
Dev Sagar Shrestha  
Iowa State University

Brian L. Steward  
Iowa State University, bsteward@iastate.edu

Follow this and additional works at: http://lib.dr.iastate.edu/abe_eng_conf

Part of the Bioresource and Agricultural Engineering Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/abe_eng_conf/37. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

Abstract
From yield monitoring data, it is well known that yield variability exists within a field. Plant population variation is a major cause of this yield variability. Automated corn plant population measurement has potential for assessing in-field variation of plant emergence and also for assessing planter performance. Machine vision algorithms for automated corn plant counting were developed to analyze digital video streams. Video streams were captured along 6.1 m long cornrow sections at early stages of plant growth and various natural daylight conditions. A sequential image correspondence algorithm was used to determine overlapped image portions. Plants were segmented from the background using an ellipsoidal decision surface, and spatial analysis was used to identify individual crop plants. Performance of this automated method was evaluated by comparing its results with manual stand counts. Sixty experimental units were evaluated for counting results with corn population varying from 14 to 48 plants per 6.1 cornrow length. The results showed that in low weed field conditions, the system plants counts well correlated to manual counts (R² = 0.90). Standard error of population estimate was 1.8 plants over 34.3 manual plant count that corresponds to 5.4% of average error.

Keywords
Machine vision, image sequencing, segmentation, plant count

Disciplines
Bioresource and Agricultural Engineering

Comments
ASAE Paper No. 011067

This conference proceeding is available at Iowa State University Digital Repository: http://lib.dr.iastate.edu/abe_eng_conf/37

Dev Sagar Shrestha
Iowa State University, Agricultural and Biosystems Engineering Department,
139-Davidson, Ames, IA 50010 USA, dev@iastate.edu

Brian L. Steward
Iowa State University, Agricultural and Biosystems Engineering Department,
206-Davidson, Ames, IA 50010 USA, bsteward@iastate.edu

Written for presentation at the
2001 ASAE Annual International Meeting
Sponsored by ASAE
Sacramento Convention Center
Sacramento, California, USA
July 30-August 1, 2001

Abstract. From yield monitoring data, it is well known that yield variability exists within a field. Plant population variation is a major cause of this yield variability. Automated corn plant population measurement has potential for assessing in-field variation of plant emergence and also for assessing planter performance. Machine vision algorithms for automated corn plant counting were developed to analyze digital video streams. Video streams were captured along 6.1 m long cornrow sections at early stages of plant growth and various natural daylight conditions. A sequential image correspondence algorithm was used to determine overlapped image portions. Plants were segmented from the background using an ellipsoidal decision surface, and spatial analysis was used to identify individual crop plants. Performance of this automated method was evaluated by comparing its results with manual stand counts. Sixty experimental units were evaluated for counting results with corn population varying from 14 to 48 plants per 6.1 cornrow length. The results showed that in low weed field conditions, the system plants counts well correlated to manual counts ($R^2 = 0.90$). Standard error of population estimate was 1.8 plants over 34.3 manual plant count that corresponds to 5.4% of average error.

Keywords. Machine vision, image sequencing, segmentation, plant count
Introduction

Many factors play a role in plant germination rate such as seed quality, soil tilth, water availability, light, temperature and diseases. Two of the overall indicators of this variability is plant emergence and growth rate. Assessing the adequacy of individual inputs, environment, and their interaction is a complex task to measure. For this reason a plant counting system could be used as a measure of variability in a field for these inputs, environmental conditions and their interactions.

Plant populations that are higher and lower than the optimal plant population can reduce the yield. Duncan (1958) found that the grain yield was maximum at a certain plant population depending upon variety and other characteristics. Wiley and Heath (1970) investigated the relationship established by different researchers between plant population density and crop yield and found the predictions had similar trends of yield maximization at a certain point.

If plant population density could be estimated automatically at its early stage of growth, early yield potential estimates could be made based on the number of plant available in the field to produce grain. Sensors have been developed to measure corn population at harvest (Sudduth et al., 2000; Nichols, 2000). Comparison of early plant population and harvesting time population may be used to estimate the population density required at planting time to achieve desired population density at harvesting time. Automated plant population estimation could also enable the evaluation of other factors like seed germination rate and land fertility variation within a field. In addition, if plant population data of subsequent years could be recorded along with other factors like moisture content, amount of precipitation, temperature, the effect of different variables determining the final yield of crop could be measured. Manual counting and recording of population density is a tedious task and subject to error. In addition, it would not be feasible to manually count a large field area.

In this research, a digital video camcorder was used to take the pictures of cornrow. Video streams were captured in the computer memory as AVI file format with digital video (DV) compression through an IEEE 1394 interface. The individual frames of this AVI file were decompressed and stored sequentially as separate images. These individual files were then processed to determine relative overlapping from frame to frame. Sanchiz et al. (1995) developed a system to sequence the image frames with assumption that there is no movement in the scene itself. However, in our case the plant leaves were fluttering rapidly with wind and inclusion of plant region for overlapping frame gave erroneous result. The overlapped images were analyzed to find the plant count, inter-plant spacing, and plant centers. Jia et al. (1990) studied the feasibility of detecting main veins along the leaves and found the intersecting point to estimate the corn plant center. However, at early growth stage of corn plant, there are no consistent distinct veins. Therefore, it was not possible to use main veins for plant center detection.

The objective of this research was to develop a methodology for counting corn population at early growth stage from a digital video stream taken from an ordinary digital video camera.

Methodology

Experimental Setting

Video sequences were collected in corn plots at the Iowa State University Agronomy and Agricultural Engineering Research Center during the third and fourth weeks of May 2001. A Sony DCR-TRV900 digital camcorder was mounted on a vehicle at 0.6 m above the ground with
a 30cm by 40 cm field of view. The vehicle was driven over a cornrow with the camera directly
over the plants at the speed of about 1 m/s. The shutter speed was adjusted to 1/1000 second;
frames were captured in progressive scan mode, and other camera settings were set to auto. A
circular polarizer filter was used to reduce glare. The corn plants were at V3 stage (McWilliams
et al., 1999). Cornrows were marked off at 6.1 m lengths by staking yellow construction tape
perpendicular to the cornrow. The plants within each section were counted manually.

The video stream was stored in a mini DV tape in the field. In the laboratory, video streams
were imported from the camera to a personal computer using an IEEE 1394 serial interface.
Adobe Premiere® 6 software was used to capture the video stream as AVI files and then to
decompress and store individual frames as color tagged image file format (TIFF) files.

**Image Sequencing**

Intensity images were derived from color images, and the amount of shift between sequential
frames was calculated. A 30-pixel by 30-pixel image patch was selected in frame (n) randomly
with the constraint that the patch did not fall outside the boundary of the subsequent frame.
Once a patch and corresponding search region was selected, it was segmented to see if it
contained any plant segments, very dark, or very bright areas. If either patch or search region
enclosed these regions, the patch was reselected. Plant regions were excluded because the
position of a plant may change from frame to frame due to wind introducing noise into the
process.

Shaded portion of frame (n+1) in Figure 1 shows the search region for patch X. In order for this
region to be completely within frame (n+1), the patch X has to be selected within the shaded
area of frame (n). The search region was set such that the patch could be moved around by 30
pixels in any direction from the center of the search region. Average amount of shifts of two
previous images were used to find the center of the search region. In order to determine the
amount of shift between the first and the second frame in the sequence, it was assumed that the
vehicle always traveled forward. Therefore, the patch was selected within lower 100 pixels width
of the first frame with 50 pixels margin from both right and left sides. Then the patch was shifted
over the entire second frame starting from the upper left corner to the bottom right corner.

![Image Sequencing Diagram](image.png)

Figure 1. For image sequencing, an image patch X in frame (n) was matched with patch X in
frame (n+1). Difference in coordinates of the patch matched to the second frame gives the
amount of shift.
Assuming patch is of size $m \times n$ and search region of size $M \times N$, the matching error for each position was determined by:

$$Err(p, q) = \sum_{i=1}^{n} \sum_{j=1}^{m} |patch(i, j) - SearchRegion(i + p - 1, j + q - 1)|$$  \[1\]

where $p$ and $q$ in $Err$ matrix varied from 1 to $(M-m)$ and 1 to $(N-n)$ respectively. The process of determining the error matrix is shown in figure 2.

The amount of shifting was determined by the significantly minimum error difference and was used to guide the succeeding searches.

To determine the validity of a match, the calculated error values were sorted in ascending order and the difference between successive values were calculated. For a valid match, the difference between the lowest error and the next lowest error value required to be less than 5 standard deviations ($\sigma$) higher than the mean of the remaining error differences. For example, the error matrix for Figure 2 was calculated as:

$$Err = \begin{bmatrix}
0.3 & 4.3 & 4.3 \\
2.8 & 0.0 & 3.0 \\
4.1 & 2.6 & 3.8
\end{bmatrix}$$  \[2\]

The matrix $Err$ were arranged in a row in ascending order and the difference was calculated as:

$$\Delta Err = \begin{bmatrix}
2.6 & 0.2 & 0.2 & 0.3 & 0.5 & 0.3 & 0.2 & 0.0
\end{bmatrix}$$  \[3\]

Since first value of $\Delta Err$ i.e. 2.6 is more than 5 $\sigma$ from mean of the rest of the differences, the minimum error 0.0 in the $Err$ matrix was considered to be a true minimum and the match was accepted. If a valid match, based on 5 $\sigma$ criteria could not be found in the specific region, then another random patch was chosen and searched for a match. From Chebyshev’s theorem, the probability that the match found in this way is not by coincidence is:
Thus the $5 \sigma$ criteria would result in at least 96% confidence in getting a correct match from frame to frame. A three dimensional view of an error surface of a typical match is shown in Figure 3.

![Error surface and its contour for a typical image matching.](image)

Figure 3. Error surface and its contour for a typical image matching. For the valid match, the difference between the minimum error and next to minimum was required to be greater than 5 standard deviations of the differences of the sorted error values.

**Image Segmentation**

Image segmentation between plant and background was performed using a truncated ellipsoidal surface. This method accomplished segmentation by using an ellipsoidal surface in RGB color space as a discrimination boundary between vegetation and non-vegetation regions. This surface was originally developed by selecting regions for a constant B value planes to separate regions perceived as green from those perceived as non-green. After stacking these decision surfaces for each B values varying from 0 to 1, it was determined that the discriminating surface could be functionally represented by a truncated ellipsoidal surface given by:

$$
\frac{R^2}{D^2} + \frac{(1 - G)^2}{(E \times B + F)^2} = 1
$$

where $R$, $G$, and $B$ values were the red, green and blue values of a particular pixel and $D$, $E$, and $F$ were the parameters describing the shape of the ellipsoid.

For a given set of parameters, the left-hand side of Eq. (5) was used to classify pixels as vegetation if $\leq 1$ or background if $>1$. Each parameter chosen above has a physical meaning. $D$ is the maximum $R$ value still conceived green when $B$ and $(1-G)$ are 0. $E$ is the slope of the ellipsoid boundary in $G$-$B$ plane at $R = 0$. $F$ is the maximum value of $(1-G)$ which, is conceived green at both $B$ and $R$ are 0. The parameters value were determined by trial and error method for an image and same values were used for all other images. The decision surface used in this paper is shown in Figure 4.
Plant Counting

After the plants were segmented, the binary segmented images were scanned through every row and two features were extracted: 1) total number of plant pixels in each image row and 2) the median position of the plant pixels along that row. Once all the image rows were scanned and extracted features were recorded, a row was either classified as a plant row or a background row. An image row was considered to be a plant row if:

1. The variation in median position from one row to next row was less than the total number of plant pixels in that row and,
2. The plant pixel count was greater than the mean value of all plant pixel counts.

Once the entire sequence of images had been classified, the adjacent plant rows and background rows were considered to be a single plant or background region. Plant centers were assumed to be in the middle row of each plant region with an offset from left side of the image equal to the median position of plant pixels in that row. This classification resulted in a crude estimate of the number of plants. Next plant and background regions were further refined using the following criteria.

1. Plant regions that were less than 20% of the average plant region lengths were considered to be a false plant region and reclassified as background.
2. Background regions that were less than 20% of the average background region length were considered to be a false background region and were reclassified as plant region.
3. If any plant center was found outside five standard deviations from average offset of plant center from the frame border then that plant was considered to be a weed.

After this refinement, the plants were counted again and if the numbers varied by more than 5 percent of the original count, the average length of plant and background were updated and above algorithm from step 1 to 3 were repeated until the plant counted before and after refinement varied less than 5 percent.

Finally, the plant regions that were more than twice the length of an average plant region were counted as doubles and more than three times as triples and so on. The plant centroid locations were assumed to be at the middle of the plant region. In case of multiple plants the centroids were assumed to split evenly along the plant region.
Results and Discussion

The segmentation quality using truncated ellipsoidal method was satisfactory for this application. The qualitative performance of segmentation result is shown in Figure 5 for three different lighting conditions. The segmentation method was robust in changing lighting conditions; however, the segmentation quality deteriorated with extreme light changes for instance if the parameters values were optimized for very dark images and if same set of parameters were used to segment the image in very bright lighting condition and vice versa, the segmentation was poor. A neural network was found effective in adjusting the values of parameters in Eq. 5 to adapt to changing lighting conditions. However, these parameters were fixed at $D = 0.9$, $E = -0.57$ and $F = 0.81$ in this entire experiment.

![Figure 5](image_url)

Figure 5. Three different lighting intensity images and their corresponding segmented image at the bottom using truncated ellipsoid method. Picture in left represents low intensity image, middle image has shaded portion and right image has medium lighting intensity.

The result of overlapped images showed that the non-moving objects like stalks and soil cracks aligned better than moving objects like plants. During image matching, it was observed that even though some pixels were perfectly positioned relative to preceding frame, other pixels were skewed by a pixel or two. Factors responsible for this error may be vibration, unequal distance between a fixed ground point to camera lens in a subsequent frame and imperfect camera alignment. Placing the camera sufficiently high above the ground may reduce this error. However, these effects were not considered in this study as the vibration was not measured and this error had no observable impact on final result.

Although the patch size of $30 \times 30$ pixels, and search region of $90 \times 90$ pixels were chosen arbitrarily, they were adequate for finding a match. Based on a spatial resolution of 16 pixels/cm, a 30-pixel margin of error in the search region allowed for an instantaneous acceleration or deceleration of $112.5 \text{ cm/s}^2$ during video acquisition.

For image sequencing, it was observed that on average a significantly minimum value in $\Delta Err$ was not obtained once in every six matching process which forced patch reselection. Similarly, on average randomly selected patches were rejected once in every 4.4 times. Even when selected patch was valid, 20% of the time the search region was invalid hence forcing the algorithm to reselect of the patch. This indicates that an image sequencing errors could be greatly minimized with only a little extra effort of reselecting a patch.
Prior to using $\Delta Err$ for determining the best match, the minimum value in $Err$ matrix was used and compared with the mean of the other error values. The minimum value was considered significant when the value was more than $5 \sigma$ below the mean. This procedure picked the lowest value even when there were values near to the absolute minimum and gave rise to frequent false matches.

Figure 6. Overlapped image of 25 subsequent frames (above). Segmented and plant region detected image (below). Grey lines are detected plant regions. Circles indicate the estimated plant centers.

The result of 25 subsequent overlapped frames is shown in figure 6. This figure provides qualitative evaluation of typical image sequencing. Although non-moving objects were aligned, we can see that some leaves were misaligned (notice choppy segmented leaves on 4th, 5th and 8th plants from left in Figure 6).

The plot of automatic count versus manual count is shown in Figure 7. The manual plant count varied from minimum of 14 to maximum of 48 plants with mean value of 33.2. Root mean square error of the actual count and automated count was 1.83. The slope and intercept of the regression line was found to be 0.93 and 1.98 respectively. However these values were not significantly different from 1 and 0 for the slope and the intercept. The coefficient of determination ($R^2$) value of linear regression was 0.90.

The main source of error was great variability in plant size and leaf orientation. This made the assumption to refine the plant and background region to be at least more than 20% of the corresponding average more susceptible. More weed and noise pixels were counted as plant when this threshold was lower and vice versa. More accurate counting was possible if this threshold was individually adjusted for each experimental unit however, it would make the whole approach more cumbersome and impractical.
Conclusion

From this experiment it was concluded that patch-matching algorithm with certain restrictions could be used to sequence the digital images captured from video camera. The truncated ellipsoid was successfully used to segment plant from background in varying lighting conditions. The sequenced images were analyzed and numbers of plants were counted. Also an algorithm was developed to locate the plant centers. In the low weed areas, the method could count the early stage corn plants with 5.4% of average counting error.

This algorithm for early stage plant counting could be used for low weed condition. More research is needed to automatically adjust the thresholding values in region refinement for better result. Implementing a neural network method to adjust segmentation parameters may improve the result. In the future Global Positioning System (GPS) signals will be recorded with the video for future geo-referencing the population counts.

References


[http://www.ext.nodak.edu/extpubs/plantsci/rowcrops/a1173/a1173w.htm](http://www.ext.nodak.edu/extpubs/plantsci/rowcrops/a1173/a1173w.htm) Accessed 23 July 2001

