Abstract. Three dimensional (3D) reconstruction of the plant or tree canopy is an important step in order to measure canopy geometry, such as, height, width, volume, and leaf cover. In this research, binocular stereo vision was used to recover the 3D information of the canopy. A revised camera calibration method was provided to calibrate the cameras in world coordinate system. Only two images were used to realize a dense reconstruction. These two images were firstly rectified to make sure the corresponding feature points in the left and right images were on the same horizontal line. An efficient large scale stereo matching (ELAS) algorithm was used to find the disparity map. The plant or tree canopy was finally reconstructed based on these calibrated camera matrices and the disparity map through a triangulation method. A plant (croton) with big leaves and a small citrus tree with small leaves were used to test this two-view dense reconstruction. It was easy to measure the geometry of the big leaf. Two big leaves from croton plant were used to measure the width and length of the leaves. The measurement from the reconstruction and manual measurement showed that this reconstruction was metric reconstruction. Another three reconstructions were completed based on a side view of the croton plant, a top view of the croton plant, and a side view of the citrus tree. All these gave good 3D visualization of the objects.

Keywords. Computer Vision, Stereo Vision, 3D Reconstruction, Precision Agriculture, Canopy Reconstruction

Introduction
Canopy width and height, and total frontal area are important measurements for precision agriculture.
Destructive and non-destructive methods were available to do this type of measurement. Usually, destructive method involved manually stripping each leaf and manually taking measurements or using machine. This job can be time consuming and tedious. Ultrasonic sensors, laser, and time-of-flight (ToF) cameras could be used in non-destructive methods. However, these sensors are expensive. For a more cost effective way, two regular webcams can be used as binocular stereo cameras to do 3D reconstruction of the plant. (Scharstein and Szeliski, 2002) summarized and evaluated different dense two-frame stereo correspondence algorithms. Different optimization methods were compared, such as winner-take-all (WTA), Graph Cuts (GC), Dynamic Programming (DP), and cooperative algorithm. All of these methods worked only at pixel level. To obtain sub-pixel level accuracy and make the disparity map look continuous, (Birchfield and Tomasi, 1998) calculated disparity by pixel-to-pixel stereo.

To realize a real-time computation, (Geiger et al., 2010) provided a method called Efficient Large-Scale Stereo (ELAS). This could be realized on a single CPU, with no GPU (Graphics Processing Unit) technology required. (Andersen et al., 2005) investigated the use of stereo vision on 3D analysis of plants and estimation of geometric properties. Simulated annealing (SA) method was used to find the dense matches. Plant height and leaf area of 10 young wheat plants under lab conditions were tested.

(Song et al., 2007) created a surface model of a plant from images taken by stereo cameras. The images of the plants were taken from top view. Three different optimization methods for stereo matching were investigated: 1) Pixel-to-Pixel method (P2P), 2) Graph cut method (GC), and 3) Multi-resolution method (MR). The height of the plant calculated by these three methods was compared with manually measured height. Then the surface of the plant was modeled by using Self-Organizing Map (Kohonen et al., 1996). (Song et al., 2011) provided a non-destructive leaf area measurement by using both stereo and ToF cameras. Two images from stereo cameras and one image from ToF camera were provided. For experiment, one leaf was segmented manually. Then this leaf was reconstructed and its area was measured. For testing, 44 plants were used. Only four leaves in the foreground were used to calculate the leaf area.

The objectives of the research were: 1). Use stereo vision to make a 3D metric reconstruction of the object, 2) Realize that the reconstruction shows the real size of the object.

In this paper, a revised 3d reconstruction method by using stereo vision was introduced. Camera matrices for left and right cameras were calibrated in world coordinate system. Image rectification method based on projective transformation was introduced. And a reliable and equally spread feature matching method was used to detect and match the key points used in image rectification. Then a fast disparity calculation method called ELAS was used to find the dense matches between rectified left and right images. Finally, the scene was reconstructed through triangulation method.

**Methods and Algorithm**

**Camera calibration**

A 2D x-z coordinate system was drawn on a paper whose size was A0 (1189 x 841 mm$^2$). The resulting y axis is perpendicular to this x-z planar and pointing into the paper. Each line was 50mm apart. The z axis (middle direction) was rotated $-10^\circ$ (left direction) and $10^\circ$ (right direction) around center point (0, 0) respectively. Then the checkerboard was placed at different locations. The overview of the image acquisition system is shown in Figure 1. The extended 2D x-z coordinate system is shown in Figure 2, which was extended from Figure 1 in z direction. So the center point (0, 0) in Figure 1 became (0, 225) in Figure 2. Two webcams from Microsoft LiftCam Studio were used in this experiment as stereo cameras. This setup would provide 3D coordinates for all the checkerboard corners in world coordinate system.
Totally, 45 images were taken for both left and right cameras. A method called Gold Standard algorithm (Hartley and Zisserman, 2003) was used to calculate left and right camera matrices from 2D image projections and 3D world coordinates. The parameters are described in (Hartley and Zisserman, 2003). The results were:

\[
P_{\text{left}} = \begin{bmatrix}
609.2397 & 94.8834 & 357.1960 & 1.8263e+04 \\
-63.8514 & 668.2270 & 168.5311 & 1.1899e+05 \\
-0.0731 & 0.1332 & 0.9884 & 24.5011
\end{bmatrix}
\]

\[
P_{\text{right}} = \begin{bmatrix}
613.1276 & 63.5891 & 348.2706 & -6.1077e+03 \\
-19.5970 & 670.3079 & 123.0921 & 1.1876e+05 \\
-0.0672 & 0.1667 & 0.9837 & 30.8089
\end{bmatrix}
\]

**Image rectification**

Projective transformation (Hartley and Zisserman, 2003) method was used to make a rectification of the epipolar line to be parallel with the baseline in both left and right images.
This method didn’t require pre-computed camera matrices. Fundamental matrix (F) for two-view geometry was required. F could be solved through RANSAC algorithm (Hartley and Zisserman, 2003). The feature points used in RANSAC algorithm were from quasi-dense matching algorithm (Lhuillier and Quan, 2005). Quasi-dense matching provided more matches, and the feature points were evenly separately over the whole image. For sparse matching method, like SIFT (Lowe, 2004), which provides less matches, and were concentrated on several places.

This algorithm was summarized as:

1. Compute the epipoles \( \hat{c}_L \) and \( \hat{c}_R \) for two images
2. Compute the projective transformation \( t_R \) which maps the epipole \( \hat{c}_R \) to the point at infinity
3. Find the corresponding projective transformation \( t_L \) which minimizes the least squares distance

\[
\sum_i d(t_L \tilde{x}_{L,i}, t_R \tilde{x}_{R,i})
\]

Then, the projective transformations \( t_L \) and \( t_R \) were applied on both left and right images to resample the images.

**Disparity calculation**

Once images were rectified, searching for correspondence in the right image would be constrained on the horizontal line.

For a point \((x, y)^T\) in the left image, the corresponding matched point in the right image would be \((x - d, y)^T\). The difference \((d)\) between the x coordinate of matched point is called disparity. This was illustrated in Figure 3. For a feature point in the left image, the matched feature point in the right image will be located in the same horizontal line (the blue line). The found feature point was the intersection of the blue and green line in the right image of Figure 3.

![Figure 3. Disparity for rectified images](image)

The Efficient Large-Scale Stereo (ELAS) matching algorithm (Geiger et al., 2010) was used to calculate the disparity for each pixel, which was called dense matching. A value less than 0 meant no matching. This algorithm didn’t require pre-knowledge of the largest disparity. And the resulting disparity was at sub-pixel level. The most important property was its fast computation, which could be used for real-time matching.

After image rectification and disparity calculation, the matches between left and right images were calculated. The matches should then be mapped back to the original images in order to do 3D reconstruction. Inverse projective transformation will be applied both on the left and right rectified matches.

**Image segmentation**

To make a clear reconstruction of the target, the image was segmented to separate the object from the background. An interactive foreground extraction algorithm (GrabCut) using Iterated Graph Cuts (Rother et al., 2004) was provided. The matlab code (Irena and Aviad) was also available.

By using this method, the user only needed to draw a rough region to mark the foreground. Then the
foreground would be separated from the background.

3D reconstruction
2D projections in image and 3D coordinates in world coordinate system are related through camera matrix,

\[ \tilde{x} = PX \]  

(3)

where, \( \tilde{x} \) and \( X \) are in homogenous form.

For left image, we have \( \tilde{x}_L = PLX_L \). And for right image, we have \( \tilde{x}_R = PRX_R \). The left camera center point and the 3D point form a line, the right camera center and the same 3D point form another line. These two lines would intersect at a common point, if there wasn’t noise or error. The 3D point could be solved through triangulation method (Hartley and Zisserman, 2003).

One kind of open source software called Meshlab (which is available at \[ \text{http://meshlab.sourceforge.net/} \]) was used to view the 3D reconstruction result.

Algorithm
The algorithm for stereo dense reconstruction was summarized in Figure 4:

![Algorithm Diagram](image)

Experimental Results and Discussion

Test 1: Laboratory Test on Potted Plant: In this test, a plant called croton was used to build the 3D reconstruction. Images of two separate leaves, a side view, and a top view were taken for reconstruction trials. The original stereo-view images are shown in Figure 5, where left column was taken with the left camera, and the right column was the right camera. The 1st leaf is shown in row 1, 2nd leaf in row 2, the side view in row 3, and the 4th row was the top view.
Once the images were taken, image reconstruction was generated using the steps described above for the reconstruction algorithm. The results of these steps are shown in figure 6, where the rectified left and right images were shown in the first and second column, the extracted foreground of left image was shown in the third column, and the final reconstructions is shown in the fourth column. It can be seen in the fourth column, that the leaves were nicely reconstructed compared with the rectified images in the first and second column.
Test 2: Field Test on Citrus Tree: In this experiment, image scenes were taken of citrus trees in situ. During this stage of the development process, it was determined that the web-cameras had a limited field of view and that it would be best to limit the target scene. The first pair of images was taken from a branch segment of a mature citrus tree. The original left and right camera images are shown in Figure 7. The foreground of the left image and the final reconstructed result were shown in Figure 8. From this result, we can see the leaves were well reconstructed.

Figure 7. Left and right images of a citrus branch

![Figure 7](image1)

Then a stereo pair of images from the side view of a small citrus tree was taken. The original images are shown in Figure 9. These two images were rectified, and the foreground of the left image was extracted (Figure 10A). The 3D reconstructed result was shown in Figure 10B. From this result, we can see that the small citrus tree was well reconstructed.

Figure 8. A) Foreground of left image, and B) reconstructed 3D image

![Figure 8](image2)

Figure 9. The left and right image of a small citrus tree

![Figure 9](image3)
In Meshlab, the distance between two points in an image can be measured using their Measuring Tools. Figure 11 showed the measured length and width for the first leaf of croton. The length estimated by the reconstruction was 164mm, and the width was 89mm. The actual length and width were measured as 166mm and 90mm, respectively. For the second individual leaf, the estimated length and width were 149mm and 65mm, while the hand measured length and width were 156mm and 72mm. We can see that the results from the 3D reconstruction and the hand measurement were very similar. When comparing these results with other reconstruction methods (Lhuillier and Quan, 2005; Snively et al., 2006), this approach was identified as a metric reconstruction.

From these two experiments, we can see that the objects were well reconstructed through the use of only two images. One of the strengths was that a good feature matching algorithm was employed to rectify the images, which is the basis for dense matching. The employed algorithm called quasi-dense matching (Lhuillier and Quan, 2005), provides an evenly distributed set of feature points, comparing favorably with SIFT, a popular and robust sparse feature matching algorithm (Lowe, 2004).

The limitations of this approach are that stereo vision is not sufficient, of itself, to insure a complete 3D reconstruction of the plant. Leaf occlusions and smooth features may cause inaccurate disparity, which limits feature point selection. In the third row (side view of croton) of Figure 6, it can be seen that only partial leaves were reconstructed due to occlusion. In addition, in the last row (top view of croton) of Figure 6, one of the rightmost leaves wasn’t accurately reconstructed. The reason being that, there were an insufficient number of strong features were found during disparity calculation to insure accurate 3D reconstruction.

**Conclusion**

In this research, a 3D reconstruction method based on two-view dense stereo vision was introduced. A camera calibration method based on world coordinate system was utilized. By applying the individual camera matrices to 3D reconstruction, the results could be projected onto and then measured in the same coordinate system. An image rectification method employing projective transformation was used with a quasi-dense matching algorithm to find evenly distributed feature points. A fast disparity calculation method was used to find dense matched features so that the canopy could be reconstructed through a triangulation algorithm.
The final reconstructed scene was a metric reconstruction, which represented the actual dimension of the object. Three dimensional reconstruction based on stereo vision is a kind of passive reconstruction. In some regions of the target image, insufficient features could be found using stereo matching, due to leaf occlusion, highly saturated reflection, as well as other reasons. In future work, images at multiple views could be taken and then merged together to form a full reconstruction. This approach might minimize the negative influence of occlusions.

Acknowledgements
The authors would like to acknowledge the contribution to this research provided by sponsors GeoSpider, Inc and their funding source through the USDA SBIR program.

References