ABSTRACT

Periodic measurements of plant height could be used to calculate growth rate and have potential for characterizing spatial variation of plant response to inputs and field conditions. A corn plant height measurement system using stereo machine vision was developed. A test apparatus was developed, which consisted of an \( x-y \) translating sensor platform with a CCD camera and an imaging stage for camera calibration and plant image acquisition. The position of the camera was controlled precisely by a microcontroller and allowed images to be taken of the top view of the plant from a canonical stereo configuration. Corresponding plant edge points in two images were found using an area-matching algorithm after plant object segmentation. Plant height was estimated by triangulation using pixel disparity between two images. Average plant height was compared to manual measurement of maximum plant height. The stereo vision system estimated maximum plant height with a RMSE of 2.55 cm. Sources of error were investigated. These results demonstrate the potential of stereo vision for a vehicle-mounted crop height measurement system.

**Keywords:** corn plant height, stereo machine vision.

INTRODUCTION

Precision agriculture is a systems approach to managing spatial and temporal variability in crop fields for improved crop performance and environmental quality. Crop yield is affected by many factors such as soil properties, water availability, pest infestations, climatic variation, and topographic features. Because these factors interact with each other, vary spatially and temporally, and stress crop plants at different times during the growing season, it is difficult to
identify the effects of these individual factors on crop growth and yield or to determine how these factors could be managed site-specifically. A major limitation to identifying and mapping yield-limiting factors in fields is the availability of appropriate on-the-go sensing technologies for plant growth (Sadler et al., 2000). Combine yield monitors are one example of sensing technology that has proven successful in measuring crop spatial variability in real-time with high spatial measurement density. However, yield monitors are relatively inaccurate for small areas (Colvin et al., 1999) and can only measure plant response after the growing season is completed. The capability to measure crop growth responses to spatially varying factors at several times during the growing season would provide the opportunity to identify, map, and manage crop stress to minimize its effect on yield. For example, by making sequential measurements of plant height across a field and calculating a growth rate throughout the crop-growing season, it may be possible to identify when and where stress is occurring and the probable causes of that stress. Sammis et al. (1988) used repeated, manual measurements of plant height as an indicator of water stress, evapotranspiration, and yield of irrigated corn, but they did not attempt to determine spatial patterns. Measuring spatial and temporal patterns of crop growth on a field scale has potential as a diagnostic tool for identifying crop stress or crop responses to spatial variability. In general, however, traditional manual crop growth measurements are too destructive or too labor intensive for sequential crop growth measurements to be used on field scale for either commercial or research applications.

Image-based crop growth measurement has been shown to be effective in measuring and modeling crop plant growth in laboratory or greenhouse applications (Morden et al., 1997; Van Henten and Bontsema, 1995; Tarbell and Reid, 1991 a,b; Tarbell et al., 1991). However, vehicle-based sensing systems have not been developed to make repeated, non-destructive, non-contact crop growth measurement on field scales. A stereo vision system should make these measurements possible.

Stereoscopic (stereo) vision has potential for vehicle-based, in-situ measurement of crop growth across crop fields. Stereo vision is a machine vision technique for recovery of three-dimensional information using two imaging viewpoints. These viewpoints may come from two cameras spaced at some fixed baseline distance or from a single camera moved from one viewpoint to another between image acquisitions. Image formation in a camera is a many-to-one mapping from 3-D real world coordinates to 2-D image coordinates and thus depth in a single image is ambiguous. With stereo vision, depth is recovered through triangulation by using imaging geometry and the disparity between common scene points in the two images (Figure 1). Stereo vision algorithms thus typically consist of three steps: a) camera calibration – relating image pixels to a collection of points in the 3-D scene, b) point correspondence – determining the pair of points in the two images which relate to a single point in the scene, and c) 3-D reconstruction – finding the 3-D location of points in the scene (Sonka et al., 1999).

Stereo vision has been applied in to several agricultural applications. Tebourbi et al. (1999) developed a stereo vision system for measurement of soil texture and recommended that stereo vision be applied to crop growth sensing. McDonald et al. (1999) estimated the soil surface roughness and hence light reflection using
stereo vision for different soil structures. They found good agreement between results obtained with the stereo vision approach and those obtained with other standard methods that were extremely time-consuming. Stereo vision has been applied to robotic harvesting applications to estimate the distance between the fruit or vegetable and the robotic manipulator (Takahashi et al., 1998; Kondo et al., 1996). Stereo vision has also been used in light interception studies of several crops (Ivanov et al., 1994; Sinoquet et al., 1998).

Figure 1. Illustration of the stereo vision concept. Depth is recovered from the scene through triangulation by using the disparity, $P_r - P_l$, of the projections of the scene point on the two image planes.

Kanuma et al. (1998) showed that a cabbage growth analysis system which employed stereo vision had improved measurement accuracy of leaf area over a single vision system. Matsuura et al. (2001) developed a transplant population growth analysis system that estimated average height, leaf area, projected leaf area, and mass volume with good correlation to destructive measurements. Lines et al. (2001) developed a stereo vision system which estimated the mass of free-swimming fish with a mean measurement error of 18%.

Will (2001) demonstrated that stereo imaging with vehicle-mounted sensors could estimate the 3-D world coordinates of control points with decimeter-level accuracy with a small number of images. The control points in Will's study were unambiguous and were determined manually in the images. However, Will demonstrated the potential of vision sensors and stereo-vision techniques on agricultural vehicles for estimating the location of physical points in field scenes. The objectives of the study described in this paper were to 1) investigate the extraction and correspondence of control points on stereo scenes containing corn plants and 2) determine the accuracy and error associated with such techniques in estimating corn height.
MATERIALS AND METHODS

Equipment

A lab-based sensor platform was developed which consisted of a motion control system on which a Pulnix TMC-9700 (Sunnyvale, CA) camera was mounted 2.1 m (6.9 ft) above and perpendicular to an imaging stage. A 6mm Tamron (Wayne, NJ) auto-iris (F1.6 to F360), varifocal (8-16mm) lens was used to focus on a 1.14 m (45 in.) by 0.86 m (33.75 in.) field of view. RGB component video from the camera was routed to a FlashBus MV color frame grabber board (Integral Technologies, Indianapolis, IN), which was installed in an Optiplex GX300 portable computer (Dell Computer Corp., Round Rock, TX) with dual 866 MHz Pentium III CPUs. The frame grabber had a resolution of 640 by 480 pixels and converted the analog video signal to 24-bit digital color images. Matlab version 6.1 (The Mathworks, Natick, MA) was used for image processing and plant height estimation.

The camera was translated in the horizontal plane through a lead screw in the y direction and with a traction roller drive in the x direction (Figure 2). Camera motion was controlled with a PK 2110 (Z-World, Davis, CA) single-board computer with a 6.144 MHz microcontroller.

Figure 2. Experimental apparatus for camera calibration and depth measurements. The controller could position the camera with 1.59 mm precision in both x and y directions.
Stereo Vision Plant Height Estimation Algorithm

Plant height was defined as the height from the imaging stage to the highest point of a plant. The stereo vision algorithm for plant height estimation consisted of the following three steps.

Camera Calibration

Camera calibration is necessary to estimate the camera parameters. When a camera is calibrated, the intrinsic and extrinsic parameters of a camera are determined and used in succeeding calculations of height determination. A point W in the scene is imaged by two cameras facing downwards (Figure 3).

Figure 3. Side view of two downward facing cameras viewing a target point W. U_{c1} is the projection of point W in camera 1, and U_{c2} is corresponding projection in camera 2. C_1 and C_2 are focal points of the camera. The two cameras are separated by baseline distance, d.

With C_1 as the origin of the camera coordinate system, the coordinates of scene point W are r_{c1} = [x_{c1}, y_{c1}, z_{c1}]^T. Then the point W will have a perspective projection onto the left image plane:

\[
\begin{pmatrix}
U_{c1} \\
V_{c1}
\end{pmatrix} = K \begin{pmatrix}
x_{c1} \\
y_{c1} \\
z_{c1}
\end{pmatrix} /
\begin{pmatrix}
z_{c1}
\end{pmatrix}
\]  

(1)

where U_{c1} is the coordinate of projected point with respect to the camera coordinate system. This physical projection is related to a particular set of pixel coordinates on the CCD sensor. There is scaling and possible shearing of the image in the transformation from physical to pixel coordinates. Furthermore, since the image pixels are measured from top left corner, there will be translation of coordinates. Therefore, for the image coordinates, let U_1, V_1 be the pixel location on the image corresponding to the point U_{c1} on CCD. Shear and scaling in the camera can be modeled as a K matrix defined by:
The matrix $K$ consists of the intrinsic parameters of the camera, where parameters $a$ and $c$ are scaling factors in pixel/mm and parameter $b$ is the shearing component. Parameter $f$ is the focal length of the camera (Sonka, 1999). To implement a stereo imaging system, the camera being used must first be calibrated to correct for lens distortions and to relate pixel coordinates to corresponding unique rays in image space. The intrinsic camera parameters were determined using camera calibration software (Bouquet, 2002) based on the work of Zhang (1999), Heikkilä and Silven (1997), and Tsai (1987). Twenty images at different viewing angles of a 9 x 18 square checkerboard with 30 mm square grid size were used for the calibration. Radial and tangential distortion coefficients of the image were also calculated.

**Point Correspondence**

Several image-processing steps were used to determine point correspondence of plants. First, image segmentation between plant and background was performed using the truncated ellipsoidal surface method developed by Shrestha and Steward (2001). This method accomplished segmentation by using an ellipsoidal surface in RGB color space as a discrimination boundary between vegetation and non-vegetation regions. Fixed surface parameter values were used with this method because the lighting source was fixed and controlled. Segmentation noise was removed through morphological opening with a 4-by-4 square structuring element. Objects segmented as plants and larger than 80 pixels were labeled as plant objects. Next, the chain code for each valid plant object boundary was determined. Points on plant object edges were selected as the candidate point for correspondence because of high intensity variance associated with the region around those points. Thus, the object chain codes were parsed in 20 code long chunks. The mean row location and minimum and maximum columns for each chain code chunk were used to define the image area that was matched with the same area in the other image to establish correspondence. Normalized cross-correlation of excess green images was used to find correspondence. The result of this search was a collection of image coordinates and associated pixel disparity between common points on plants in two images.

**Height estimation**

Once the intrinsic parameters of the camera were known, depth information was calculated using disparity in image coordinates taken from the camera located
at two different viewpoints as shown in fig. 3. It was assumed that the two camera viewpoints were on the same horizontal plane and separated only in the y direction by baseline distance, d. It was also assumed that the image coordinates for camera 1 and 2 were parallel, that is, there was no rotation between camera positions 1 and 2. With these assumptions, the perspective projection of point W on camera 2 was given by:

\[ U_{c2} = K \begin{pmatrix} \xi_{c1} \\ \eta_{c1} \\ 1 \end{pmatrix} \begin{pmatrix} x_{c2} \\ y_{c2} \\ z_{c2} \end{pmatrix} \begin{pmatrix} d \\ 1 \end{pmatrix} \]

(3)

Since the depth was equal for both cameras and \( U_{c1}, U_{c2}, K \) and d were all known, equation 1 and 3 could be solved for unknown \( \xi \). This was accomplished by subtracting Equation (2) from (3), and substituting \( \xi = H - h \), where \( H \) is the vertical distance from camera to a reference plane and \( h \) is the height of an object being measured from the reference plane (Fig. 3). After simplification, the height of a plant was calculated using:

\[ h = H - \frac{f_c}{P} \frac{d}{P} \]

(4)

where \( P \) is the disparity in pixels of a control point in the two images. The highest heights for each object were used. If a plant consisted of more than one object then the highest object was assumed to be the highest point in the plant. If more than one image pair had a height estimate for the same plant, then the average height was taken and compared with measured height of the plant.

**Experimental Design**

Sixteen corn plants at V3-V4 growth stage were placed on the imaging platform, and the camera was moved in the y direction at 5.1 cm increments with an image acquired at each point. Ten images were acquired. Then the camera was moved 37.5 cm in the x direction and another set of 10 images was acquired at 5.1 cm intervals. The FOV was 114 cm in the x direction resulting in an overlap of 75 cm from one camera location to the next. The maximum height of each plant was measured manually to compare the results with calculated height. Plant height ranged from 45 cm to 77 cm including the pot height.

**RESULTS AND DISCUSSION**

Camera calibration revealed that there was both radial and tangential distortion in the images. The K matrix in equation (3) was determined. The value of \( f_c \) in equation 4 was found to 1163.86 \( \pm \) 5 pixels.

The image had pin cushion distortion. A rectified image is shown in figure 4. The barrel-shaped image outline is due to distortion correction. The height was calculated with the maximum disparity detected for an object. When the height was calculated using many pairs of images for the same plant, it was found that calculated height from some pairs of image was very different than others. In those cases, the values were considered to be outliers due to noise and were not included in the calculation. For all other values, the average height was taken as an estimate of plant height from different pairs of images. When there were more
than 1 object detected in segmentation and later it was determined those objects belonged to the same plant, then the maximum height of all the objects belonging to that plant was used as the height for that plant. In figure 4, 3 objects were detected for 2 plants. We then took plant height as 52.7 cm for plant on top row and 65.8 cm for plant in the middle of the bottom row. The plot of actual height and estimated height is shown in figure 5.

Figure 4. Undistorted image showing plant detection chain coding, disparity measurement and height detection.

The $R^2$ value of the linear regression line was 0.84. The slope was 0.88 and the y-intercept was 7.46. There was no evidence that the slope and the y-intercept were significantly different than 1 and 0, respectively. The RMSE of the model was found to be 2.55 cm. The mean actual plant height was 59 cm. Therefore, RMSE was 4.3% of the mean value.

For larger plants (V4 –V5), there was poor correlation when we estimated the height of plants that are touching each other. When one plant was underneath another plant, they were detected as a single object in segmentation and only the higher plant height was measured. Therefore, the plant height calculated for smaller plants were biased. Interplant leaf overlap was the main reason for poor correlation.
During stereo matching, the disparity measurement $\delta P$ may vary slightly because of several error sources, such as image distortion and camera calibration error. Height measurement resolution under the assumption of one pixel matching resolution can be calculated an equation derived by differentiating equation 4 and after manipulating:

$$\frac{dh}{\delta P} \frac{(H - h)^2}{(fc)d}$$

Equation 5 shows that the depth resolution is directly proportional to the square of the distance of an object from the camera and inversely proportional to the camera separation and resolution. For this reason, alternate images were used to calculate the height of the plant from images taken at 5.1 cm apart. For height measurements of 45 to 78 cm in this experiment, the depth resolution due to 1 pixel matching resolution is shown in figure 6.
Figure 6. Uncertainty in height estimation from pixel disparity. The line in the middle is the estimated height. Dots above and below the middle line shows the confidence interval of estimated height due to scaling uncertainty. The actual height may be anywhere between outer lines due to height measurement resolution.

As pixel disparity increases, the depth resolution decreases and the confidence interval of depth measurement decreases as we have larger disparity. If we have used the 5.1 cm baseline difference, then this interval would have doubled as well. As we try to measure higher objects the slope of the curve in fig. 6 decreases. This indicates that there is a limitation on maximum height we can measure from this technique. For an image size of 480 × 640, the maximum theoretical discrepancy in the y direction is 639 pixels. However in practice, there should be a considerable amount of overlap in two images for robustness in correspondence detection.

CONCLUSIONS

A stereo vision-based plant height estimation system was developed. A single camera was used on an imaging platform. The movement of the camera was precisely controlled using a stepper motor controlled by a digital controller. Early stage corn images were acquired with the camera with known camera movements.

The area-matching algorithm worked well to find plant point correspondence. Estimated plant heights were well correlated for smaller plants, but were poorly correlated in the case of large plants. This poor performance was due to
overlapped leaves. Improvement in the 3-d reconstruction and height estimation algorithms should result in future improvements in height estimates. Pixel matching resolution is a major source of uncertainty in the height measurements. Subpixel interpolation for pixel matching will be investigated for future improvements in measurement accuracy. The preliminary results presented here reveal the potential that stereo vision has for crop height measurements.

REFERENCES


