Challenges and Opportunities for Autonomous Micro-UAVs in Precision Agriculture

Xu Liu and Steven W. Chen, *University of Pennsylvania, Philadelphia, PA, 19104, USA*

Guilherme V. Nardari, *University of São Paulo, São Paulo, 13566590, Brazil*

Chao Qu, Fernando Cladera Ojeda, Camillo J. Taylor, and Vijay Kumar, *University of Pennsylvania, Philadelphia, PA, 19104, USA*

Mobile robots such as unmanned ground vehicles (UGVs) and unmanned aerial vehicles (UAVs) are increasingly used for precision agriculture. While UGVs have larger payload capabilities and longer operation time, they are limited to 2-D space. This makes UAVs better suited for tasks that require fast coverage, harsh terrain traversal, and high altitude or multilevel operation. However, it remains a challenging task to develop a reliable yet fully autonomous UAV system that can actively extract actionable information in large-scale cluttered agricultural environments. Such a system will have to estimate its own poses, build a map of the environment, navigate through obstacles, and act to gather information with limited onboard computation and battery life. In this survey, we first review recent advances in UAV hardware and software, ranging from novel platforms and sensors to state-of-the-art autonomous navigation, object detection and segmentation, robot localization, and mapping algorithms related to agriculture. We then provide a list of challenges in each field and potential opportunities for the broader adoption of UAVs in precision agriculture.

The world is facing a unique combination of very challenging problems[1]: There are over 700 million malnourished people; most land that can be cultivated is already in use; agriculture accounts for over 70% of freshwater usage; pathogens can increase the inefficiency and losses in farming; modern agriculture needs to adapt to challenges related to climate change. While humans excel at making scientific assessments (e.g., measuring sizes, identifying species, infestation, or investigating damage to the roots/soil) at a relatively small scale, accurately performing such assessments at a large scale is still challenging. For example, data gathering forms a large part of the expenditure on forest management but is carried out in a rudimentary fashion with human labor. Development and cross-fertilization of technologies in robotics, artificial intelligence, and agriculture will have a tremendous impact.

0272-1732 © 2022 IEEE
Digital Object Identifier 10.1109/MM.2021.3134744
Date of current version 1 February 2022.

Many commercial solutions for precision agriculture exist today, including satellite-based, UGV-based, and UAV-based technologies. Although unmanned ground vehicles (UGVs) have advantages in operation time and payload capabilities, they also have intrinsic limitations. First, ground robots can only see objects of interest close to the ground or from a viewpoint close to the ground, reducing the amount of information gathered. Second, ground robots cannot traverse steep and harsh terrains, commonly found in agricultural environments (rice farms, terrace farms, and forests). Finally, UGVs are unable to survey large agriculture fields rapidly.

Recently, we have seen remarkable advances in unmanned aerial vehicles (UAVs) that weigh less than 5 kg. UAVs can hover and fly fast in 3-D environments. They are becoming popular in precision agriculture for yield estimation, crop fertilization, and crop monitoring. Most applications, however, focus on overhead flight

FIGURE 1. Example forest and orchard environments. The left-hand panel shows the over-canopy view of a forest, the middle panel shows the under-canopy view of the same forest, and the right-hand panel shows the view from a UAV flying in an apple orchard. Navigating under tree canopy is challenging for UAVs due to moving and repetitive features, tiny branches, undergrowth, and unreliable GPS. Accurate semantic mapping is also hard because the environment is unstructured and the operating conditions (illumination, occlusions) are quite varied.

through wide-open space.[2] This drastically simplifies operations, but over-canopy data severely limit what is possible to measure. For example, it is difficult to assess individual fruits' sizes and health conditions, or get measurements on the diameters of trees from overhead data alone.

IN THIS ARTICLE, WE WILL DESCRIBE THE STATE OF THE ART IN UAVs FOR PRECISION AGRICULTURE, AND THE CHALLENGES REQUIRED TO BUILD AUTONOMOUS UAV SYSTEMS THAT CAN ACTIVELY EXTRACT ACTIONABLE INFORMATION IN CLUTTERED AND HIGHLY UNSTRUCTURED AGRICULTURAL ENVIRONMENTS ACROSS MULTIPLE SQUARE KILOMETERS, ADDRESSING CHALLENGES ON BOTH HARDWARE AND SOFTWARE.

UAVs capable of under-canopy flights can address these issues. Under-canopy UAVs can achieve a good tradeoff between coverage rate and sensor resolution while keeping the labor cost modest. However, developing an autonomous UAV system that can operate at large scale and multiple altitudes, between rows of trees, or even go under the tree canopy is still very challenging: First, global positioning system (GPS) is not always reliable due to canopy occlusion. Second, the environment is very unstructured and dynamic (e.g., leaves or grass blowing in the wind), which poses significant challenges on robot odometry systems that rely on static geometric features. Third, the environment is cluttered with many tiny objects, which requires a very accurate and dense mapping system.

There is prior work that addresses under-canopy data collection but mostly with human pilots.[3,4] Only a few under-canopy autonomous flights have been demonstrated so far,[5,6] but the environments are much sparser than those shown in Figure 1. Moreover, those flights are only performed at a relatively small scale.

In this article, we will describe the state of the art in UAVs for precision agriculture, and the challenges required to build autonomous UAV systems that can actively extract actionable information in cluttered and highly unstructured agricultural environments across multiple square kilometers, addressing challenges on both hardware and software. We will mainly address micro-UAVs, with a focus on operation between trees and under the canopy flight, which is more challenging and the opportunity for impact is more significant.

UAV HARDWARE AND AUTONOMY

There has been a considerable progress in the past decade with UAVs that can operate with sophisticated sensory and computational payloads, despite the stringent power constraints.[7,8] Today, UAV technologies for agricultural applications are available as commercial products. However, these solutions focus on relatively less complex missions such as overhead flights with GPS. Even though there are commercial UAV systems that can perform fully autonomous flights in cluttered environments, most of them still rely heavily on GPS if long-range operation is required. Moreover, they are not designed to accomplish high-level missions in complex agricultural environments, such as building large-scale high-resolution semantic maps of fruits or trees.

Platform	Year	Computation power	Storage	Weight	Battery	Size (tip to tip)	Flight time
	2013-2015	Intel NUC (i5)	512 GB	4.5 kg	10000 mAh (6S)	1.18 m	~15 min
	2013-2015	Intel NUC (i3)	512 GB	1.9 kg	4000 mAh (4S)	0.80 m	~8 min
	2016-2018	Intel NUC (i7)	1 TB	3.5 kg	6000 mAh (6S)	0.76 m	~10 min
	2018-2019	NVDIA Jetson TX2 + FPGA	32 GB	1.0 kg	4000 mAh (4S)	0.40 m	~15 min
	2018-2019	Qualcomm SnapDragon 801	32 GB	250 g	910 mAh (2S)	0.32 m	~8 min
	2019-2021	Intel NUC (i7) + FPGA	1 TB	4.2 kg	17000 mAh (6S)	1.10 m	~30 min
	2019-2021	Qualcomm SnapDragon 821	32 GB	185 g	650 mAh (2S)	0.24 m	~10 min

FIGURE 2. Generations of autonomous UAV platforms developed by our lab. The need for autonomy in GPS-denied environments leads to bigger sensor and CPU payloads, and the need for longer flight times leads to bigger batteries. Despite the significant decrease in price/performance and weight/performance ratios for LiDARs and CPUs/GPUs, sub-500-g platforms have to be based on camera-IMU sensing packages and smartphone processors—see, for example, the smallest platform (0.12-m radius, 185 g). Faster flight requires long-range sensing and therefore platforms based on LiDARs and move powerful CPUs, which are bigger and heavier (over 3.5 kg).

UAV Platforms and Autonomy

An autonomy stack for UAVs usually consists of state estimation, planning, and control modules, as shown in Figure 3. The goal of the mission and the properties of the environment such as scale, structuredness, clutteredness, and access to GPS have a significant influence on the design of each module, which directly impacts the onboard computing and sensor platforms. Figure 2 shows various platforms that have been used to demonstrate safe, high-speed flight (2–10 m/s) in moderately cluttered environments without human oversight, GPS, or radio communications for up to 1 km.

In Mohta *et al.*'s work,[7] a stereo visual-inertial odometry (VIO) algorithm was used for state estimation, a 2-D light detection and ranging (LiDAR) mounted on a nodding gimbal was used for mapping and obstacle avoidance, and a search-based motion planner was used for motion primitives plans collision-free and dynamically feasible trajectories. A vision-based autonomous flight system was proposed by

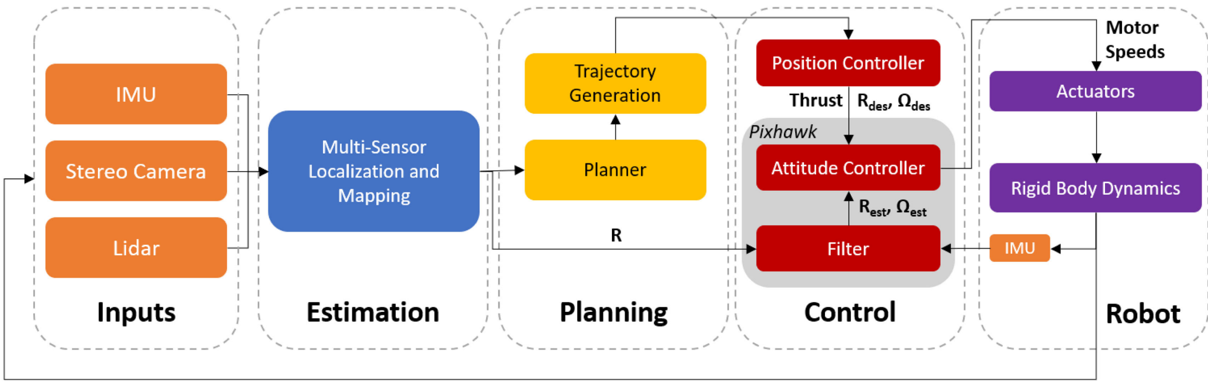

FIGURE 3. Autonomous UAV system consists of state estimation (simultaneously estimating robot poses and building a map from raw sensor data), planning (computing a collision-free and dynamically feasible trajectory from current pose to a given goal), and control (tracking the trajectory considering the UAV model) modules.

Oleynikova *et al.*[8] for dense mapping in GPS-denied scenarios. The system can navigate in structured and moderately unstructured environments, such as collapsed buildings. While this system was tested in simulated forest environments, no real-world experiment for such highly unstructured environments was shown. A conservative local planning strategy was used with unobserved space regarded as nonfree space, which potentially limits the flight speeds as well as the optimality of local planning. A system of multiple UAVs collaboratively surveying and mapping under forest canopy was presented by Tian *et al.*[6] A 2-D LiDAR is used for state estimation and mapping. A collaborative simultaneous localization and mapping (SLAM) scheme with loop closure capabilities was proposed where object-based maps were built by detecting trees and using them as landmarks. This representation significantly decreases communication bandwidth between robots and the base station. However, the proposed clustering- and filtering-based tree detection method may not be able to reliably detect trees for large-scale complex forests with thick undergrowth or tiny branches. In addition, the UAV is operating in a 3-D environment, much of the space will not be observed by the 2-D body-fixed LiDAR.

A common limitation for these systems is that they are unable to perform long-range (several kilometers) missions in complex agricultural environments, as illustrated in Figure 1. In addition, they also do not have the capability of detecting and modeling objects of interest such as fruits or trees in real time and at scale, which is important for guiding the robot to plan more informative trajectories.

Sensor Configuration

Due to the limited payload capability, sensors in UAVs serve a dual purpose: they are used both for autonomy, as well as for gathering mission-specific data. Most commonly used sensors in UAVs for precision agriculture include: cameras, IMUs, LiDARs, and global navigation satellite systems (GNSS). Because multirotor UAVs consume around 100–200 W/kg,[9] lightweight sensors naturally enable longer missions.

Cameras and IMUs provide a lightweight low-power sensor combination for navigation and obstacle avoidance.[2] For navigation purposes, cameras and IMUs can provide fast odometry updates in GPS-denied conditions. However, cameras suffer from intrinsic drawbacks, such as scale ambiguity, the requirement for calibration, limited dynamic range, and high computation requirements. For example, the limited baseline of stereo cameras has a direct impact on the depth error. Direct sunlight or darkness affects the performance of cameras. For instance, sunlight coming through canopies can create patches of bright spots on the ground, making it difficult to adjust exposure. Some of these problems may be solved with the use of event-based cameras,[10] but they are heavier and more expensive than conventional cameras.

Three-dimensional LiDARs provide rich information for autonomous robotic platforms, with the ability to perceive objects at tens to hundreds of meters. This makes them an essential sensor for obstacle avoidance flights at high speed in very cluttered environments or when accurate measurements are required. While still expensive, 3-D LiDAR prices have dropped significantly in the recent years, and LiDAR-based algorithms are a fast-growing field of research. Unfortunately, the sensor weight is still considerable. Solid-state LiDAR which replaces the moving parts of traditional LiDARs with semiconductors, may offer a significant improvement over the current technologies in term of robustness, weight, and power consumption.

Both cameras and LiDARs have their unique advantages, and combining them provides a good balance between the different capabilities of the sensors. For such an approach, there will be challenges regarding meeting size and weight constraints as well as sensor synchronization and calibration.

Finally, GNSS sensors such as GPS allow robots to obtain geospatial positioning. These sensors are inexpensive, and their accuracy can be as high as 1 m. However, if there are obstacles between the receiver and the satellite, the accuracy of GNSS decreases, which is the case for between-tree or under-canopy environments. Enhancements to GNSS using ground stations, such as differential GPS (DGPS) and real-time kinematics (RTK), allow improvements in accuracy up to the centimeter range. When using corrections, it is necessary to have a radio link between the UAV and the base station. Postprocessing techniques, such as postprocessed kinematic (PPK), can help get accurate measurements when there is no available link between the base station and the robot, but it cannot be used for real-time control. For these reasons, a reliable onboard state estimation system without relying on GPS is key to UAVs in precision agriculture.

Challenges

Obstacles like thin branches are notoriously hard to detect, yet failure to avoid them will result in catastrophic crashes. An obvious solution is to increase the resolution of the onboard sensors. Kong *et al.*[11]

proposed a fully autonomous quadrotor system with a forward facing solid-state LiDAR that can fly safely in cluttered environments while avoiding small obstacles. By focusing the LiDAR beams to the front, the sensor effectively gains resolution, but at the cost of the full 360-degree spatial awareness provided by a spinning LiDAR with lower resolution. As UAVs usually have limited payload and onboard computation, this problem cannot be solved by adding more sensors without sacrificing flight time. The tradeoff between sensing capability and power consumption should be carefully investigated when designing such systems.

Smaller UAVs are safer, able to fly through narrower gaps, more agile, and easier to deploy. Although small and lightweight UAVs are available in the market, their autonomy stack capabilities, computation power, and data storage for performing large-scale autonomous missions in GPS-denied conditions is still limited.

Running onboard deep learning detectors is especially challenging because of the computational needs. While some of the data may be postprocessed, inference at the edge may be required when deep learning is used in the robot control loop. Network discretization[12] and AI accelerators such as application-specific integrated circuits (ASICs)[13] and embedded general-purpose graphics processing units (GPGPUs) may be the key toward intensive machine learning onboard UAVs.

A friendly and intuitive user interface is required to democratize the access to complex agricultural UAV systems, and abstracting this complexity to the user is challenging. When trying to operate in a new environment, UAVs usually require tuning and configuration from knowledgeable users.

OBJECT DETECTION AND SEGMENTATION

Object detection and semantic segmentation are crucial to precision agriculture because actionable information is usually related to semantic features such as fruits or trees, or phenotype data. In this section, we will review methods to obtain semantic features based on both 2-D images and 3-D LiDAR data.

Image-Based (2-D)

Various imaging techniques including RGB, multispectral and hypersectral, thermal, and near-infrared images have been used for agricultural applications such as image classification, anomaly detection, and yield estimation.

Early attempts use classical machine learning methods, such as K-means clustering and support vector machines, to solve agriculture-related object detection and segmentation problems, using hand-crafted features. Later efforts shift toward a data-driven paradigm,[4,14] accelerated by the massive progress in deep learning. These techniques have also been used in agriculture-related tasks such as weed detection, plant stress assessment, leaf area index evaluation, soil segmentation, or moisture distribution modeling. We refer the readers to a recent survey on this topic.[15]

LIDAR-Based (3-D)

The most commonly used 3-D representation for object detection include voxel grid, point cloud, multiview, and spherical image.

Volumetric convolutional neural networks (CNNs) represent data as 3-D occupancy grids, and directly apply 3-D convolution on this representation. Multiview-based approaches project 3-D data from different viewing angles to a 2-D image, which can then be fed to existing image-based CNNs. However, the projection process is sensitive to noisy or incomplete input, and the voxelization process induces information loss. For this reason, some argue that it is better to directly use point clouds. Other approaches also combine 2-D multiview and 3-D point cloud representations. Another solution to this problem is projecting range data of each LiDAR scan into a spherical image (i.e., range image), which can also leverage existing image-based CNNs while avoiding information loss. For further references, we refer our readers to a recent survey paper by Guo *et al.*[16]

The application of these techniques in agriculture is a relatively new area of study. Some recent work use methods such as clustering, filtering, circle fitting, and arc extraction to detect trees.[6] However, their performances are parameter-sensitive, which require expert knowledge to tune. Therefore, a data-driven approach is preferred for large-scale complicated agriculture environments. For instance, Chen *et al.*[3] used a data-driven approach to detect trees from point clouds using LiDAR range images.

Challenges

Access to large high-quality labeled data sets remains a significant challenge. While there exist many labeled data sets in the public domain, very few are agriculture specific.[17,18] Transfer learning can be used to adapt models trained on general data sets to agriculture settings.[4] Another approach is to leverage synthetic data sets. However, the so-called sim-to-real gap is still quite significant. Finally, for tasks such as yield estimation, acquiring accurate ground-truth harvest data still remains labor intensive and time-dependent.

Occlusion poses a big challenge to yield estimation and semantic mapping in orchards. Simple statistical tools such as linear regression can be used to account for the difference between estimated and actual yields. However, these calibration procedures are limited in that they require the presence of previous harvest yield data. Possible solutions to deal with heavy occlusions include aggregating information from multiple views and multiple sensors, or utilizing robotic manipulators to effectively expose the occluded objects.

ROBOT LOCALIZATION AND MAPPING

Traditional Localization and Mapping

Robot localization and mapping is a key module in autonomous navigation system. Mapping is the task of accumulating readings from one or more sensors over time into a representation of the space observed by a robotic agent. The agent may also be required to estimate its position with respect to this map.

Traditional approaches address this problem relying on geometric features (corners, lines, and planes). Plane and line features are hard to find. Corner features are repetitive, hard to match, and sometimes dynamic (e.g., corners on the grass or undergrowth move with the wind), which leads to large drifts or even failure in robot localization. A potential solution to these problems is exploiting semantics or prior knowledge about the structure of the field.

Semantic Localization and Mapping

The importance of semantic features is twofold: first, they allow us to generate a meaningful map represented by models of objects of interests; second, they enable more reliable robot pose estimation. Semantic objects are sparser and more unambiguous and thus less susceptible to mismatching and semantic information can indicate whether the object is static.

For agriculture-related localization and mapping, some prior work also utilizes semantic information. For example, although trees look largely similar to each other, they are much sparser compared to corner features. Thus, it is easier to track them across a data sequence (e.g., image frames, LiDAR scans). In addition, the spatial relationship of semantic features, such as polygons formed by multiple trees, can be used as descriptors for local regions.[19] Such descriptors can be utilized to help robot correct drift, detect loop closures, and build a better map over a long range.

Semantic LiDAR odometry and mapping (SLOAM) is proposed by Chen *et al.*,[3] which generates high-resolution forest maps and extracts timber volume estimates. SLOAM detects trees and ground using a neural network and explicitly models them. These models are used for mapping and state estimation.

A monocular-camera-only CNN-based system that counts and maps fruits from image sequences was proposed by Liu *et al.*,[20] where the reconstructed 3-D fruit locations are used to reject outliers that cannot be identified in 2-D images. Semantic information including tree trunks and ground were used by Dong *et al.*[21] to align and merge views from two opposite sides of the tree rows. Chebrolu *et al.*[22] augmented classic features with semantic data, using the segmentation of crops, weeds, and their stem positions to improve data association between an orthomosaic map and images captured from a ground robot.

Challenges

Agricultural environments pose additional challenges to the robot's localization and mapping capability compared to urban ones, especially at a larger scale. Typical urban environments have plenty of man-made objects with well-defined geometric, visual, and semantic features. Such features enables state-of-the-art SLAM algorithms to produce high-quality maps, which in turn allows accurate robot localization. However, these features are rare in an agricultural setting. For example, in a dense forest or orchard, many trees have similar textures and shapes, which causes perceptual aliasing for both vision and LiDAR sensors. If not carefully dealt with, they will give rise to incorrect data association and result in estimator drift, or even complete failure.

> *WE IMAGINE A FUTURE WHERE AUTONOMOUS UAVs WILL HAVE A TREMENDOUS IMPACT IN PRECISION AGRICULTURE, ENABLING FAST INFORMATION ACQUISITION IN LARGE-SCALE COMPLEX ENVIRONMENTS.*

Semantic mapping consists of adding semantic concepts to geometric map. Building a semantic map is an important step toward a more efficient, task-based representation for human interaction. In addition, it leads to a hierarchical map representation in which the top level only incorporates sparse semantic information for action guidance, and a smaller geometric map can be built around the robot for local motion planning and obstacle avoidance. This facilitates real-time computation for autonomous operation on resource-constraint robots.

Active mapping is the process of robot actively choosing actions to reduce uncertainties in the map based on its sensor measurement model. However, it is hard to directly apply such methods to agricultural environments where the scale is large and information is densely distributed. One potential solution is to leverage the semantic map since it is usually sparser than the metric map. This in turn requires measurement models that can account for uncertainties in semantic information in a meaningful and efficient manner.

CONCLUSION

We imagine a future where autonomous UAVs will have a tremendous impact in precision agriculture, enabling fast information acquisition in large-scale complex environments. In this survey, we have summarized the recent advances regarding UAV hardware and software, and highlighted challenges and opportunities toward ubiquitous adoption of UAVs in precision agriculture. It is our hope that this work provides a roadmap for both academic and industrial efforts, such as NSF Engineering Research Center for the Internet of Things for Precision Agriculture (IoT4Ag).[1]

ACKNOWLEDGMENTS

This work was supported in part by the IoT4Ag Engineering Research Center funded by the National Science Foundation (NSF) under NSF Cooperative Agreement EEC-1941529 and in part by C-BRIC, a Semiconductor Research Corporation Joint University Microelectronics Program cosponsored by DARPA.

REFERENCES

1. "The Internet of Things for Precision Agriculture (IoT4Ag), an NSF Engineering Research Center." Accessed: Aug. 20, 2021. [Online]. Available: https://iot4ag.us
2. P. Radoglou-Grammatikis, P. Sarigiannidis, T. Lagkas, and I. Moscholios, "A compilation of UAV applications for precision agriculture," *Comput. Netw.*, vol. 172, 2020, Art. no. 107148.
3. S. W. Chen *et al.*, "SLOAM: Semantic lidar odometry and mapping for forest inventory," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 612–619, Apr. 2020.
4. S. W. Chen *et al.*, "Counting apples and oranges with deep learning: A data-driven approach," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 781–788, Apr. 2017.
5. K. Mohta *et al.*, "Fast, autonomous flight in GPS-denied and cluttered environments," *J. Field Robot.*, vol. 35, no. 1, pp. 101–120, 2018.
6. Y. Tian *et al.*, "Search and rescue under the forest canopy using multiple UAVs," *Int. J. Robot. Res.*, vol. 39, no. 10/11, pp. 1201–1221, 2020.
7. K. Mohta *et al.*, "Experiments in fast, autonomous, GPS-denied quadrotor flight," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 7832–7839.
8. H. Oleynikova *et al.*, "An open-source system for vision-based micro-aerial vehicle mapping, planning, and flight in cluttered environments," *J. Field Robot.*, vol. 37, no. 4, pp. 642–666, 2020.
9. Y. Mulgaonkar, M. Whitzer, B. Morgan, C. M. Kroninger, A. M. Harrington, and V. Kumar, "Power and weight considerations in small, agile quadrotors," *Proc. SPIE*, vol. 9083, 2014, Art. no. 90831Q.
10. G. Gallego *et al.*, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, Jan. 2022.
11. F. Kong, W. Xu, Y. Cai, and F. Zhang, "Avoiding dynamic small obstacles with onboard sensing and computating on aerial robots," *IEEE Robot. Autom. Lett.*, vol. 6, no. 4, pp. 7869–7876, 2021.
12. C.-J. Wu *et al.*, "Machine learning at Facebook: Understanding inference at the edge," in *Proc. IEEE Int. Symp. High Perform. Comput. Archit.*, 2019, pp. 331–344.
13. S. Cass, "Taking AI to the edge: Google's TPU now comes in a maker-friendly package," *IEEE Spectr.*, vol. 56, no. 5, pp. 16–17, May 2019.
14. S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2017, pp. 3626–3633.
15. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agriculture*, vol. 147, pp. 70–90, 2018.
16. Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3-D point clouds: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 12, pp. 4338–4364, Dec. 2021.
17. S. Bargoti and J. Underwood, "Image segmentation for fruit detection and yield estimation in apple orchards," *J. Field Robot.*, vol. 34, no. 6, pp. 1039–1060, 2017.
18. N. Häni, P. Roy, and V. Isler, "MinneApple: A benchmark dataset for apple detection and segmentation," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 852–858, Apr. 2020.
19. G. V. Nardari *et al.*, "Place recognition in forests with Urquhart tessellations," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 279–286, Apr. 2021.
20. X. Liu *et al.*, "Robust fruit counting: Combining deep learning, tracking, and structure from motion," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 1045–1052.

21. W. Dong, P. Roy, and V. Isler, "Semantic mapping for orchard environments by merging two-sides reconstructions of tree rows," *J. Field Robot.*, vol. 37, no. 1, pp. 97–121, 2020.
22. N. Chebrolu, P. Lottes, T. Läbe, and C. Stachniss, "Robot localization based on aerial images for precision agriculture tasks in crop fields," in *Proc. Int. Conf. Robot. Autom.*, 2019, pp. 1787–1793.

XU LIU is currently a Ph.D. student in mechanical engineering and applied mechanics with the General Robotics, Automation, Sensing and Perception Laboratory (GRASP Lab), University of Pennsylvania, Philadelphia, PA, USA. His research focuses on large-scale autonomous navigation and semantic mapping in complex 3-D environments, such as farms and forests. Liu received an M.S.E. degree in robotics from the University of Pennsylvania in 2019. He is a member of IoT4Ag. Contact him at liuxu@seas.upenn.edu.

STEVEN W. CHEN is currently a Ph.D. student in computer and information science with the General Robotics, Automation, Sensing and Perception Laboratory (GRASP Lab), University of Pennsylvania, Philadelphia, PA, USA, and the Co-Founder and CEO of Treeswift, Philadelphia. His research interests include robotics, control theory, and machine learning, applied to perception, action, and communication using ground and aerial robots. Chen received an M.S.E. degree in robotics from the University of Pennsylvania in 2019. He is a member of IoT4Ag. Contact him at chenste@seas.upenn.edu.

GUILHERME V. NARDARI is currently a Ph.D. student with the Institute of Mathematics and Computer Science, University of São Paulo (ICMC-USP), São Paulo, Brazil, with funding from the São Paulo Research Foundation (FAPESP—2017/17444-0). His research interests include the use of semantic information in robot localization and mapping in cluttered and unstructured environments such as forests. Nardari received a B.S. degree in information systems with an emphasis on intelligent systems from ICMC-USP in 2016. Contact him at guinardari@usp.br.

CHAO QU is currently a Ph.D. student in computer and information science with GRASP Lab, University of Pennsylvania, Philadelphia, PA, USA. His research interests include depth completion, state estimation, and mapping. Qu received an M.S.E. degree in robotics from the University of Pennsylvania in 2014. Contact him at quchao@seas.upenn.edu.

FERNANDO CLADERA OJEDA is currently a Ph.D. student in computer and information science with GRASP Lab, University of Pennsylvania, Philadelphia, PA, USA. His research interests include perception and energy-efficient hardware architectures for robotics. Ojeda received an M.Sc. degree in embedded systems from the University of Rennes 1, Rennes, France, in 2013, and an M.S.E. degree in robotics from the University of Pennsylvania in 2019. He is a member of IoT4Ag. Contact him at fclad@seas.upenn.edu.

CAMILLO J. TAYLOR is with the Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA, USA. His research interests include computer vision and robotics, reconstruction of 3-D models from images, vision-guided robot navigation, and scene understanding. Taylor received a Ph.D. degree from Yale University, New Haven, CT, USA, in 1994. He is the recipient of numerous awards. Contact him at cjtaylor@seas.upenn.edu.

VIJAY KUMAR is the Nemirovsky Family Dean of Penn Engineering with appointments in the Departments of Mechanical Engineering and Applied Mechanics, Computer and Information Science, and Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. Kumar received a Ph.D. degree from The Ohio State University, Columbus, OH, USA, in 1987. He is a Fellow of the American Society of Mechanical Engineers (ASME) and IEEE. He is the corresponding author of this article. Contact him at kumar@seas.upenn.edu.