Autonomous Guidance of a Corn Harvester using Stereo Vision

F. Rovira-Más[1], S. Han[2], J. Wei[3], and J. F. Reid[4]

[1]Polytechnic University of Valencia. 46022 Valencia, Spain.

[2]John Deere Agricultural Management Solutions. Urbandale, IA 50322, USA.

[3]Kansas State University. Manhattan, KS 66502, USA.

[4]John Deere Technology Center. Moline, IL 61265, USA.

E-mail: frovira@dmta.upv.es

ABSTRACT

The growing demand for applications in Precision Agriculture has been proportional to the emergent interest in automatic guidance of agricultural machines. The difficult, and at the same time strenuous, task of driving a harvester during the prolonged working days of the harvesting season justifies the special attention paid to this technology. The development of Global Navigation Satellite Systems has meant an important push for autonomous navigation, especially in the case of off-road vehicles. However, experience has shown that automatic guidance inside the tight arrangement of crop rows is difficult to achieve unless local positioning sensors take part in the localization unit of auto-steered machines. This research develops a perception engine based on stereoscopic vision to detect the cut-uncut edge of corn during harvesting operations. The stereo camera was mounted on the head of an agricultural combine. The algorithm implemented was capable of finding the edge of corn, allowing the system to automatically guide the combine at regular harvesting speeds.

Keywords: Autonomous navigation, corn harvester, edge detection, guidance, Precision Agriculture, stereoscopic vision, visual perception

1. INTRODUCTION

Among typical agricultural operations, harvesting is one of the most delicate because it usually needs to be carried out in a narrow period of time when ambient conditions are favorable. Total benefits obtained from the crop depend on the yield harvested, which is sensitive to the quality of the harvesting operation. If maize, for example, is the crop to be reaped, a combination of circumstances has to be met: adequate moisture content in the corncob; good weather conditions with absence of rain or snow; dry soil to allow the circulation of the machines; accessible combine and loading truck; and operators' availability. When all these constraints are taken into consideration, it is not rare to request combining from dawn to dusk, even during the night, for several days in a row. Under these demanding conditions, the operator is forced to work under stress and fatigue that can result in poor performance or what is much worse, in a serious accident. If the ultimate goal of technology is to improve life, harvesting operations bring a breeding ground for its application. A natural and logical use of technology in this case would be through assisting in the vehicle's navigation. There is no need to attain a complete automatic guidance solution in order to be considered a satisfactory outcome; a semi-autonomous result would be helpful in managing the combine since it would relieve the operator from some hours of strenuous work. Furthermore, it would be beneficial for acquiring experience in guidance

systems, which, eventually, will lead to fully autonomous prototypes. This research investigates in such direction, trying to bring some light to the challenge of autonomous navigation for agricultural combines.

Since the incorporation of sensors, electronics, and computers to agriculture at the end of the eighties, and more generally after the nineties, many automatic guidance architectures for off-road vehicles have been proposed, both in industry and academia. Machine vision solutions seem to be the most appropriate for the situations faced during harvesting, however, they are not exempt from complexities and challenges. One of such intricacies deals with the camera location, for which there is no definite solution. Benson et al. (2001) studied four camera locations for combine guidance, and arrived to the conclusion that the best position for the camera is above the cabin. One of the positions discarded in that work placed the camera attached to the combine head, pointing at the cut-uncut edge. The main cause for rejecting the low camera position on the combine's head was due to sparse crop and shadows. In spite of the flawed results experienced with such location in previous research projects, the solution proposed in this article places the camera on the head of the combine, but adds two significant changes: first, the camera was mounted on a sliding bar on the combine's head in such a way that a better view of the cut-uncut edge was acquired by rotating the camera towards the crop; and second, stereovision was employed as the main perception method. The advantages of stereo over monocular vision can, to some extent, account for the shadows issue: a change in ambient illumination will affect both stereo sensors simultaneously. Therefore, the texture patterns that allow stereo matching will remain unaltered, as long as there is enough light for the camera to detect distinct features in both images of the stereo pair. This property of stereo vision, together with the fact that stereo provides the range or third coordinate, has motivated the advent of multiple applications for robot perception and autonomous navigation in the last five years. Murray and Jennings (1997) demonstrated that stereo is an alternative to laser and sonar, and built a lab robot that used stereo for localization and mapping of an unknown environment. The scenes found during harvesting present certain structure as corn is planted following regular rows separated evenly a known distance. There is, therefore, an *a priori* knowledge about the scene to be acquired by the stereo camera. If landmarks can be placed in the field of view of the camera, the location of a vehicle can be determined by means of stereo (Wang et al., 2005); and if a solid model of the target object is available, a robotic manipulator will have at its disposal a modeled environment for automatic tasks (Lee et al., 2005). While agricultural applications are still scarce (Rovira-Más and Reid, 2004; Kise et al., 2005; Rovira-Más et al., 2005; Wei et al., 2005), the possibilities of perception via stereo are expanding in all directions; from 3D mapping for underwater autonomous vehicles challenged by limited visibility caused by turbid waters (Khamene and Negahdaripour, 1999) to humanoid robots that walk around unknown home environments (Sabe et al., 2004).

The object of this study is to develop a perception engine based on stereoscopic vision to detect the cut-uncut edge of corn during harvesting operations. Several hardware configurations and different edge detection algorithms were evaluated in the course of the experiments.

2. SYSTEM ARCHITECTURE AND TEST PLATFORMS

The perception sensor was a binocular stereo camera (BumbleBee, Point Grey Research Inc., Vancouver, Canada) with a 120 mm baseline and 6 mm focal length lenses yielding a horizontal

field of view of 42 º. The camera communicated with the on-board processing computer (Motion Computing Inc, Austin, TX, USA) via an IEEE 1394 connection. The edge detection algorithm was implemented in the processing computer and the resulting output, namely the vehicle's offset, was sent to the control unit of the combine. The corn harvester selected to perform navigation tests was a John Deere 9660 STS (Deere and Company, Moline, IL, USA) equipped with AutoTrac^{TM}, an auto-steering system that provides positioning corrections using a GPS receiver as the principal localization sensor. The offsets (lateral misalignments) calculated by the AutoTrac algorithm were substituted by the offsets sent by the stereo processor, being the steering controller and mechanism those of the original AutoTrac.

One of the key parameters in the system was the camera location. Several options were tested before attempting to guide the combine in the field. The monocular camera approach reported by Benson et al. (2001) found a low position on the head to be problematic, but when a stereo camera was employed as the main perception sensor, the possibility of rendering a 3D view of the scene resulted in several advantages for that placement. Nevertheless, different alternatives were studied within the head before deciding a permanent mount. Figure 1 illustrates the general layout of the system: (a) side view; (b) top view. The first experiments were conducted in Weslaco, Texas (USA), on 14 July 2005. The objective of that preparatory phase was to scrutinize the potential of stereo for combine guidance. Stereo images were taken with the following parameters, as defined in figure 1: h_c = 1.61 m, θ = 0 º, L_r = 0.965 m, L_W = 5.842 m, and L_c = 0. Difficulties encountered detecting the edge, as explained in the following sections, led to a reformulation of some parameters, being the camera position for the automatic guidance tests, carried out in Des Moines, Iowa (USA) in October 2005, given by h_c = 1.66 m, θ = 30 º, L_c = 1.42 m. The camera was always mounted flat except for one test where it was tilted down 5°.

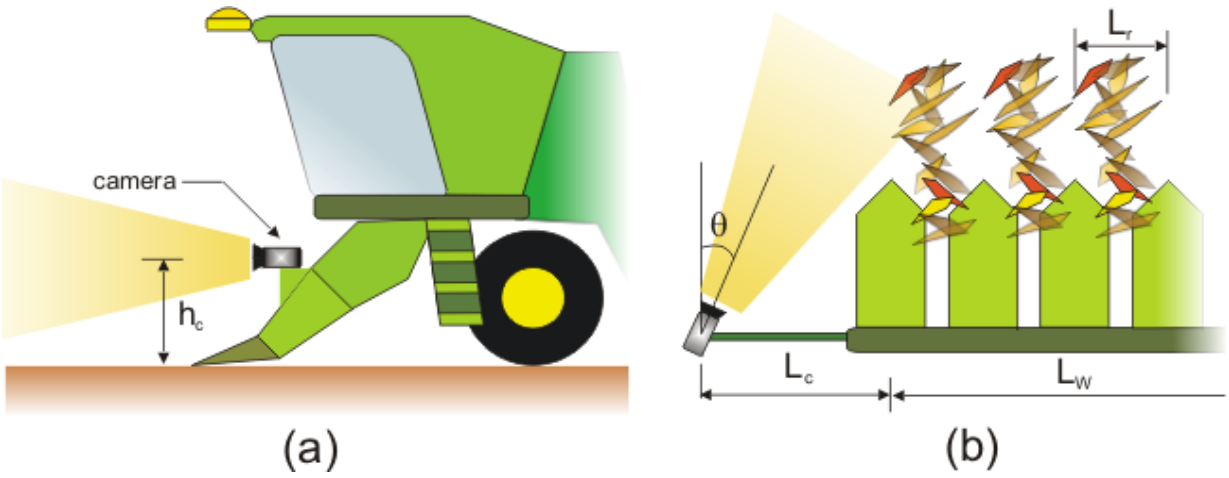


Figure 1. System design for autonomous guidance of a combine using stereo vision: (a) side view, and (b) top view.

3. FINDING THE EDGE

Before obtaining high quality point clouds, it is necessary to acquire satisfactory stereo images, that is, to get acceptable disparity images. In the first configuration of the system the camera was mounted just on the head's left side without using the extension arm and looking ahead (θ = 0 º

and $L_c = 0$) at 1.61 m high. Figure 2 shows a raw left image (a) and its corresponding disparity image (b) where dark intensities represent far objects and black stands for filtered mismatches.

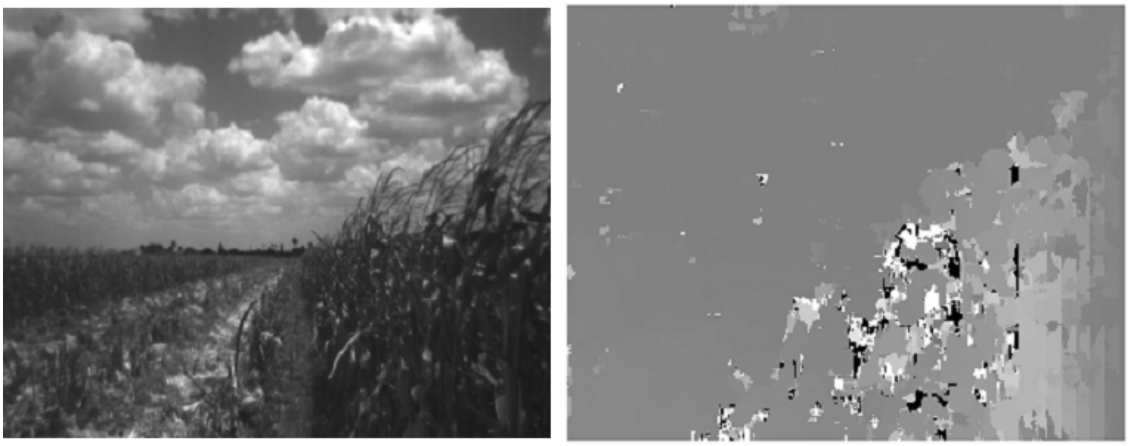


Figure 2. Outputs from the stereo camera: (a) left image, and (b) disparity image.

The disparity image stores the 3D information of the sensed scene in camera coordinates. The first step in the data process is a transformation from the original camera coordinates to the ground coordinates as defined by Rovira-Más and Reid (2004). The result of that transformation is a point cloud of the scene given in a more intuitive frame. Figure 3 represents the 3D cloud of the scene given in figure 2: (a) side view YZ; (b) front view XZ; and (c) top view XY.

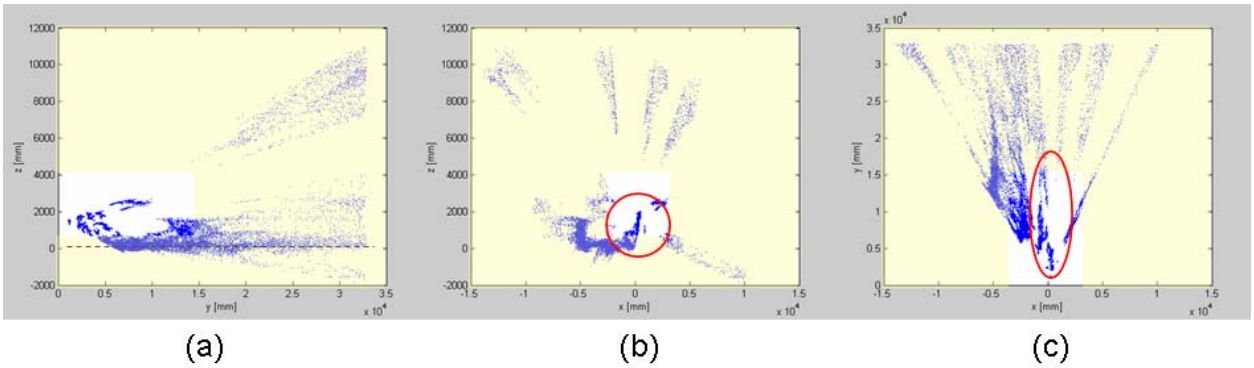


Figure 3. 3D point cloud of the scene given in figure 2: (a) side view, (b) front view, and (c) top view.

The ground system of coordinates places its origin at ground level right under the optical center of the left lens. The approximate expected position for the cut-uncut edge can be determined beforehand, and therefore 3D information can be restricted to important zones. Apart from reducing the size of the point clouds, the processing time is also reduced. The side view of the point cloud given in figure 3 (a) shows wrong 3D locations for heights (Z) over 4 m caused by the cloudy sky. The representation of the ground level is rather flat, and it is located at Z = 0, which indicates a sound coordinate transformation; however the 3D information close to the ground level is not important to find the edge, and can be neglected for such purpose. The

depiction of the edge can be easily identified in the front view of figure 3 (b). An edge remaining from previous passes is also noticeable around 5 m left. As shown in the side view, heights over 4 m rendered incorrect information. The top view of the scene (fig. 3 (c)) demonstrates low data reliability for ranges greater than 15 m. For the scene studied in figure 3, a reasonable three-dimensional space for data processing can be defined by (1).

$$x \in [-2500, 2500]\, mm\,; \quad y \in [0, 15000]\, mm\,; \quad z \in [1000, 4000]\, mm \tag{1}$$

The cloud point represented in figure 3, under spatial restrictions similar to the ones stated in expression (1), was converted into a *density grid*, as defined in Rovira-Más and Reid (2004). The Z coordinate was limited to the interval [500, 2000] mm, yielding the 37 x 50 grid of figure 4 (a). The variation of the 3D density (d3d) as a function of the range Y is plotted in figure 4 (b).

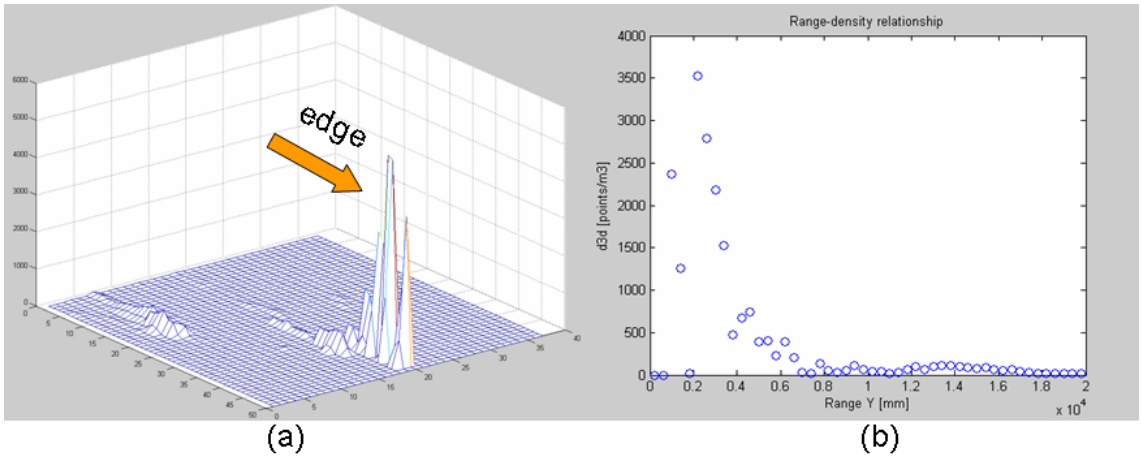


Figure 4. Density grid representation: (a) grid, and (b) density-range relationship.

The graphic of figure 4 (b) indicates that the 3D density drops very fast in the first 8 m, being very low beyond 10 m. In spite of this unbalance, the position of the edge is well determined, as indicated in figure 4 (a). In this particular application, where the goal is to detect the cut-uncut edge, since the key information resides in those cells located close to the camera, there is, apparently, no urgent need to range-compensate the 3D density.

With the aim of enhancing the position of the cut-uncut edge so that it can be detected more reliably, the gradient mask (V) of equation (2) was applied to a density grid defined around the location of the edge. The resulting graphical representation is plotted in figure 5: (a) gives a three-dimensional view of the value of the gradient for every cell, and (b) provides the same data through a two-dimensional color map.

$$V = [-1 \quad -2 \quad 0 \quad 2 \quad 1] \tag{2}$$

The gradient mask marked transitions rather intensely, producing a large increment negative maximum- positive maximum where the edge lays. Once the gradient grid was calculated, the maximum value for each row was registered, as well as its cell location. After the computation of the maximum gradient values, a threshold (5% of the maximum value in the grid) was applied to

eliminate low-reliability locations. Finally, the X coordinates of the remaining cells were stored to estimate the vehicle's offset.

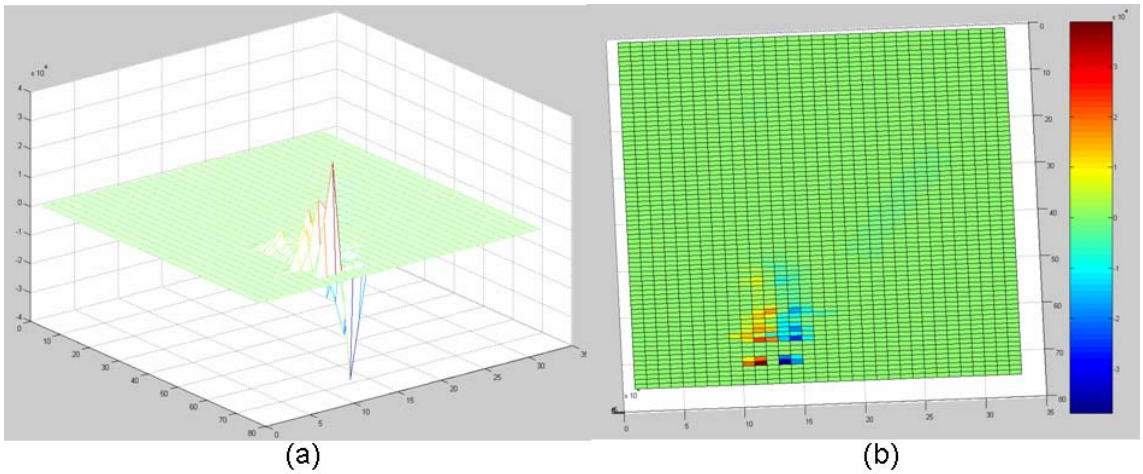


Figure 5. Application of Gradient Operation: (a) 3D view, and (b) 2D color map.

When only the first row was detected (row that signals the position of the cut-uncut edge), the gradient operation yielded similar results to the outcomes obtained after detecting the edge directly from the density grid. However, the gradient became more valuable when special situations occurred. One such circumstance took place when several rows were detected in the same scene. The gradient procedure marked the largest transition between the maximum and minimum peaks for the edge. Figure 6 (a) represents the gradient response when more that one row is captured in the image. Figure 6 (b) provides the locations of the found rows. The highest peak to peak difference determined the placement of the edge, and the density of those rows detected on the right side of the chosen row were weighted by a reduction coefficient before calculating the final position of the edge.

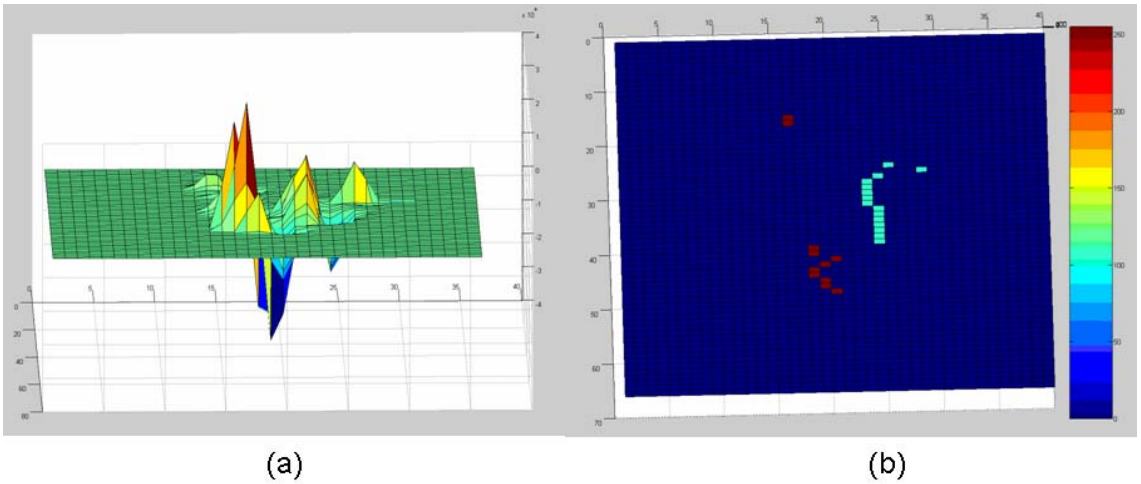


Figure 6. Edge detection when several rows are captured in the scene:
(a) 3D view, and (b) 2D map.

Even though the camera was initially installed on the left side of the combine head with $\theta = 0$ º and $L_c = 0$ (fig. 1), the navigation tests were performed with a significantly different configuration: $\theta = 30$ º and $L_c = 1.42$ m. The reason for such modifications is due to the need to generate more robust stereo images and avoid the detection of multiple rows as shown in figure 6. An even more problematic situation arose when the corn leaves got so near to the camera that they touched the lenses and blocked part of the field of view. In that case, the cut-uncut edge was not always properly detected, causing instability problems. Figure 7 illustrates the situation of a camera mounted too close to the actual crop: (a) left image of the stereo pair; (b) gradient representation of the scene; and (c) approximate position of the edge after the inner rows have been penalized.

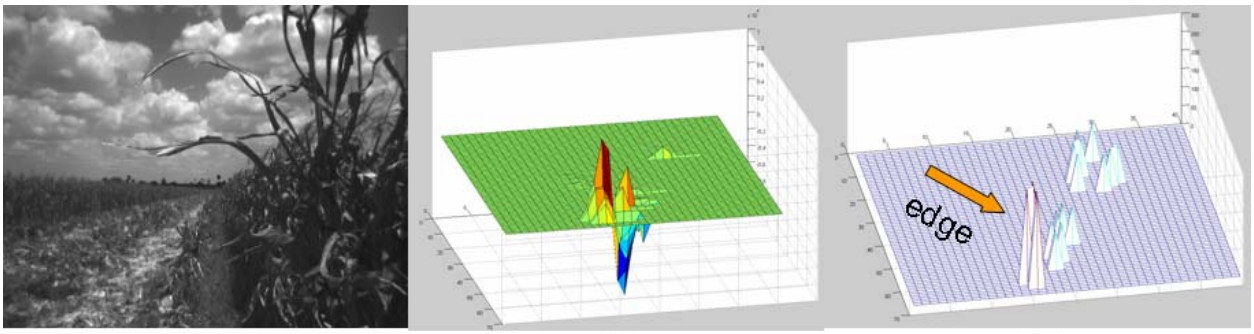


Figure 7. Inappropriate distance camera-crop: (a) raw image, (b) gradient plot, and (c) estimated edge position.

4. CONFIDENCE METRICS

With the second configuration for the camera position ($h_c = 1.66$ m, $\theta = 30$ º, $L_c = 1.42$ m.), the estimates for the edge's position improved, yet sometimes there were appreciable differences between consecutive estimates. The objective of the *confidence index* is to assess the quality of the edge position estimation in such a way that every result (estimate) is accompanied by a quality indicator, representing a confidence value for the solution. The expression for the Confidence Index (*CI*) is defined in (3), where M_x is the moment index and C_x is the clustering index, whose definitions are included in the next paragraphs.

$$CI = M_x \cdot C_x \tag{3}$$

4.1 Moment and Moment Index

Figure 8 represents an archetypal output of the edge detection algorithm: a set of cells arranged approximately drawing a line. For every single valid cell, three parameters are known: *x* coordinate, *y* coordinate, and 3D density. Valid cells are those whose density exceeds certain threshold, determined after the gradient operation. The optimal position of the edge is approximated by a regression line applied to the valid cells.

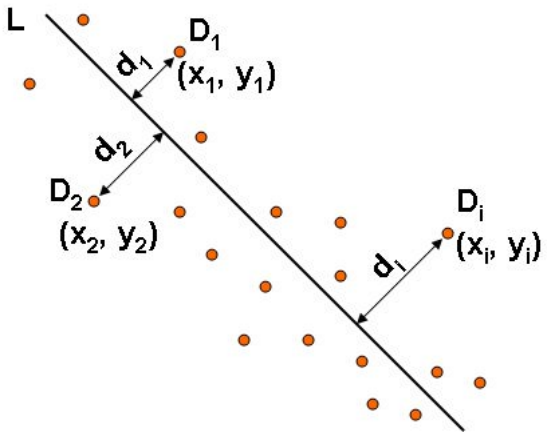


Figure 8. Confidence index calculation: Moment definition.

The *Moment* for a particular image *k* is defined in equation (4), where *n* is the number of valid cells after thresholding, d_i is the perpendicular distance from valid cell *i* to the fit line *L*, and D_i is the density of the cell number *i*.

$$M_k = \sum_{i=1}^{n} \frac{d_i^2}{D_i} \tag{4}$$

If the fit line *L* is determined by the equation $L \equiv ax + by + c = 0$, the perpendicular distance for the valid cells d_i can be calculated following the conventional expression of equation (5):

$$d_i = \frac{|a \cdot x_i + b \cdot y_i + c|}{\sqrt{a^2 + b^2}} \tag{5}$$

Due to the fact that the *Moment Index* was intended to range between 0 and 1, a maximum value reachable by the *Moment* had to be deduced. Such value was named the *estimated maximum moment of the image k* ($MoMax_k$), and was defined according to equation (6). Note that an adjustment constant equal to 2 has been arbitrarily added to the denominator in order to obtain more rational values for the *Confidence Index CI*. The other parameters taking part in the equation are d_{max} as the maximum distance considered in the image *k* after the threshold has been applied, and D_{min}, the non-zero minimum density recorded among the *n* valid cells.

$$MoMax_k = \frac{n \cdot d_{\max}^2}{2 \cdot D_{\min}} \quad ; D_{\min} \neq 0 \tag{6}$$

The expression of the *Moment Index* is given in equation (7). It is, by definition, a number comprised in the interval [0, 1], and it will decrease when the valid cells are located far from the regression line and having low densities, since that specific combination will enlarge M_k.

$$M_x = 1 - \frac{M_k}{MoMax_k} \tag{7}$$

4.2 Clustering Index

An alternative way to evaluate the capacity of the valid cells to delimit the cut-uncut edge is by quantifying their tendency to cluster around the optimal line L. A basic procedure to quantify that clustering ability is by analyzing the distribution of the perpendicular distances d_i calculated with equation (5). The standard deviation of the perpendicular distances d_i for a given image is shown in equation (8), where $\overline{d}$ is the average perpendicular distance:

$$\sigma_d = \sqrt{\frac{\sum_{i=1}^{n}(d_i - \overline{d})^2}{n-1}} \tag{8}$$

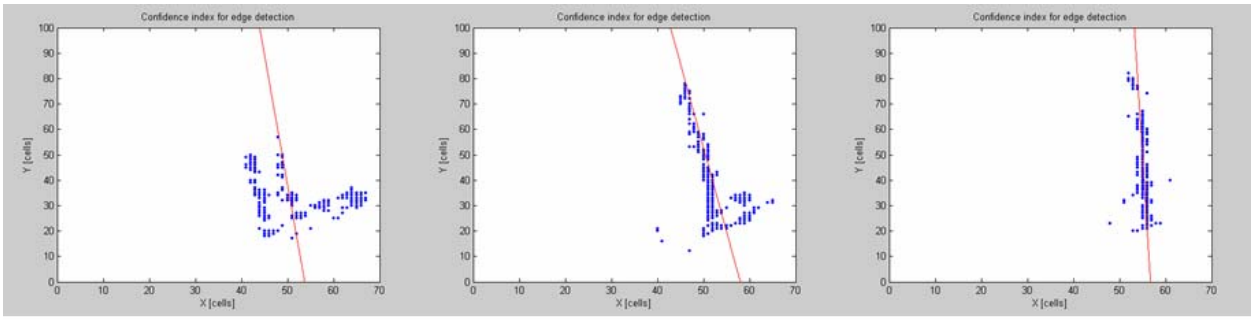
The *Clustering Index* C_x is finally defined in equation (9a), with the conditional constraint stated in equation (9b), which assures that C_x is never above 1.

$$C_x = 0.8 + \frac{0.2}{\sigma_d} \tag{9a}$$

$$If\ C_x > 1 \quad \Rightarrow \quad C_x = 1 \tag{9b}$$

The ultimate goal of the confidence metrics is to rate the quality of the estimate (offset) and weight its contribution to the final control command accordingly, mainly if sensor fusion is employed for automatic guidance. In the navigation tests performed with the combine, the *Confidence Index* was calculated but no action was taken in the steering based on such index.

The plots represented in figure 9 give three examples of the application of the confidence metrics to the edge detection algorithm: (a) CI = 60%; (b) CI = 80%; and (c) CI = 94%. Table 1 lists the key parameters involved in the calculation of CI for the cases represented in figure 9.



(a) (b) (c)

Figure 9. Confidence Metrics for edge detection: (a) CI = 60%, (b) CI = 80%, and (c) CI = 94%.

Table 1. Basic parameters obtained in the calculation of the *Confidence Index CI*.

Case	*n*	*M*	d_{max}	D_{min}	*MoMax*	M_x	σ_d	$\overline{d}$	C_x	*CI*
(a)	154	112	15	45	385	0.71	4.7	5.3	0.84	60%
(b)	168	33	13	34	417	0.92	2.9	1.5	0.87	80%
(c)	121	2	6	44	49	0.96	1.1	0.2	0.98	94%

5. OFFSET ESTIMATION

The methodology covered thus far tries to identify the approximate location of the cut-uncut edge and, at the same time, provide an estimate of the quality of the solution found. Once the grid cells that bear the key information have been isolated, the offset of the vehicle is computed from the position of the selected cells. The *offset* can be defined as the lateral distance between the actual position of the combine and its optimal position determined by the location of the cut-uncut edge. The offset is the output of the edge detection algorithm, and is sent out to the steering control unit of the AutoTrac system. The first step in the calculation of the offset is to generate the profile of cumulative densities by summing up the densities for every column of the grid, eliminating the data stored in the boundaries of the grid (20%). Figure 10 (a) represents the profile of cumulative densities for several images whilst figure 10 (b) plots the profile for a single image. In the former case, it is difficult to decide where the best placement for the edge is. The latter gives a clearer idea of the possible situation for the searched edge, however, there are two distinct peaks separated by 10 cells, and an extra verification would be helpful before making the final decision. The application of the gradient operation, defined in equation 2, to the profile of figure 10 (b) gave the profile of figure 10 (c), where the biggest peak to peak difference provides a redundant value for the approximated position of the edge. The most accurate estimates were obtained with the gradient operation, although most of the times the results from both plots coincided. After the final position for the solution cell has been chosen, the last step consists of transforming the position of that cell to conventional length units, taking into account the extension of the arm L_c and the width of the combine head fingers.

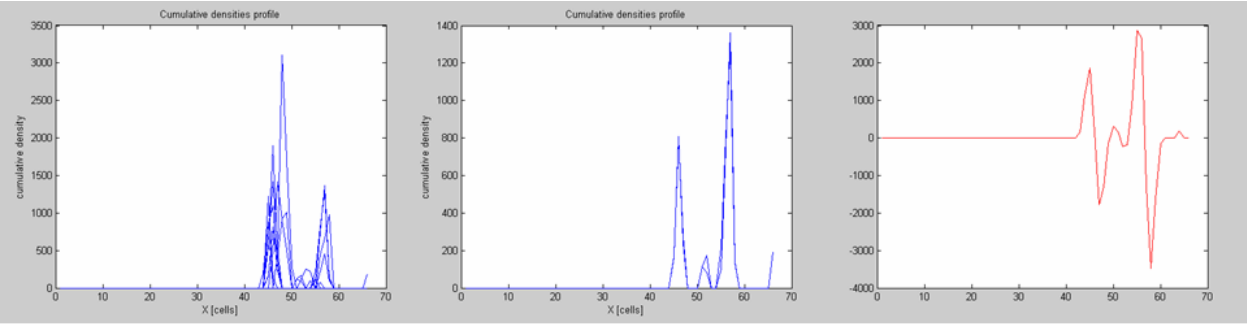


Figure 10. Offset estimation: (a) cumulative densities of several images, (b) cumulative densities of a single image, and (c) gradient applied to the profile of cumulative densities of (b).

6. RESULTS

The results obtained during the preparatory tests led to the final configuration camera-vehicle defined by the parameters h_c = 1.66 m, θ = 30 º, and L_c = 1.42 m (fig. 1), as well as to the development and tuning of the algorithm. The conclusive experiments to examine the behavior of the entire system were carried out guiding the combine autonomously through a cornfield in central Iowa (USA), on 21 October 2005. Since the auto-steering operation was executed by the AutoTrac system, it was not possible to introduce modifications in the steering controller or vehicle dynamics model; correct offsets given by the algorithm were supposed to produce stable and precise navigation. The assessment of the quality of the automatic guidance was realized

visually, paying attention to possible damage caused to the crop. Corn usually leans as the snouts of the combine head run over the plants laterally, which is easily detected from the cabin. The harvester was automatically guided through the field for several passes at speeds ranging from 2.5 to 4 km/h, a representative velocity interval for corn harvesting operations. The stability of the vehicle greatly depended on the quality of the images and, consequently, the edge detection results. Figure 11 shows two results of the edge detection algorithm when the edge was not the only feature detected: in (a) there were some leaves registered in the grid on the left of the edge; and in (b) there were plants belonging to the inner rows detected on the right side of the cut-uncut edge.

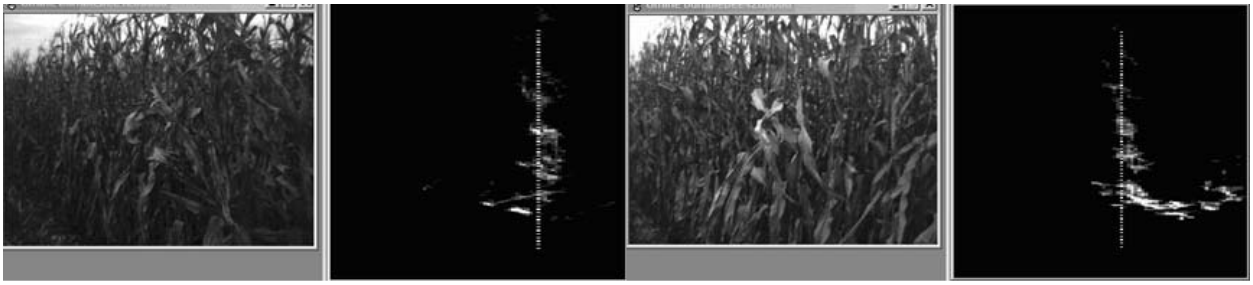


Figure 11. Edge detection results with challenging grids.

Figure 12 represents the opposite case, when the edge is basically the unique feature detected in the density grid. Even though there were some cells providing information on the right side of the edge in (b), they had low density values, and basically did not perturb the estimation of the line. In the four cases presented in figures 11 and 12, the position of the estimated edge (marked by a dotted vertical line) gives the impression of being properly situated.

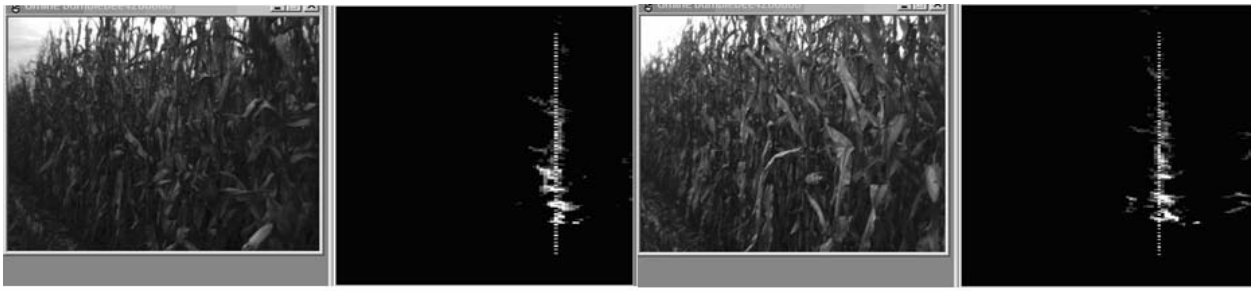


Figure 12. Edge detection results.

7. CONCLUSIONS

The numerous benefits of assisting combine drivers in the monotonous task of harvesting motivated the development of an algorithm for autonomous guidance of harvesters. The elaborated methodology was based on the detection of the cut-uncut edge of corn plants using stereoscopic vision. Several operations and techniques such as density grids and gradient profiles resulted very helpful in estimating the position of the edge. A confidence index to evaluate the

probability of obtaining good results was also developed, although it was not used in the actual control of the vehicle. One of the most sensitive parameters in the configuration of the system was the location of the camera. The best results were achieved with a low camera position on the combine head, and by separating the camera from the cut-uncut edge with an extension arm. Even though this solution turned out to be optimal from the image processing standpoint, it is not the most adequate from the structural point of view: a metal bar with a camera and its cables protruding from the combine head does not seem to be an efficient design. The choice of stereo for the perception of the edge produced good results due to the advantage of having the three dimensions of a scene where the wall created by the edge can be easily detected. The combine was capable of driving autonomously at regular harvesting speeds; however, more guidance tests are necessary to challenge the system and develop a more robust algorithm that can work for a broad variety of conditions. A methodology to evaluate the quality of the automatic task would also be beneficial for comparing the results found with different configurations of the developed system.

8. ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Chenghai Yang and Alfredo Gómez from the USDA Integrated Farming and Natural Resources Research Group (Weslaco, Texas, USA) for their assistance in the image acquisition stage of the research. Appreciation is also conveyed to Deere and Company for the support throughout the entire duration of the project.

9. REFERENCES

Benson, E. R., J. F. Reid, and Q. Zhang. 2001. Machine vision based steering system for agricultural combines. ASAE Paper No. 011159. St. Joseph, Mich.: ASABE.

Khamene, A., and S. Negahdaripour. 1999. Building 3D elevation maps of sea-floor scenes from underwater stereo images. In *Proc. OCEANS '99 MTS/IEEE*, 1: 64- 70.

Kise, M., Q. Zhang, and F. Rovira-Más. 2005. A stereovision-based crop row detection method for tractor-automated guidance. *Biosystems Engineering*, 90(4): 357-367.

Lee, S., D. Jang, E. Kim, S. Hong, and J. Han. 2005. A real-time 3D workspace modeling with stereo camera. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 2140 – 2147.

Murray, D., and C. Jennings. 1997. Stereo vision based mapping and navigation for mobile robots. In *Proc. IEEE International Conference on Robotics and Automation*, 2: 1694 – 1699.

Rovira-Más, F., and J. F. Reid. 2004. 3D density and density maps for stereo vision-based navigation. In P*roc. Automatic Technology for Off-Road Equipment Conference (ATOE 2004)*, 24 – 35.

Rovira-Más, F., Q. Zhang, and J. F. Reid. 2005. Creation of three-dimensional crop maps based on aerial stereoimages. *Biosystems Engineering*, 90(3), 251 – 259.

Sabe, K., M. Fukuchi, J.-S. Gutmann, T. Ohashi, K. Kawamoto, and T. Yoshigahara. 2004. Obstacle avoidance and path planning for humanoid robots using stereo vision. In *Proc. IEEE International Conference on Robotics and Automation (ICRA '04)*, 1: 592 – 597.

Wang, L. K., S. Hsieh, E. C. Hsueh, F. Hsaio, and K. Huang. 2005. Complete pose determination for low altitude unmanned aerial vehicle using stereo vision. In *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 108 – 113.

Wei, J., F. Rovira-Más, J. F. Reid, and S. Han. 2005. Obstacle detection using stereo vision to enhance safety of autonomous machines. *Transactions of the ASABE*, 48(6): 2389 – 2397.