J.; and Chen, L. 2020. Path Tracking Control of Field Information-Collecting Robot Based on Improved Convolutional Neural Network Algorithm. *Sensors*, 20(3): 797.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.

Higuti, V. A.; Velasquez, A. E.; Magalhaes, D. V.; Becker, M.; and Chowdhary, G. 2019. Under canopy light detection and ranging-based autonomous navigation. *Journal of Field Robotics*, 36(3): 547–567.

Kayacan, E.; Zhang, Z.; and Chowdhary, G. 2018. Embedded High Precision Control and Corn Stand Counting Algorithms for an Ultra-Compact 3D Printed Field Robot. *Proceedings of Robotics: Science and Systems. Pittsburgh, Pennsylvania*.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

McAllister, W.; Whitman, J.; Axelrod, A.; Varghese, J.; Chowdhary, G.; and Davis, A. 2020. Agbots 2.0: Weeding Denser Fields with Fewer Robots. In *Proceedings of Robotics: Science and Systems*. Corvalis, Oregon, USA.

Mueller-Sim, T.; Jenkins, M.; Abel, J.; and Kantor, G. 2017. The Robotanist: a ground-based agricultural robot for high-throughput crop phenotyping. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 3634–3639. IEEE.

Reid, J. F.; Zhang, Q.; Noguchi, N.; and Dickson, M. 2000. Agricultural automatic guidance research in North America. *Computers and electronics in agriculture*, 25(1-2): 155–167.

Shamshiri, R. R.; Weltzien, C.; Hameed, I. A.; Yule, I. J.; Grift, T. E.; Balasundram, S. K.; Pitonakova, L.; Ahmad, D.; and Chowdhary, G. 2018. Research and development in agricultural robotics: A perspective of digital farming. *International Journal of Agricultural and Biological Engineering*, 11(4): 1–14.

Uppalapati, N. K.; Walt, B.; Havens, A.; Mahdian, A.; Chowdhary, G.; and Krishnan, G. 2020. A Berry Picking Robot With A Hybrid Soft-Rigid Arm: Design and Task Space Control. In *Proceedings of Robotics: Science and Systems*. Corvalis, Oregon, USA.

Velasquez, A.; Higuti, V.; Guerrero, H.; Gasparino, M.; Magalhães, D.; Aroca, R.; and Becker, M. 2020. Reactive navigation system based on H control system and LiDAR readings on corn crops. *Precision Agriculture*, 21(2): 349–368.

Velasquez, A. E. B.; Higuti, V. A. H.; Gasparino, M. V.; Sivakumar, A. N.; Becker, M.; and Chowdhary, G. 2021. Multi-Sensor Fusion based Robust Row Following for Compact Agricultural Robots. In *arXiv*.

Xue, J.; Zhang, L.; and Grift, T. E. 2012. Variable field-of-view machine vision based row guidance of an agricultural robot. *Computers and Electronics in Agriculture*, 84: 85–91.

Young, S. N.; Kayacan, E.; and Peschel, J. M. 2019. Design and field evaluation of a ground robot for high-throughput phenotyping of energy sorghum. *Precision Agriculture*, 20(4): 697–722.

Zhang, Q.; Reid, J. F.; and Noguchi, N. 1999. Agricultural vehicle navigation using multiple guidance sensors. In *Proceedings of the international conference on field and service robotics*, 293–298. August.

Acknowledgments

This project was supported in part by NSF STTR #1820332, USDA/NSF CPS project #2018-67007-28379, USDA/NSF AIFARMS National AI Institute USDA #020-67021-32799/project accession no. 1024178, NSF IIS #2007035, and DARPA Machine Common Sense. We thank Earthsense Inc. for the robots used in this work and we thank the Department of Agricultural and Biological Engineering and Center for Digital Agriculture (CDA) at UIUC for the Illinois Autonomous Farm (IAF) facility used for data collection and field validation of CropFollow. We thank Vitor Akihiro H. Higuti and Sri Theja Vuppala for their help in integration of CropFollow on the robot and field validation.