



2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 152181371

Plant Recognition through the Fusion of 2D and 3D Images for Robotic Weeding

Gai, Jingyao¹; Tang, Lie¹; Steward, Brian¹

¹Iowa State University.

**Written for presentation at the
2015 ASABE Annual International Meeting**

**Sponsored by ASABE
New Orleans, Louisiana**

July 26 – 29, 2015

Abstract. *In crop production systems, weed management is vitally important. But both manual weeding and herbicide-based weed controlling are problematic due to concerns in cost, operator health, emergence of herbicide-resistant weed species, and environment impact. Automated robotic weeding offers a possibility of controlling weeds in a precise fashion, particularly for weeds growing near crops or within crop rows. However, identification and localization of plants have not yet been fully automated. The goal of this reported project is to develop a high-throughput plant recognition and localization algorithm by fusing 2D color and textural data with 3D point cloud data. Plant morphological models were developed and applied for plant recognition against different weed species at different growth stages.*

Keywords. *Robotic weeding, computer vision, sensor fusion*

Introduction:

The control of weeds in vegetable crops, particularly in the intra-row region, can be accomplished with lower levels of labor, energy, and chemical inputs with the engineering development of field robotic automation technology. This will lead to improve agricultural sustainability and improve nutritional benefits through expanded production of these crops. However, to be able to realize intra-row weed controlling automatically, which is more complicated than weeding between rows, right treatment must be applied at the right place and right time. Thus it is important to realize in-situ plant discrimination and localization.

In order to realize in-situ plant discrimination and localization, computer vision technic is a good choice. Computer vision has been shown to provide an option in inspection of agricultural products, particularly when color and shape need to be analyzed at high speed. Many applications in agricultural robotics like plants discrimination

The authors are solely responsible for the content of this meeting presentation. The presentation does not necessarily reflect the official position of the American Society of Agricultural and Biological Engineers (ASABE), and its printing and distribution does not constitute an endorsement of views which may be expressed. Meeting presentations are not subject to the formal peer review process by ASABE editorial committees; therefore, they are not to be presented as refereed publications. Citation of this work should state that it is from an ASABE meeting paper. EXAMPLE: Author's Last Name, Initials. 2015. Title of Presentation. ASABE Paper No. ---. St. Joseph, Mich.: ASABE. For information about securing permission to reprint or reproduce a meeting presentation, please contact ASABE at rutter@asabe.org or 269-932-7004 (2950 Niles Road, St. Joseph, MI 49085-9659 USA).

and self-guidance can be realized with the ability of computer vision (ASTRAND & BAERVELDT, 2002). And plant morphology and structure have been focused on, which remains one of the most consistent methods of plants identification (Du, Wang, & Zhang, 2007).

At present, 2D vision based approaches are very common for environment sensing in agricultural robotics. But the disadvantage of 2D vision is it is sensitive and unstable to changing light condition. 3D sensors are not widely used due to its low resolution and high cost. However, one obvious advantage of 3D sensor over 2D vision is its simplicity in discovering morphology and structure data from the data acquired. As a result, it would benefit if both 2D vision and 3D vision are utilized together.

Thus the objective is to investigate the use of 2D textural data and 3D spatial data with plant morphological models for high-throughput in-situ plant discrimination and localization.

This paper is organized as follows. In the first part of this paper, we discuss the Kinect v2 sensor which fusions 2D textural data and 3D spatial data. We also compare this sensor with other 3D sensors. The second part of this paper is focused on an algorithm performing plant detection and localization using both 2D and 3D data from Kinect 2 sensor. The proposed algorithm is based on an approach detecting the ground to remove the points from soil. In the remaining point cloud it searches the clusters represent plants, and extract plants features for classification. In the third part of this paper, it shows the experimental results using these algorithms. Finally, the conclusion and future works are presented.

Sensing in agricultural robotics

1. 3D sensors in robotics

2D sensing is justified to some extent for indoor sensing applications like object recognition, building maps and self-localization. However, these approaches will not work properly in outdoor applications like agricultural robotics where there are no flat grounds or upright walls, as well as complex environment, especially light conditions. Here, 3D sensors promise to give the required information to perform plant discrimination, self-localization, mapping, etc. It is becoming popular to apply 3D computer vision in agriculture applications like (Li J. , 2014), (Jin & Tang, 2009) and (Nakarmi & Tang, 2012). The advantages of 3D sensors for plant discrimination and localization are obvious: It is much easier to get the 3D structural and morphological data of the plants. On the other hand, 3D sensor can provide reliable distance information. And they are useful for save and robust applications. Up-to-date 3D sensors like photon mixer devices (PMD) and Laser sensors also outputs intensity values, stereo vision cameras even provide color values.

Today three sensor technologies are mainly in use for 3D sensing on mobile robots: stereo vision, laser sensors and PMD time-of-flight cameras. Works like (Sanson, Trebeschi, & Docchio, 2009), (Weiss & Biber, 2011) compared and evaluated applications for 3D sensors on mobile robots:

To receive 3D data from normal cameras, typically stereo systems with two cameras or structure from motion techniques are used. In the work of (Jin & Tang, 2009), a real-time corn sensing system was developed using stereo vision. However, due to their passive operation mode, it is hard for both to provide reliable data for accurate sensing. To receive 3D data from stereo vision, structures or features in the images are required and used to calculate disparity. Further, the precision and maximum depth is limited by the baseline between the cameras, and the quality of the distance values decreases very fast as depth increases. Big advantages of the stereo vision are the high resolution images and the availability of color values. Further, stereo systems are relatively small and low priced, because one can use standard vision components. On the other hand, the cameras can also be modified to near-infrared cameras by replacing the filters with NIR filters to get high contrast images where living plant material and soil is easily discriminated (Jr. & Hively, 2010).

Among the 3D laser sensors the price range as well as the accuracy, resolution and frame rate is wide spread. Some of 3D laser sensors are implemented from 2D laser scanner. One example is Kurt3D (Surmann & Nüchter, 2003) project who equips a rotating 2D SICK laser to realize 3D laser scanning. However it needs a stop-and-go mode for travelling to receive consistent 3D data, as a reason of the 2D line sensor was not built for 3D applications. There are other groups using highly precise 3D laser sensors. However, these sensors usually have the properties of high weight, high power consumption and expensive prices.

The semiconductor based PMD cameras are the latest technique. It measures distance in addition to the common grayscale intensity information based on the time-of-flight by a modulated light source. The modulated source light signal is reflected by the environment and absorbed from this sensor (Li L. , 2014). Then the distances are calculated by the phase shift ϕ of the signal and the reflection intensity by the signal amplitude. Most light sources

used by the time-of-flight cameras are buildup using LEDs with a modulation of $f_{mod}=20$ MHz which allows an unambiguous range measuring up to 7.5 m calculated by equation below. There are many research applications using PMD as sensing devices in agriculture, especially in phenotyping (Alenyà, Dellen, Foix, & Torras, 2012). The limitations for current PMD sensors are: the quality of the depth values depends on the color of the material reflecting the emitted light, and some sensors have blur problems with moving objects. Although there are limitations, PMD cameras have not yet reached their full potential: from version to version the resolutions are increasing while the prices decrease and they are also getting more robust, like Kinect 2 developed by Microsoft.

$$d = \frac{c * \varphi_0}{4\pi * f_{mod}} \quad (1)$$

2. Kinect 2:

Kinect v2 (Fig.1 (a)) is a new version of Kinect sensor, provides RGB, IR (Infrared) and depth image like its predecessor Kinect v1. The 2D color information can be registered to the 3D space point with sensor's relationship provided in Kinect SDK (see Fig.1 (b)). It's amazing because it offers a way to combine the advantages of both 2D color camera's high resolution as well as 3D depth sensor's light changing insensitiveness together. Due to its low price, it is widely adopted for various computer vision applications, including 3D reconstruction, object recognition and some Human-Interface applications. Further, the resolution for depth is 512x424, which is much higher than other common commercialized depth sensor, for instance Swiss Ranger. Another advantage of the Kinect v2 sensor is that it uses 3 strong IR emitters as light sources. They enable the Kinect v2 to work outdoor, with indirect sunlight.

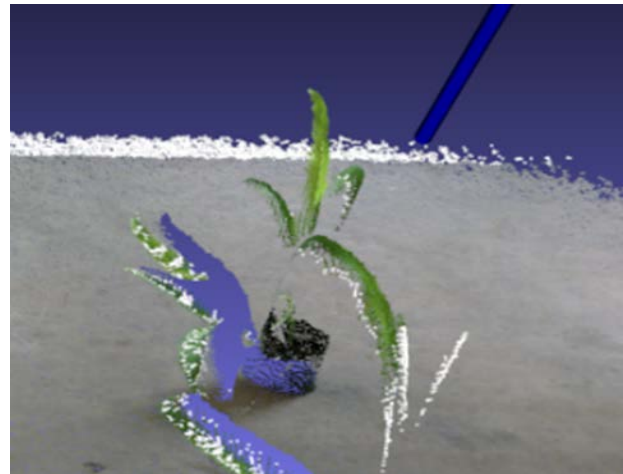
Table 1 compares a typical stereo vision camera and a laser sensor with the Kinect v2 sensor, used by us. The main advantage of the Kinect v2 sensor, compared with both other technologies, is the low price, with high-speed, high resolution 3D depth image fusing with 2D color information.

Table 1. Specs of 3 typical 3D sensors

Sensor type/ name	Stereo vision/ Bumblebee2	Laser/ InfiniSoleil FX8	PMD/ Kinect v2
Manufacturer	Point Grey	Nippon Signal	Microsoft
Operation mode	Passive	Active(laser)	Active (IR illumination)
Resolution	640 x 480	100 x 60	512 x 424
Dimension (mm)	157 x 36 x 47.4	62 x 66 x 89	260x60x70
Price (USD)	2,000	6,000	199



(a) Kinect v2 sensor



(b) 3D point cloud fused with color

Figure 1. The picture shows (a) Kinect v2 sensor used in this project. (b) The colored point cloud output from fusing 3D depth information and 2D color information.

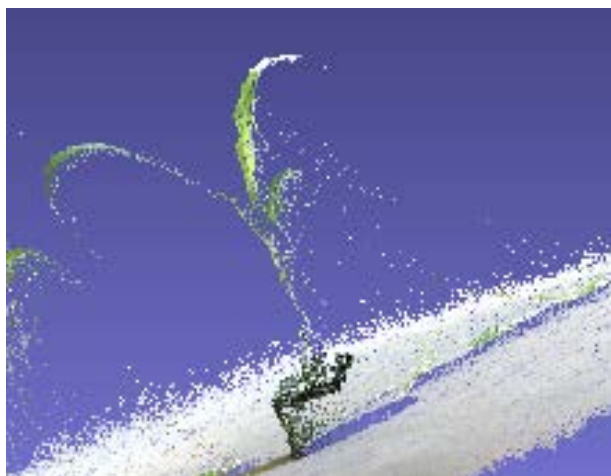
Plants identification and localization:

With 3D sensors, many applications can be realized in agricultural robotics. The main objective of our research is plants identification and localization. Our algorithm discriminates the ground and plants using morphological models of plants, and localizes the plants according to the relationship between the sensors and the operating mechanism. Therefore, the algorithm takes a 3D point cloud with 2D color information, the pose and the velocity of the instrument as input.

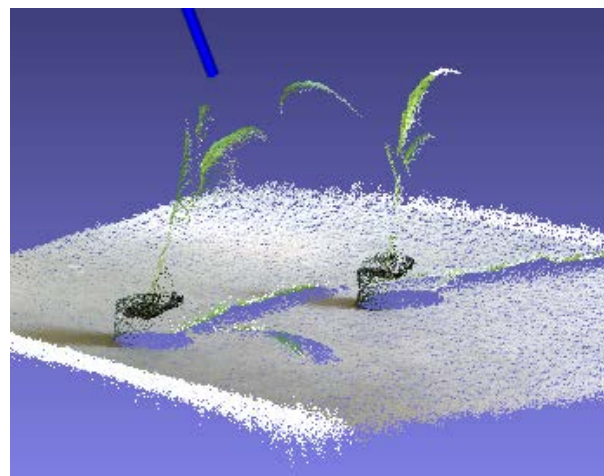
Above all, preprocessing is needed to roughly remove the background and remove noise points from the sensors. This procedure is done by using a cut-off filter and a neighbor count filter. After preprocessing, the ground is detected using both 2D color and 3D depth information, and a plane equation is determined. The remaining points which are belonging to plants are extracted into a new point cloud. In the point cloud, the algorithm searches for clusters representing plants, and labels are added to the clusters. For each of these clusters the probability to be a crop is calculated using a statistical model of a crop row. Then crop classification is applied to the plant clusters after extracting features of the plants. The positions of the crops are determined according to the filtered ground plane and transformed into the world coordinate system using the robot pose. Finally, the positions are stored in the map, as soon as the tracked plant disappears from view. Below, the single steps will be described in more detail.

1. Preprocessing

The output images of Kinect v2 sensors (Color, depth, Intensity) are stored, and the 2D color and 3D depth images are fused to a colored points cloud using functions provided by Kinect SDK. Before actually analyze the data, preprocessing is needed as a preparation procedure. In this procedure, useful data is extracted from the raw data, with the disturbance from noise and information useless. At the same time, data size is also reduced. In our algorithm, first is applying a cut-off filter to remove points which are outside a defined bounding cube. With this filter, bad points from the sensor and points too far away are removed. After that, a filter to remove sparse noise is applied. In this case, we are using neighbor count filter, which searches the surrounding points to find neighbor points, and calculates the neighbor count for every point in the cloud. Then it removes the point with less neighbors than threshold (see Fig 2). This algorithm is a simplified version of statistical outlier filter stated by Rusu (Rusu, 2009), in order to get better performance.



(a) Point cloud with sparse noise



(b) Point cloud after filtering

Figure 2. Pictures showing differences before and after preprocessing. The point cloud is generated from corn crops in laboratory. After applying preprocessing procedure, spare noise and bad points are removed.

2. Ground detection

The first step of the algorithm is to detect the plants in the 3D sensor output by segmenting the point cloud into background set containing the ground, and the plants. We are able to see the ground because the sensor is looking down at the plants vertically or at a tilt angle of 30 to 45 degree. To detect the ground the Hessian plane equation is fitted with d , the distance to the plane and n , the plane normal, as a model into the data.

$$n \cdot x - d = 0 \quad (2)$$

For the plane estimation the Random Sample Consensus (RANSAC) Algorithm is used, as its ability to deal with a large number of outliers and its computational simplicity. Therefore, three points from the point cloud are

selected by random to set up Eq. (2). Afterward, the number of inliers, by counting the points whose distance to the plane are within certain limits, is determined. This step is shown in Fig.3 (a). This step is iterated several times and the randomly chosen plane with the largest number of inliers is used for a refinement step. Those inliers are considered as ground (Fig.3 (b)). The refinement of the plane is done by using a least square fit, and the resultant plane is also used to calibrate the pose of the robot, as well as compensate the disturbance of vibration. Sometimes the algorithm may fail due to too many outliers, which happens when the ground is covered by weeds. This failure can be identified with the sensor position and robot pose, which are known in this case. If the normal and distance are far away from the expected values, the model is considered to be false. And the expected model calculated with sensor position and robot pose is applied.

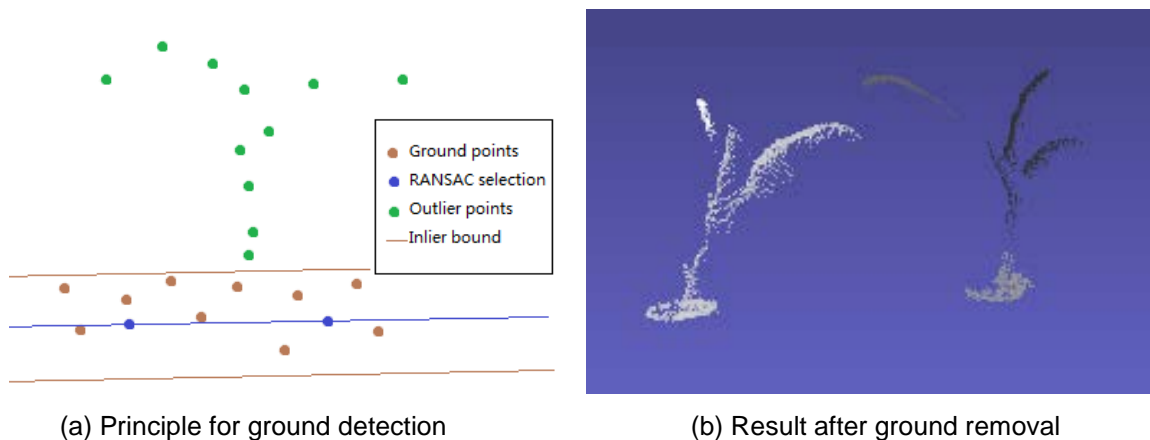


Figure 3. Fig (a) shows a synthetic 2D point cloud of a plant with ground plane to visualize the principle of the ground detection. With RANSAC algorithm, the blue points are randomly selected and a plane is fitted using these points. The points between the brown lines become inlier standing the ground. Fig (b) shows a point cloud with only outliers after ground detection. The points of the ground are successfully removed.

3. Plants extraction refinement

After removing the ground, the complement points should be from plants and those points should be separable. However, this may fail if the ground is not smooth or there are some other disturbing objects in the view. Then 2D color information from the sensor can be utilized to compensate the error in this case. Generally, the objects with green color can be considered as plants. However, the problem is the difficulty to extract green pixels in non-uniform illumination, for instance a green plant with shadow on it.

In this case, illuminant-invariant image space is used to extract green pixels with the disturbance of non-uniform illumination. Finlayson et al. have indicated that if the lighting is approximately Planckian and having Lambertian surfaces imaged by three delta-function narrow-band sensors, it is possible to generate an illuminant-invariant image (Drew, Finlayson, & Hordley, 2003). Under these assumptions, a log-log plot of two dimensional $\{\log(R/G), \log(B/G)\}$ values for any surface forms a different parallel straight line under different illumination. Thus, lighting change is reduced to a linear transformation along an almost straight line in the plot. And the characteristic straight line, which is perpendicular to the parallel lines, with its characteristic angle are obtained by calibrating the camera using method stated by Alvarez (J.M. Alvarez, 2008). The reduced values are used to characterize different colors.

During the experiment, the characteristic angle is optimized to be 40 degree, and the range for green plant color in illuminant-invariant image space is determined to be -0.9 to -0.2 using statistical methods. Points with color in range are extracted as points belonging to green plants (Fig.4).

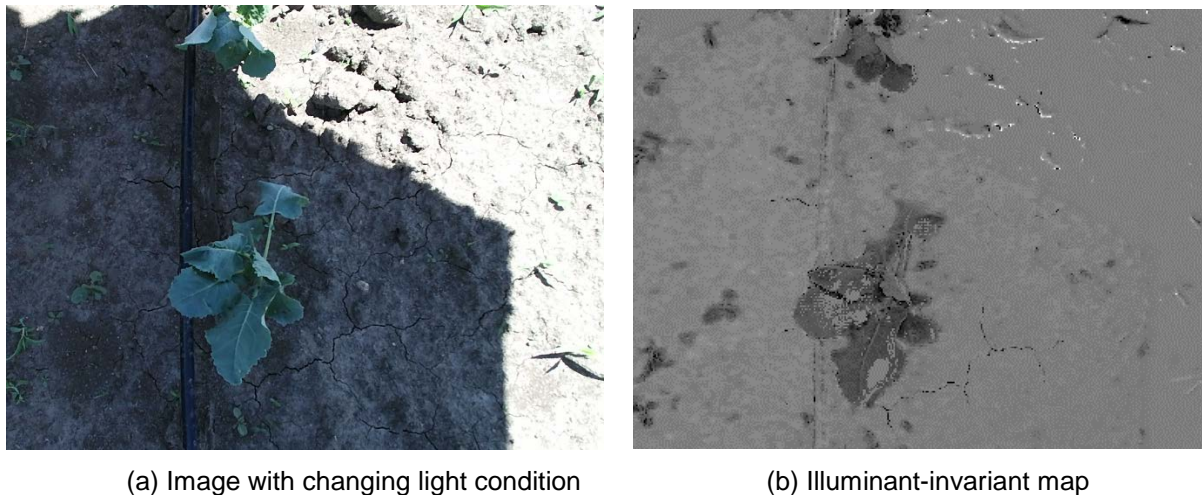


Figure 4. Figures showing the effect of illuminant-invariant map. In Fig (a), the light conditions are different between shaded and unshaded area. And in Fig (b), the illuminant-invariant map is generated, and compensates the effect of changing light conditions. It simplified the extraction of green points.

4. Plants detection and localization

With the plants extraction using the combination of RANSAC ground detection and illuminant-invariant method, points from plants in green can be extracted. Then clustering is applied to the remaining point cloud, in order to separate the point cloud into different clusters indicating different plants.

The algorithm of clustering is based on region growing in 3D. Therefore the point cloud is organized in a k-d tree for a fast nearest neighbor search. The region growing starts at a randomly selected seed point. The neighbors of this seed point within a smaller distance than the threshold are selected into the same cluster. Then the search continues on the newly joined points iteratively, until no more neighbors found. After finishing on this cluster, a seed point is randomly selected in the remaining points. Then the searching starts again in the remaining points, until every point is assigned into a cluster. In the final step the center and the standard deviation for each cluster are calculated. Clusters with a distance smaller than the added standard deviation between the centers are merged. And clusters with few members will be removed as noise or weeds. (Fig.5)

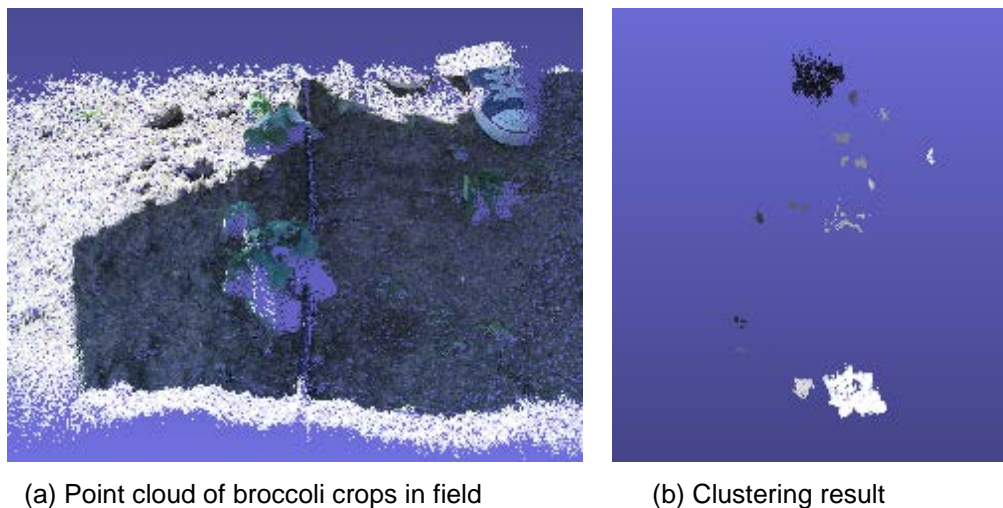


Figure 5. Figures showing plants clustering. Fig (a) is the point cloud generated from Kinect v2 sensor after preprocessing. And Fig (b) is the result after background removal and clustering. Different clusters are colored with different grey scale values. Each cluster stands for a plant. Larger clusters are identified as potential plants, and smaller ones as weeds.

After all points are clustered, the points are transformed into world coordinate system. Thus the positions of the plants are determined using the medians of the remaining clusters into the x and y direction. Then a statistical model is applied to these clusters, and for each of these clusters the probability to be a crop is calculated using the information of the crop rows. The reason is that the crops are assumed to be in lines, with normal vertical and horizontal error in position. Then a test is performed for each cluster with a null hypothesis: the cluster is not crop. For those clusters failed to be rejected to be not crop, combination will be applied with corresponding

clusters from previous frame for more information. The detail for detecting plants statistically will be presented in future papers. After testing, the positions of the potential crops are determined and some of weeds are removed.

5. Feature extraction

After detecting and localizing the plants, as well as removing some of non-crop plants, features are extracted from the plant clusters for recognition. There are many features can be used to distinguish crops from weeds, which in several categories: color based, plant morphology based, and plant structural based features.



Color based features are not reliable for green leaf plants, as they are sharing similar colors with weeds. As a result, plant morphology and structural based features are utilized in our research. Plant structural based feature classification is suitable for plants like corns (Jin & Tang, 2009), as a reason of the obvious tree-shape. However, it is not commonly used, because the features need to be extracted from a relative comprehensive 3D model.

Morphology based features are most commonly used in agricultural computer vision due to its reliability and simplicity. Especially for size features and leave features, as they are planer and easy to be input into computer using 2D or 3D sensors. There are many application extracting morphology features for plants classification. In the work of (Wu, Zhou, & Wang, 2006), the author stated many features can be extracted for plants classification. In our case, 3 features are currently extracted: plant height, plant size and leaf size for corns. More features will be extracted in future work, and classification using these features will be stated in future publications.

Experiment

The experiment is carried out by applying the algorithms stated above to 2 data sets. One of the data set is of corn plants acquired in laboratory, and the other is broccoli plants acquired in filed. The result for ground detection, localization, and feature extraction are evaluated. The ground detection is evaluated by successful rate, the localization and feature extraction is evaluated by average error between calculated one and measured ground truth. The experimental field set-up and results are listed in table 2.

Table 2. Experimental field set-up and evaluation results

	Laboratory	Field
Plants	Corn 	Broccoli 
Plants height (in)	8-12	4-6
Plants size diameter (in)	9-20	6-9
Sample size	10	20
Ground detection rate (%)	100	85
Localization average error distance (in)	1.79	0.48
Height estimation error (in)	-0.24 ± 0.45	-0.17 ± 1.73
Diameter estimation error (in)	-1.22 ± 0.70	-0.72 ± 0.94
Leaf length error (in)	-1.43 ± 0.27	Not feasible

Discussion

The ground detection algorithm work well in laboratory flat ground. But still not stable enough for uneven field ground. It is also unstable with short plants. The localization algorithm is sensitive to plant shape and wind, it works better for short plants. For tall plants like corns, the localization should be performed by using other methods, e.g. searching the stalks. Morphology features extraction accuracy is acceptable. However, the height estimation accuracy is greatly affected by ground detection results, and the diameter and leaf length estimation accuracy is affected by sensor noise level, due to the “flying points” issue of PMD sensors.

Conclusion

In this work we discussed the advantages of 3D sensors in agricultural applications, and evaluated Kinect v2 sensor with some other common 3D sensors. Kinect v2 sensor seems to be a promising and reliable systems for autonomous agricultural robots to sense the environment. It allows the weeding robot to take use both 2D textural information and 3D depth information to realize plants discrimination and localization.

With the fusion of 2D textural and 3D point cloud data, an algorithm has been developed, and the performance is acceptable. With the algorithm, the robot is able to detect ground and detect single plants in crop rows. The location can be determined and some features can be extracted for classification. And operation can be applied precisely to the crops to eliminate weeds.

In the future work, we intend to advance the plant detection using an improved localization and identification method to discriminate several kinds of plants. More kinds of features especially structural models will be extracted for better identification accuracy. And machine learning methods will be applied in classification and also in unexpected conditions like multiple plants are gathered into one cluster.

Acknowledgments

The research is sponsored by Leopold Center for Sustainable Agriculture, Grants No. M2009-23, M2012-24. USDA NIFA Foundational Programs, Grant No. 20136702121126

References

- Alenyà, G., Dellen, B., Foix, S., & Torras, C. (2012). Robotic Leaf Probing Via Segmentation of Range Data Into Surface Patches. *IROS Workshop on Agricultural Robotics: Enabling Safe, Efficient, Affordable Robots for Food Production*, (pp. 1-6). Vilamoura, Portugal.
- ASTRAND, B., & BAERVELDT, A.-J. (2002). An Agricultural Mobile Robot with Vision-Based Perception for Mechanical Weed Control. *Autonomous Robots*, 21-35.
- Drew, M. S., Finlayson, G. D., & Hordley, S. D. (2003). *Recovery of Chromaticity Image Free from Shadows via Illumination Invariance*. Nice, France: Workshop on Color and Photometric Methods in Computer Vision.
- Du, J.-X., Wang, X.-F., & Zhang, G.-J. (2007). Leaf shape based plant species recognition. *Applied Mathematics and Computation*, 883–893.
- J.M. Alvarez, A. L. (2008). Illuminant-Invariant Model-Based Road Segmentation. *IEEE Intelligent Vehicles Symposium*.
- Jin, J., & Tang, L. (2009). Corn Plant Sensing Using Real-Time Stereo Vision. *Journal of Field Robotics*, 591–608.
- Jr., E. R., & Hively, W. D. (2010). Acquisition of NIR-Green-Blue Digital Photographs from Unmanned Aircraft for Crop Monitoring. *Remote Sensing*, 290-305.
- Li, J. (2014). *3D machine vision system for robotic weeding and plant phenotyping*. Iowa State University.
- Li, L. (2014). Time-of-Flight Camera – An Introduction. *Technical White Paper*.
- Nakarmi, A., & Tang, L. (2012). Automatic Inter-plant spacing sensing at early growth stages using a 3D vision sensor. *Computers and Electronics in Agriculture*.
- Rusu, R. B. (2009). *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. Institut für Informatik der Technischen Universität München.
- Sansoni, G., Trebeschi, M., & Docchio, F. (2009). State-of-The-Art and Applications of 3D Imaging Sensors in Industry, Cultural Heritage, Medicine, and Criminal Investigation. *Sensors*, 9, 568-601.
- Surmann, H., & Nüchter, A. (2003). An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, Volume 45, Issues 3–4.
- Weiss, U., & Biber, P. (2011). Plant detection and mapping for agricultural robots using a 3D LIDAR sensor. *Robotics and Autonomous Systems*, 265-273.

Wu, Q., Zhou, C., & Wang, C. (2006). Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network. *Advances in Artificial Intelligence* , 3-10.