Applied Mechanics and Materials
ISSN: 1662-7482, Vol. 884, pp 77-85
doi:10.4028/www.scientific.net/AMM.884.77

Submitted: 2018-02-26
Revised: 2018-06-29
Accepted: 2018-07-01
Online: 2018-08-28

Overview of Sensor Technologies Used for 3D Localization of Asparagus Spears for Robotic Harvesting

Matthew Peebles, Shen Hin Lim , Mike Duke and Chi Kit Au

Science & Engineering
Mechanical Engineering Dept.
University of Waikato

Keywords: Asparagus harvesting, robotic harvesting, machine vision, detection systems, 3D sensors

Abstract. Advances in agricultural automation, coupled with a general decline of available labour has generated interest in automated harvesting of various crops. Paramount to the success of such systems is the development of accurate, robust detection technologies and localization strategies. This paper presents an overview of sensor technologies used in the detection and localization of green asparagus spears for robotic harvesting. Tactile, photoelectric, machine vision and time-of-flight sensors are investigated and their applicability for use in robotic asparagus harvesting is evaluated. Investigation of previous asparagus harvesting devices has revealed that no such device has yet achieved commercial viability. It was identified that this is likely due to weaknesses in currently employed detection technologies, namely slow response times, high sensitivity to changes in ambient lighting conditions and requirement for frequent manual calibration. Of the sensor technologies investigated it was found that time-of-flight cameras, such as the Microsoft Kinect V2 are the most feasible for the detection of asparagus spears for robotic harvesting. It was concluded that further research would be conducted into the application of such sensors into a commercially viable harvester.

Introduction

Agricultural automation is becoming increasingly popular in the world [1]. The prospect of cheap, reliable labour provides industry with a large incentive to invest in the development of robotic harvesters and other automation technologies. Despite the surge of research and development the asparagus industry has yet to receive significant benefits from automation. This is likely due to the complexity of the asparagus harvest. Asparagus spears emerge randomly distributed around their parent plants and typically grow at sporadic rates. Additionally, the beds are relatively unstructured and often contain significant weed growth and foreign debris [2]. Such randomness presents a real challenge for any automated system [3].

Attempts to automate the asparagus harvest have been made as early as 1952 [4] with machines like the Matteoli harvester, however, no machine has yet to see success as a commercially viable tool. This is usually because such devices either operate at slow speeds or cause excessive collateral damage to neighboring crops. Finding a balance between reliable detection, speed, and accurate localization seems to be key for realizing a commercially viable harvester. Paramount to achieving this is the development of an appropriate detection system.

This paper will present the results of an investigation into the applicability of several different sensor technologies for the detection of asparagus spears.

Tactile and Photoelectric Sensors

Tactile Sensors. Early selective asparagus harvesters such as the Matteoli and Turkington harvesters, as well as early versions of the Haws harvester utilized tactile sensors for the detection of spears [4, 5, 6]. Little information is known about the specific performance of these early harvesters, likely due to the commercial sensitivity of their respective developments, however an overview of mechanical asparagus harvesters performed in 2010 by Du et al. [3], reported that such devices tended to produce unreliable results. The researchers argued that the widely varying diameters and stiffness of individual spears resulted in a wide range of different response characteristics from the sensors. Additionally, it was found that due to the delicate nature of the spear tips such detection methods were not desirable because of the damage caused by contact with the sensors. It was also reported that the rigid nature of such devices made it hard to incorporate much flexibility in the design. This resulted in the developed systems performing poorly under real-world conditions due to changes in operating conditions like weather or field quality. Such inflexibility makes tactile sensors ill-suited for detecting asparagus spears, due to the crops propensity to grow unpredictably, and the difficulty of structuring its cultivation.

Photoelectric Sensors. Photoelectric sensors were implemented in the harvesters of the early 1990's to replace contact sensors on many of the early designs. Machines such as the CAMIA (Centre of Advanced Manufacturing and Industrial Automation) harvester, modern Haws harvester and Geiger-Lund Harvesters utilized photoelectric sensors in such a way as to replace direct contact from tactile sensors. These devices utilized photoelectric sensors coupled with various light sources to detect the passing of qualifying spears [7, 8, 9] . This allows for the elimination of the aforementioned spear tip damage. In the spring of 2006, the Geiger-Lund harvester underwent field trials in the USA. Clary et al. [10] performed an economic analysis of the harvester based on the results of these trials . The researchers reported that the Geiger-Lund harvester was able to harvest up to 70% of the harvestable spears per row. It was also reported that the harvester was able to detect, but failed to harvest an additional 9% of the marketable spears. This means that the harvester had a total detection rate of 79%. Such systems do not require sophisticated data processing to localize spears. Typically several gates are used to segment the asparagus bed longitudinally. Each of these gates is equipped with a sensor to detect the passing of qualifying spears within each gate. This simplicity is advantageous because it allows the device to operate quickly, with a maximum reported speed of 2.8m/s. The major flaw of this detection method is the extremely poor localization accuracy. Since the sensors can only provide pass/go type outputs the spears can only be located with the resolution of the size of each gate, resulting in a clumsy area harvest type approach by the harvesting effector. This causes unnecessary collateral damage, hurting the commercial viability greatly.

Machine Vision

Image Processing. Standard image processing techniques are common in machine vision applications. Studies such as that conducted by Humberg et al. [11] have used monochromatic CCD sensors to take grayscale images of asparagus beds and used standard grey level segmentation to separate the spears from the background. Algorithms such as these do not have a semantic understanding of the scene, rather they rely on the physical aspects of the scene to be constructed in such a way as to predictably ensure that the objects of interest are illuminated primarily. The unstructured nature of asparagus beds however makes this difficult to achieve in practice. Humberg's system required manual calibration of the threshold parameters, as well as per-row recalibration of the perspective transformation matrix used to localize spear bases based on their image-space XY coordinates. For this reason, the data was processed offline. Using this method it was reported that 86% to 95% of all spears were located within a 2.97mm by 5.39mm window [12]. This is an impressive detection rate, however the requirement for manual calibration makes it very hard to implement in a real-time system.

Attempts have been made to develop systems to support standard image processing techniques in order to mitigate their reliance of sensor calibration. Irie et al. [13] developed a vision system that used two slit laser projectors to project two horizontal laser lines on images of asparagus beds taken from a horizontal perspective . Although this work was only ever done under lab conditions. Using image processing techniques intersection of the projected laser and qualifying spears were found. Since the physical separation distance between the projected laser lines was known the spears were able to be localized trigonometrically. Using such an approach could reduce the required output quality of processed images, thereby reducing the sensors dependence on initial calibration. However, similarly to the work conducted by Humberg and his team, the system was reported to require rigid, and frequent calibration in order to achieve reasonable accuracy.

Both of these systems are hindered in application by the unstructured nature of real-world asparagus beds. Variable day-to-day lighting, as well as the need for the scenes to be rigidly structured, make unsupported machine vision of this type difficult to achieve in the real world.

A brief investigation into using standard image processing techniques for the detection of asparagus was conducted. A Basler ACE series camera was used to image asparagus spears in a lab environment. The images were illuminated using an array of 12V LEDs to provide frontal lighting similar to the work done by Grattoni et al [18]. Ambient lighting was minimised by cloaking the scene in PVC backed canvas. This also provided a dark background for the images, greatly improving the contrast between asparagus spears and background. OpenCV was used to binarize the image using standard grey level segmentation. This was done using several different threshold levels; the results can be seen in Figure 1. From the figure, it is clear that correct spear segmentation is only achievable when the threshold values are correctly set. Similarly to Humberg and his team, it was found that the ”correct” thresholding level varies greatly with changes in ambient lighting conditions and the geometry of the scene. This makes automatic calibration of the system difficult because the computer does not have a semantic understanding of what it should be looking for.

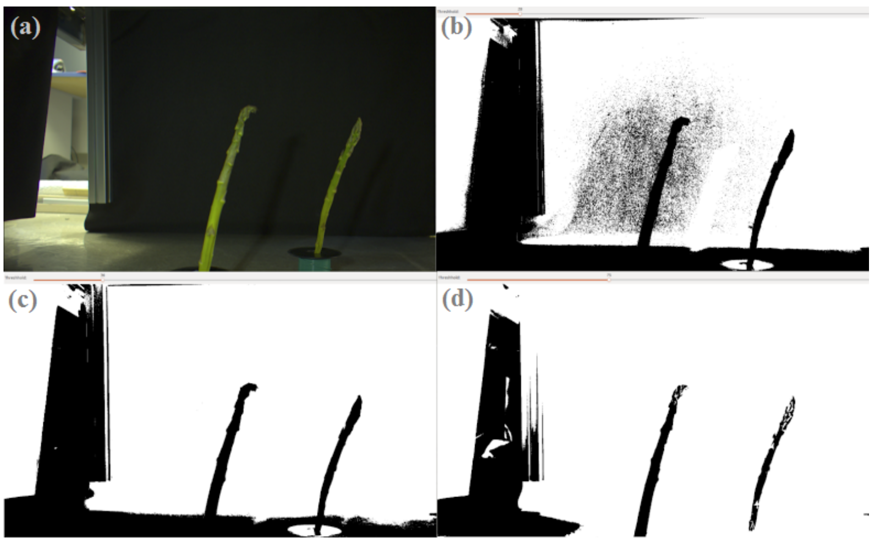

Fig. 1: Asparagus spears segmented using various threshold values: (a) Original image, (b) Threshold to low, (c) Good segmentation, (d) Threshold to high.

The same imaging apparatus was used for field testing during December of 2016. Images in Figure 2 are typical of the data collected from these tests. Even though the canvas cover provided excellent rejection of direct sunlight, ingress of diffuse sunlight and changes in scene perspective due to variations in row straightness generated a significant amount of variance in the observed pixel intensities. Similar results to those found in the lab testing were observed during the processing of these images. Changes in relative image intensity values between spears in different images resulted in a requirement for frequent recalibration of the binarization threshold. Additionally, weed growth and foreign debris present in the field were found to be difficult to segment from the target spears.

It was concluded that techniques to automatically calculate the threshold value based on feature detection algorithms would need to be developed in order to make this segmentation strategy viable. Personal communication with industry experts [15] revealed a typical issue in applying such techniques to unstructured environments. The computational cost of feature matching algorithms is generally prohibitivly high when applied to unstructured scenes due to the large amount of filtering and other compensatory work that is required. This means that such processing is often very difficult to achieve in real time. Artificial neural networks and other machine learning techniques can be applied to provide the feature matching functionality. Although these techniques are generally slower than standard feature matching algorithms, their inherent robustness and insensitivity to unstructured environments can often mean that they outperform purely image processing based approaches.

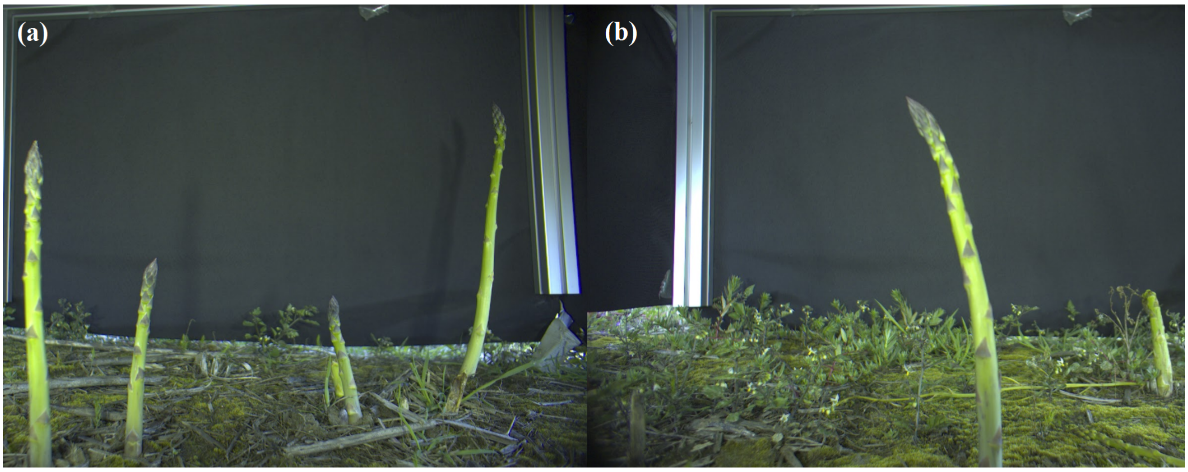

Fig. 2: Comparison of observed light intensity between spears during field testing: (a) Low intesnity image due to long range between spears and camera, (b) High intensity due to close range

Stereoscopic Vision. Stereo vision is commonly used for 3D localization in agricultural automation. Two camera stereo-pairs are typical, however multi-camera systems such as that developed by Mehta et al. [14] have also seen some success. Such systems generally require considerable computational power in order to achieve reasonable response times. Typically such systems would operate on a per-pixel basis, finding discrepancies between individual pixel datasets, however, for online applications this approach is too slow. Instead a feature matching type approach is used where features are detected in each of the images using more standard image processing techniques, the key points of which are then localized stereoscopically [15]. This however, introduces the inherent weaknesses of image processing described in the previous section making such a system much more susceptible to erroneous data due to fluctuating ambient light conditions and incorrect calibration. As such stereo pairs used in online applications tend to use structured lighting to mitigate these effects.

Several groups have worked on applying stereo vision to the localization of asparagus. The earliest work found was that conducted by Baylou et al. [16] in 1984. Their work was concerned with the identification and localization of emergent tips of white asparagus. The researchers used strong backlighting to produce high contrast silhouettes of the beds, allowing a lot of the problems associated with variable lighting to be eliminated. However, in only imaging the silhouettes of the spears much of the foreground information and high-resolution details of the scene were lost. This made problems such as spear occlusion very difficult to solve. The researchers used a statistical algorithm for feature detection in the images. Congruent features were then used to localize the identified spear tips stereoscopically. Although the success of this approach was not reported a patent for using such a system in a robotic harvester was filed in 1984 [17].

Grattoni et al. [18] also developed a stereo-vision system for detection of asparagus. Similarly to the work conducted by Baylou et al, the researchers aimed to reduce the amount of variability in the lighting conditions by using powerful frontal lighting coupled with a dark background. This provided a saturated, high contrast image of the spears, which could more reliably be segmented using standard image processing techniques. Although easier to segment, such images lose significant information regarding the higher resolution detail and form of the spear. This resulted in ambiguities in the image processing and subsequent localization, resulting in situations where multiple spear distributions could be explained by the same input data. The researchers used an XYZ manipulator to take several stereo images from a series of known perspectives to narrow down the possible and most likely spear distributions. On average Grattoni et al reported a detection time of 3.5 seconds, however they speculate that with some improvements and under ideal conditions that this number could be as low as 0.36 seconds. Although this acquisition time is promising it is unlikely that such a method could be implemented into a continuously operational harvester. Significantly decreasing the commercial viability of any robotic harvester which utilizes this system.

A collaboration of industry partners and the University of Bremen have been working on a robotic harvester (AmLight project) with funding from the European Commision. One of the main companies involved, Strauss Verpackungsmaschinen GmbH, has released a report outlining some of the operational principles of the device [19]. The device uses two CMOS sensors in a stereo pair to identify and localize asparagus spears. A series of normalizing algorithms and white balancing methods are used to reduce the effect of variable ambient lighting conditions. Feature detection based on standard image processing techniques is then used to identify key points in the images and localize the spears stereoscopically. Results of field testing in 2014 revealed that slow image processing times rendered much of the data unusable in real-world environments. The researchers developed a mock-up field and tested the device under laboratory conditions where it was reported that the system cold achieves a detection rate of 70% while traveling at 0.5m/s.

Time-of-Flight Technologies

Scanning Laser Rangefinder. Time-of-flight technologies are promising because of their ability to provide accurate range data even under variable ambient lighting conditions. However such systems are typically sensitive to the reflectivity of objects in the scene.

Sakai et al. [20] utilized a Hokuyo URG_04LX scanning laser rangefinder to develop a system for detecting asparagus spears for indoor greenhouse conditions. The system used the rangefinder to generate several 2D planar scans while translating the scanner vertically. These scans were collated to generate a 3D point cloud. The geometry of the asparagus beds seen in this study differs significantly to those typically seen in New Zealand. The greenhouse environment was structured such that a row of mature parent plants were present in the center of each row. This means that the detection system needed to be able to discern qualifying spears from parent plants. The point clouds were analyzed to both identify qualifying spears and to localize the point of harvest. Using the reported methods the researchers were able to achieve a detection rate of 75%. However, due to the method by which the point cloud is generated it was reported that the total detection time was about 2 seconds per spear. The stop-start nature of this method is not applicable to continuous harvesters and greatly reduces the efficiency.

Irie & Taguchi [21] tried to improve upon this research by having the laser scanner rotate upwards, rather than translate to develop the point cloud. It was found that such advances did not greatly improve the detection times, but rather only simplify the mechanics of harvesting the spear.

Lab testing of a Hokuyo URG_04LX scanning laser range finder was done to investigate the feasibility of such sensors for asparagus detection. Figure 3 shows a typical scan result from the Hokuyo sensor. Using a simple clustering algorithm, which separated objects based on both the scan angle and range, it is possible to generate reasonable estimates of the location of asparagus spears. Taking the centroid of points in each cluster gives an approximation that is within 10mm of the actual spear center in most situations. This sensor did not perform well outdoors. This is likely due to interference of near-infrared light from the sun. This interference resulted in scan data that was very noisy and in some cases failed to detect spears at all.

The laser scanner has a reasonably fast response time of 10ms, however the amount of information obtained with each scan is minimal (approximately 720 points). The sparsity of the scans meant that in many cases identification of target spears was ambiguous, especially when large amounts of weed or foreign objects were present in the scene. It was considered that using multiple sensors in various configurations might be able to rectify this issue, however simultaneously running multiple sensors

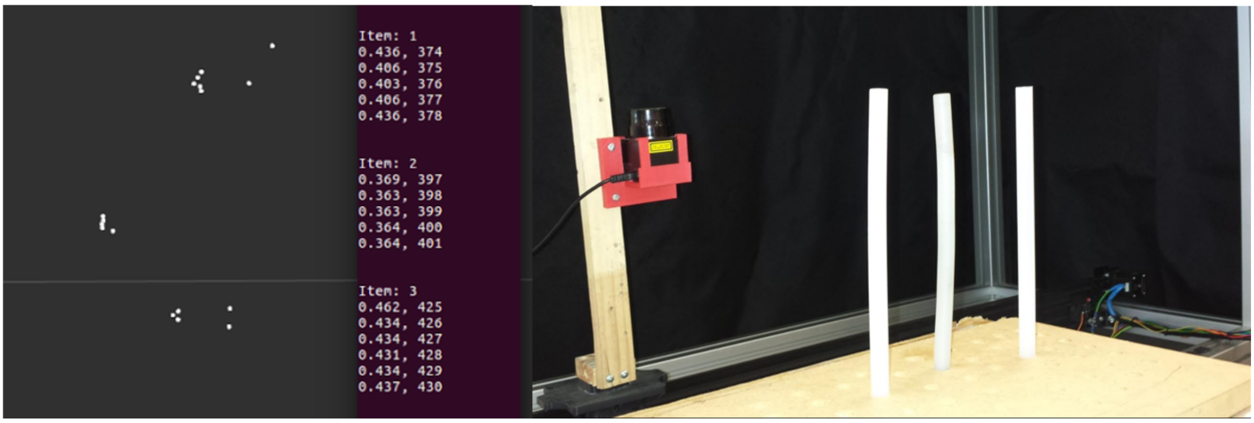

Fig. 3: Seperating spears based on range and angle

resulted in noisy data due to interference from the multiple light sources. This problem could not be solved without more sophisticated hardware. If the hardware allowed, it would be possible to time the scans such that they were interlaced and provided less interference. This would however reduce the achievable framerate of the scan significantly.

Time-of-Flight Cameras. Time-of-flight cameras have yet to be applied to the identification and detection of asparagus spears. However the technology is relatively prevalent for other detection applications

A lot of work has been done on using time-of-flight technology for the identification of humans and gesture recognition [22, 23, 24]. These applications require high accuracy, insensitivity to changes in ambient lighting intensities and robustness. These attributes make time-of-flight sensors an attractive option for the detection of crops for robotic harvesting.

Sun et al. [26] utilized time-of-flight cameras in the detection of citrus fruit. Their system used grayscale intensity images and standard segmentation techniques to detect potential citrus fruit. Using the segmented fruit as a mask, relevant depth information was extracted from the depth map provided by the time of flight sensors. A detection rate of up to 81% was achieved when the sensors were within a range of less than 1.5m, however at longer ranges this detection rate fell to under 66.7%. This is likely due to the low-resolution nature of time of flight images and the inverse square decrease in incident light intensity of the camera with range. One of the biggest troubles reported in the study was the ability to segment the depth information pertaining to target fruit from the noise readings associated with the background. In this case the background was a tree canopy. Such backgrounds can be geometrically complex. It was suggested that major improvements would be seen if this problem could be solved.

A feasibility study of the use of time-of-flight cameras in the detection of asparagus was made. A Microsoft v2 Kinect time-of-flight camera was used to image asparagus spears in both lab and field conditions. The Kinect camera has a depth map resolution of 512x424, and good background light rejection capabilities. These factors mean that unlike laser scanners the Kinect is able to receive a much larger amount of data per scan. The background light rejection allows the device to perform reasonably well in outdoor environments. These factors allowed relatively high quality images of the field to be obtained. Figure 4 show typical images captured at a range of approximately 2 meters using the Microsoft Kinect.

The figure shows two views of a group of asparagus spears, as well as an image of a section of typical weed growth. It was found that the range precision of the pointcloud was approximately 5-10mm. This was deemed acceptable for robotic harvesting.

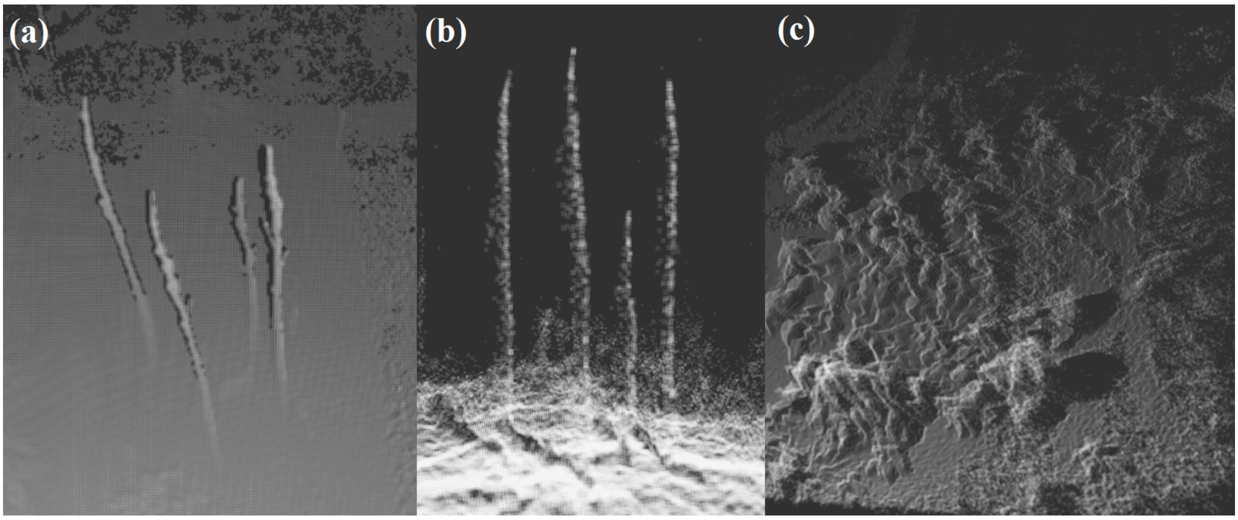

Fig. 4: Images using Microsoft Kinect v2 time-of-flight camera: (a) Top-down view of spears in real-world conditions, (b) Frontal view of the same group of spears, (c) Typical weed growth.

Conclusion

Based on the investigation many advantages and disadvantages of various sensors have been identified. Image processing, while being relatively inexpensive is extremely sensitive to calibration and ambient light intensity. This means that it is not robust which is very undesirable for application in unstructured environments such as asparagus beds. Similarly stereo vision technologies are plagued with the same problems as standard machine vision. This is because typically to achieve the required processing speed, feature-based detection, based on the same flawed image processing techniques must be made. It has been shown that some of these dependencies can be eliminated by producing high contrast, or backlight images, however the trade-off is a complete loss of information regarding higher resolution details of the scene. This, while improving robustness and ranging capabilities greatly, also makes the problem of spear occlusion very hard to solve.

Time-of-flight technologies seems to be the most resilient to unstructured and variable ambient lighting conditions. Since these sensors do not require persistent day-to-day calibration, they are also more robust in the localization of spears. Commercial viability requires that the sensors can report at a high rate. This allows more continuous harvesting which greatly improves harvesting efficiency. Consequently, this requires sensors operating from time-of-flight principals, such as time-of-flight cameras and scanning laser rangefinders to accurately account for motion in real time. This is a non-trivial task for application in field environments where the exact motion of the sensor may not be known.

Scanning laser rangefinders were found to provide fast and accurate measurements of depth, however the sparsity of the input scans was deemed to be unsuitable. This resulted in ambiguities in identification of target spears.

Time-of-flight cameras seems like the most promising technology due to their high frame rate, relative insensitivity to ambient lighting conditions and large amount of data included in each scan. Additionally, because depth calculations are performed on the cameras hardware, minimal processing power is required to localize points of interest. Initial testing has revealed that multi-path interference and motion artifacts could be problematic, however further testing is required to quantify their impact.

The physiology of asparagus lends itself well to detection via time-of-flight cameras. This is because asparagus spears grow in such a way that it is possible to control the range and ambient lighting conditions. Additionally, it is possible to position the camera in a way that minimises the complexity of the background. This results in a clear distinction between target spear and the asparagus bed in the pointcloud images. The point-cloud data obtained is sufficient for both localization and classification of spears, allowing a high degree of control over which spears are selected for harvest.

Based on the results of this investigation time-of-flight sensors will be investigated further as a potential detection system for use in a commercially viable asparagus harvester.

References

[1] LLP, A.R., *Agricultural Robots Market Analysis &; Trends - Product, Technology - Forcast to 2025.* 2016. p. 113.

[2] Lewis, G., *Personal Communication*, M. Peebles,2016.

[3] Du, C., Z. Qin, and W. Shumao, *Current Status and Technical Challenges of Asparagus Mechanical Harvesting.* 2010 Pittsburgh, Pennsylvania, June 20 - June 23, 2010.

[4] Matteoli, A.J., *Asparagus harvester*. 1952, Google Patents.

[5] Turkington, J.O., *Asparagus harvesting machine*. 1956, Google Patents.

[6] Haws, S.K., *Stalk selective harvesting machine*. 1977, Google Patents.

[7] Arndt, G., R. Rudziejewski, and V.A. Stewart, *On the future of automated selective asparagus harvesting technology. Computers and Electronics in Agriculture*, 1997. 16(2): p. 137-145.

[8] Lund, W.J., *Asparagus harvester*. 1985, Google Patents.

[9] Haws, S.K.R., WA, US, *Selective Harvester*. 2010: United States.

[10] Clary, C.D., et al., *Performance and economic analysis of a selective asparagus harvester. Applied Engineering in Agriculture*, 2007. 23(5): p. 571-577.

[11] Humburg, D.S. and J.F. Reid, *A machine vision algorithm for identification of harvestable spears of asparagus*. 1990. p. 13pp.

[12] Humburg, D.S. and J.F. Reid. *Field performance of machine vision for the selective harvest of asparagus*. in International Off-Highway and Powerplant Congress and Exposition, September 9, 1991 - September 12, 1991. 1991. Milwaukee, WI, United states: SAE International.

[13] Irie, N., et al., *Asparagus Harvesting Robot Coordinated with 3-D Vision Sensor*. 2009 Ieee International Conference on Industrial Technology, Vols 1-3, 2009: p. 408- 413.

[14] Mehta, S.S. and T.F. Burks, Multi-camera Fruit Localization in Robotic Harvesting. IFAC-PapersOnLine, 2016. 49(16): p. 90-95.

[15] Jones, M., *Personal Communication*, M. Peebles,2016.

[16] Baylou, P., et al., *Detection and three-dimensional localization by stereoscopic visual sensor and its application to a robot for picking asparagus*. Pattern Recognition, 1984. 17(4): p. 377-384.

[17] Bousseau, G., et al., *Automatic asparagus picking machine.* 1984, Google Patents.

[18] Grattoni, P., et al. *Automatic harvesting of asparagus: an application of robot vision to agriculture*. 1994.

[19] Strauß, J., *Development of an Automatic harvesting system for green asparagus with stalk detection in Ambient Light*. 2014. p. 1-14.

[20] Sakai, H., et al., *Accurate position detecting during asparagus spear harvesting using a laser sensor*. Engineering in Agriculture, Environment and Food, 2013. 6(3): p. 105-110.

[21] Irie, N. and N. Taguchi, *Asparagus harvesting robot.* Journal of Robotics and Mechatronics, 2014. 26(2): p. 267-268.

[22] Luna, C.A., et al., *Robust people detection using depth information from an overhead Time-of-Flight camera*. Expert Systems with Applications, 2017. 71: p. 240-256.

[23] Stahlschmidt, C., et al., *Applications for a people detection and tracking algorithm using a time-of- flight camera*. Multimedia Tools and Applications, 2016. 75(17): p. 10769-86.

[24] Kapuciski, T., M. Oszust, and M. Wysocki. *Hand gesture recognition using time-of- flight camera and viewpoint feature histogram*. in 11th International Conference on Diagnostics of Processes and Systems, DPS 2013, September 8, 2013 - September 11, 2013. 2014. Lubuski, Poland: Springer Verlag.

[25] Oprisescu, S., C. Burlacu, and V. Buzuloiu. Action Recognition using Time of Flight Cameras. in 2010 8th International Conference on Communications (COMM), 10-12 June 2010. 2010. Piscataway, NJ, USA: IEEE.

[26] Sun, L., J.-R. Cai, and J.-W. Zhao, *Vision System Based on TOF 3D Imaging Technology Applied to Robotic Citrus Harvesting*. Intelligent Automation &; Soft Computing, 2015: p. 1-10.