Stereovision Data Processing With 3 d Density Maps For Agricultural Vehicles

Article *in* Transactions of the ASABE (American Society of Agricultural and Biological Engineers) · July 2006

DOI: 10.13031/2013.21721

3 authors:

Francisco Rovira-Más
Polytechnic University of Valencia
82 PUBLICATIONS **2,186** CITATIONS
SEE PROFILE

John Franklin Reid
University of Illinois Urbana-Champaign
156 PUBLICATIONS **3,573** CITATIONS
SEE PROFILE

Quanyi Zhang
Northwestern Polytechnical University
62 PUBLICATIONS **1,755** CITATIONS
SEE PROFILE

STEREOVISION DATA PROCESSING WITH 3D DENSITY MAPS FOR AGRICULTURAL VEHICLES

F. Rovira-Más, J. F. Reid, Q. Zhang

ABSTRACT. *Along with the advent of compact stereo cameras, stereovision information has become an essential part of many perception units of robots and autonomous vehicles. The ease of capturing stereo data clouds is in contrast with the difficulty of processing them reliably in real-time operations. This research developed a methodology to deal with stereo point clouds by extracting useful information for an autonomous vehicle to travel between two points avoiding the obstacles interfering with its trajectory. The novel concept of 3D density and its embodiment through density grid was introduced and verified. Experimental results demonstrated that the application of 3D density and density maps is a suitable technique to cope with complex stereo data, providing useful information to aid in the path-planning routine of a mobile platform.*

Keywords. *3D point clouds, 3D density, Autonomous navigation, Density grids, Obstacle avoidance, Stereovision.*

The combination of fast computer processors and low-cost electronics has undoubtedly played an important role in the popularization of compact stereo cameras. One of the challenges in making a stereoscopic camera useful and practical for agricultural applications is the processing and extraction of useful information from the acquired images. Several technical issues, such as noise caused by stereo mismatches, excessively sparse three-dimensional (3D) clouds, and massive arrays of data, are critical for robotics applications. The main objective of this research is to obtain useful perception information from the output of a compact stereo system, which includes, but is not limited to, finding a navigation path, sensing free space, and detecting obstacles for safeguarding, in addition to the important requirement of real-time processing that allows guiding an agricultural vehicle autonomously in a field.

Various areas of expertise converge in the present application, leading to an overlap of such fields as agricultural machinery, field robotics, and machine vision. As so richly described by McCorduck (2004), artificial intelligence has reached a complete state of maturity, and its applications are spreading out through all potential fields. One such field with a long experience in machine perception is field robotics and mobile equipment. Reid (2004) states that the building blocks of intelligent mobile equipment are machine control, machine awareness, and intelligence. Two basic components of machine awareness are perception and localization, which include the awareness of the posture of the vehicle and its surroundings. Reid also pointed out that conventional wisdom in the machine awareness community was that vision-based technologies were essential for perception and provided the needs of vehicle safeguarding. But what were those needs? What kind of challenging scenes were expected in a traditional agricultural field? Noguchi et al. (2002) provided some insights into these questions, concluding that when the environment was an outdoor space with many disturbances resulting from variable soil and weather conditions, there were many problems in developing a robust vehicle guidance system for crop production. Even the case of row detection by image processing, a technique widely investigated, presented fairly difficult problems when dealing with such disturbances as shadows, weeds infestation, and a variety of soil color and types. Despite the broadly shared belief in the importance of vision-based technologies for machine perception, machine vision is not exempt from difficulties within agricultural applications. As affirmed above, shadows, weeds, and soil variations can make a camera-based implementation very difficult (Noguchi et al., 2002). Stereovision, on the other hand, provides what is most wanted in robot perception: range to target.

It is evident that a mobile vehicle with autonomous capabilities requires a satisfactory perception system to sense its surroundings and ensure safe operations. The most common sensors used to date for vehicle awareness comprise laser range finders, sonar, and machine vision. However, the possible combination of sensors for that purpose is innumerable: from inertial measurement units, radars, and global positioning systems to such special cases as the composition of perspective 3D views of urban environments by merging road maps, 3D models, and aerial photographs (Brünig et al., 2003). Stopp and Riethmüller (1995) found an original application of a diffusion algorithm (the algorithm simulates a physical diffusion process) on a multi-layer grid, where every layer contained specific information such as fixed obstacles, mobile obstacles, special spatial points (doors, elevators, etc.), unexplored regions, and so on, for path planning. Re-planning in the event of unexpected obstacles

Submitted for review in September 2005 as manuscript number IET 6087; approved for publication by the Information & Electrical Technologies Division of ASABE in June 2006.

The authors are **Francisco Rovira-Más, ASABE Member Engineer,** Contract Professor, Polytechnic University of Valencia, Spain; **John F. Reid, ASABE Fellow,** Product Technology Manager, Deere & Company, Moline, Illinois; and **Qin Zhang, ASABE Member Engineer,** Associate Professor, Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois. **Corresponding author:** Qin Zhang, Department of Agricultural and Biological Engineering, University of Illinois at Urbana-Champaign, 1304 W. Penn. Ave., Urbana, IL 61801; phone: 217-333-9419; fax: 217-244-0323; e-mail: qinzhang@uiuc.edu.

  ISSN 0001–2351

interfering with the planned trajectory was also considered in this algorithm. Okubo et al. (1997) specified the tasks for an autonomous mobile robot needing sensor information: searching landmarks or a goal, and searching a path. In order to achieve the second task, a stereo system was developed to look for passable free space, taking into account that the larger passable spaces may exist in the direction where the farther objects are.

In spite of their limitation for dusty environments, laser range finders have turned out to be one of the most used sensors for perception and localization in experimental autonomous agricultural vehicles. Ahamed et al. (2004) mounted a laser range finder on a tractor to detect reflectors strategically located within the vicinity of a moving vehicle. The main advantage of laser scanners is their accuracy, and the fact that these sensors only scan in one plane does not necessarily limit their scope to 2D perception. Yokota et al. (2004) created an elevation map by merging local maps generated by a tractor-mounted laser range finder. The range finder was installed in such a way that the motion of the tractor produced a screening with 3D capabilities. The selection of elevation maps as a way of yielding 3D information was due to the author's belief in the inappropriateness of rendering global maps with point clouds to describe 3D environments. The elevation map was divided into uniform grids, and object position errors were determined within 20 cm. This application conveniently stored the 3D information collected with the laser sensor in regular grids, a procedure that is becoming a general practice. This practice commenced with the "evidence grids" approach, reported in the pioneering work by Moravec (1996), and was originally applied to fuse sonar and stereo (Martin and Moravec, 1996). Clark and Lee (1998) pursued another grid implementation to create a topographic map from a moving vehicle based on RTK-GPS localization data. Through this approach, different grids of 3 m, 10 m, and 30 m grid spacing were devised. Wallner et al. (1995) described a mapping procedure for robot navigation based on the idea of local probability grids. Sonar and stereo data were fused to generate the grids and integrate 3D information; however, the mapping was not reliable when there were many dynamic obstacles within the vicinity of the robot. Schultz and Adams (1998) integrated a set of 16 sonar sensors in a mobile robot in an attempt to obtain "continuous localization." The registration was capable of correcting odometry with the support of *a priori* long-term maps.

Long before its introduction into agricultural engineering, stereo perception was applied to aid robot navigation. Countless references report its use in field robotics; for example, Murray and Jennings (1997) mounted a trinocular stereo system on a laboratory robot. Obstacles detected through that vision system were mapped into an occupancy grid, which was updated continuously as stereo data were being acquired. Path planning was achieved by minimizing a cost function that was a weighted representation of the proximity to obstacles and the distance to the goal. Grid locations were classified into three basic types: blocked, clear, and unknown. Another laboratory mobile robot was equipped with a binocular camera to build 3D maps to aid autonomous navigation (Kim and Kim, 2003). The feature points were detected with an algorithm based on the Harris' corner detector and the SVD matching method. Beside navigational assistance, stereo is a potential tool for mapping terrain. Van der Mark and Van den Heuvel (2001) converted stereo disparities into height maps to be used for path planning to avoid both positive and negative obstacles. They used grid values to define the average height of small regions of space in front of the vehicle. Negative obstacles generated empty spaces, as well as spatial occlusions in the grid. Several examples of agricultural robots with diverse sensing engines have been reported in the last five years. Some of them implemented a navigation system fusing GPS, IMU, and GIS maps (Noguchi et al., 2002), a stereovision localization system allowing the location of a vehicle within the field (Takahashi et al., 2002), and an automatic tractor guidance system using stereo disparity images (Rovira-Más et al., 2004).

The specific objectives of this research can be summarized as follows:

- Develop a platform to host a stereo system capable of detecting obstacles and finding a path to reach a target point.
- Elaborate a methodology to deal with stereo 3D information based on the novel concept of 3D density and its practical application through density grids.
- Conduct field tests to validate the system developed.

Materials and Methods

Principles of Stereoscopic Vision

The main benefit of stereoscopic vision over conventional monocular vision is the capability of detecting ranges, that is, distances between scene objects and the camera. Monocular cameras create planar images where every pixel is the result of the two-dimensional projection of the 3D world. Stereovision adds the third coordinate, or range, that completes the full localization of any point in a 3D Cartesian frame.

With the calculation of the camera coordinates (x_C, y_C, z_C) for the objects in the scene, the acquisition phase ends at the same time as the processing, analysis, and decision-making operations begin. The main focus of this article is how to extract useful information for autonomous vehicle navigation from an irregularly distributed 3D cloud of points, some of which will come from noisy mismatches. Once the camera coordinates are known, the first subject that arises is the adequacy of the camera coordinates for vehicle navigation perception. As illustrated in figure 1, the camera coordinates can be described as follows: the X_C Y_C plane is coincident with the imagers' plane, the range Z_C is perpendicular to the X_C Y_C plane; the Y_C axis follows the vertical dimension of the image, increasing downwards; and the X_C axis coincides with the horizontal dimension of the image, with positive values increasing from left to right. The camera coordinates are centered at the optical center of one of the lenses, arbitrarily determined by the camera's manufacturer. In figure 1, the origin of the camera coordinates is set at the optical center of the left lens, but in the course of this research different stereo sensors with different coordinate origins have been employed. Point P belongs to an object sensed in the scene, and its camera coordinates are given by (x_C, y_C, z_C).

In many agricultural applications, the stereo camera needs to be located at a high position, such as at the top of the tractor cab, in order to capture sufficient information from the scene. These situations often require the camera to be inclined

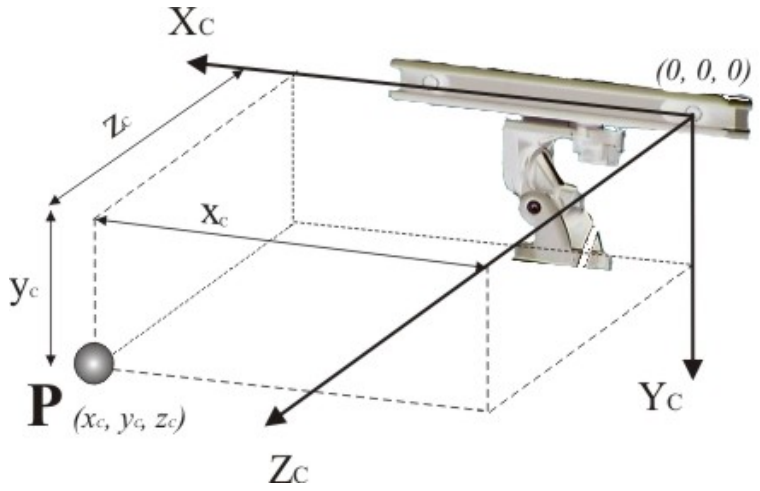

Figure 1. Definition of camera coordinates.

downwards to avoid distracting features in the scene, typically the sky or an irrelevant background. When the stereo camera is inclined at an angle (ϕ), the camera coordinates are difficult to manage, and the range will not be the horizontal distance between the objects and the camera. For this reason, a new frame of ground coordinates was defined. The ground coordinates introduce a more convenient and intuitive definition of coordinates where, independently of the camera position and orientation, the *Z* axis indicates the height of the objects, the *Y* axis gives the distance between the camera and the object, and the *X* axis registers the horizontal position with respect to the center of coordinates. Figure 2 displays the ground coordinates of a stereo system mounted on a utility vehicle. The origin of the ground coordinates is located at ground level, keeping the same *XY* position as the center of the camera coordinates, but ensuring that the *Z* coordinates represent the actual height of objects.

The transformation between the camera coordinates and the ground coordinates is formulated in equation 1. The diagram in figure 3 clarifies the relationship between the two types of coordinates, showing the geometry involved in the transformation for point *P*. The origin of the camera coordinates is $(0, 0, 0)_C$, and its axes are given by X_c Y_c Z_c. Similarly, the ground coordinates are centered at (0, 0, 0) and the frame is defined by *XYZ*. Figure 3 illustrates the difficulties of dealing with the camera coordinates in comparison with the ground coordinates. Since *x* and x_C are coincident, the transformation can be simplified by considering the plane *YZ* (or Y_C Z_C):

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\cos\phi & \sin\phi \\ 0 & -\sin\phi & -\cos\phi \end{bmatrix} \cdot \begin{bmatrix} x_C \\ y_C \\ z_C \end{bmatrix} + h_C \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

where (x_C, y_C, z_C) are the camera coordinates, (x, y, z) are the ground coordinates, h_c is the camera height measured at the optical center of the lens, and ϕ is the camera inclination angle.

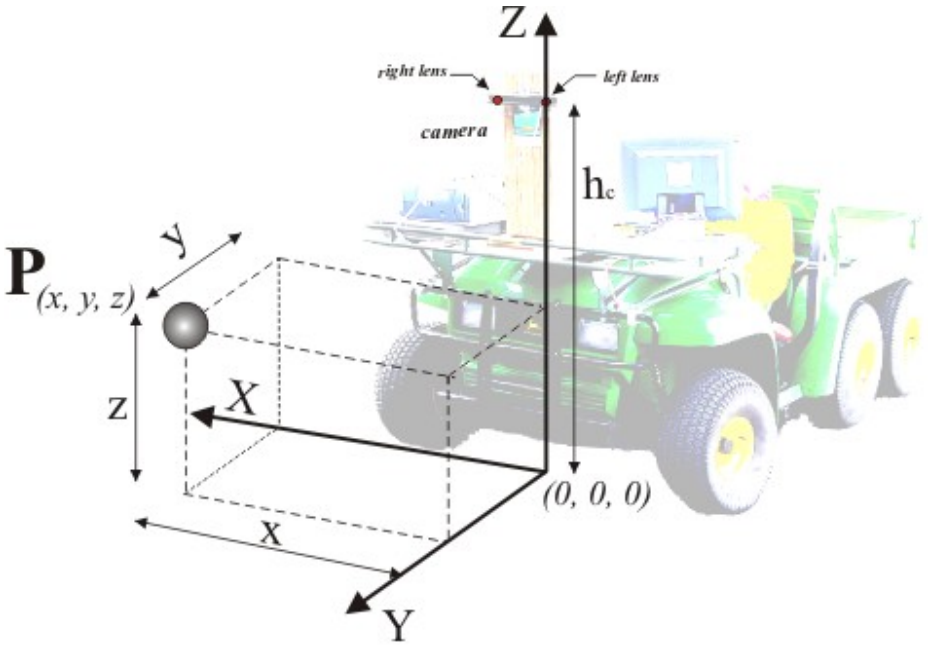

Figure 2. Definition of ground coordinates.

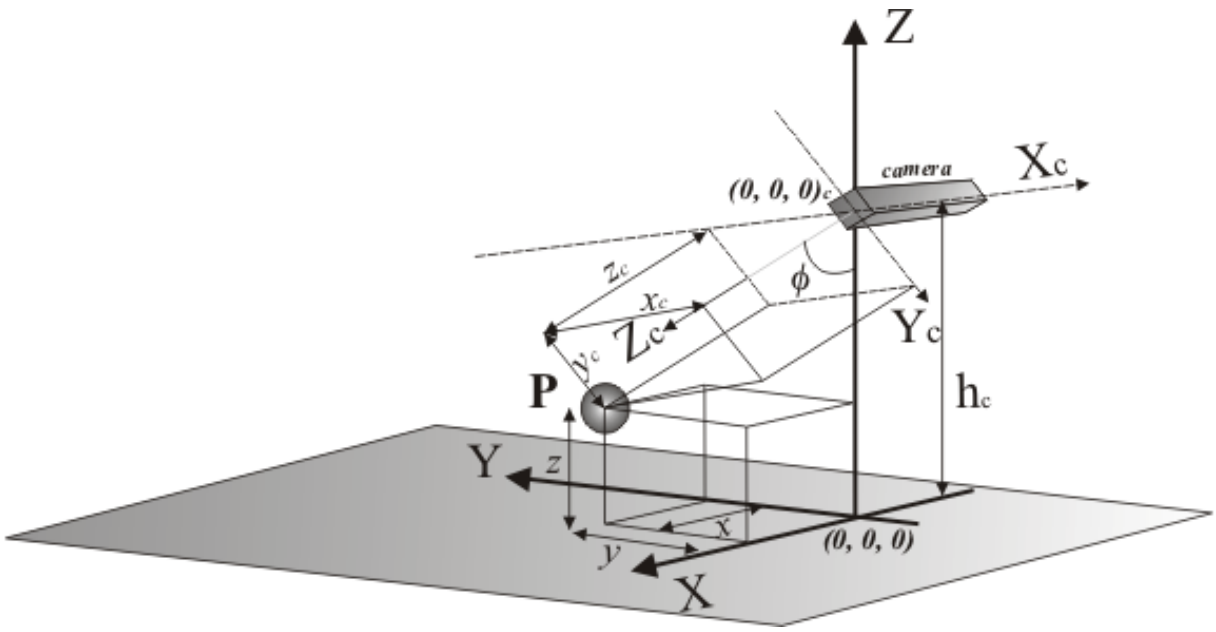

Figure 3. Coordinate transformation between camera and ground coordinates.

The natural outcome of a stereovision sensor is a 3D cloud of points that renders the captured scene with a degree of detail proportional to the resolution of the acquired images. Every single point of the 3D cloud comes from a stereo-matched pixel and will be endowed with three coordinates that locate its position in the space. The initial system of coordinates employed for the point cloud is the camera coordinates, as illustrated in figure 1. After the appropriate transformation and analysis, the initial cloud will provide the significant information for vehicle perception that results from eliminating invalid noise and a distracting background. The generation and manipulation of 3D clouds possesses several intricacies that must be addressed before integrating stereo data in a perception and navigation engine. The most notable issues encountered during the experimental phase were the following:

Stereo mismatches between the right and the left images. When a pixel is wrongly paired, the coordinates of the corresponding point will be incorrect, often pointing at unlikely, if not impossible, locations. This common problem was handled in this application by the developed 3D density concept together with the validity box limitation approach, which consists of defining bounds for the targeted space in front of the camera. For example, if a vehicle is going to navigate inside an orchard where the height of the trees is around 3 m, and the height of the vehicle is 2 m, any point standing higher than 5 m will be neglected and eliminated from the initial cloud. Off-the-shelf cameras typically include software libraries for disparity calculation that incorporate filters for mismatches.

Weakly defined texture. This is an important handicap for indoor applications, where uniform and artificial structures such as walls, doors, or floors present even surfaces with a poor texture pattern. An area with scanty texture creates a complex scenario for pixel matching because adjacent pixels have similar intensities and a singular position cannot be uniquely determined. On the contrary, this problem is not common for outdoor agricultural applications, where the texture found in trees, plants, and typical orchard scenes often ensures rich disparity images. A scene with deprived texture will certainly produce a mediocre disparity map.

Deficient illumination. Changes in illumination affect left and right images simultaneously. Taking into account that the matching routine compares the intensity pattern distribution between left and right images, a similar variation of intensities will keep the same structure and therefore will not have a dramatic effect on the stereo calculations. Just the opposite happens with monocular cameras, which are very sensitive to ambient illumination changes. In spite of the robustness of stereo cameras to adapt to lighting conditions, poor illumination results in a lack of texture, and consequently a weak disparity image leading to a sparse 3D cloud. The selection of a stereo sensor must consider the type of illumination expected: pre-calibrated or changeable optics cameras. The former are convenient because they do not require the delicate process of calibration although they typically include fixed optics with no possibility of adjustment and control. To the other extreme, stereo cameras with interchangeable lenses allow for a customized aperture control that can improve disparity images in difficult situations, but a careful calibration is necessary every time a lens is removed or the baseline is modified.

Massive 3D cloud size. A typical disparity image of average resolution (e.g., 320 × 240 or 400 × 300) can easily include 60,000 disparity points after noise filtration. When several images are processed together, as in the case of assembling a global 3D map, the magnitude of the data files grows considerably, complicating the manipulation and storage of 3D information. The problem becomes more critical when real-time processing is required. In these cases, the solution is often to process one image at a time and delete it after the information has been extracted, but even in these situations the time needed for the stereo calculations can be determinant.

Object Awareness: Concept of 3D Density (d3D)

Once the camera has been appropriately set up and the filtering process optimized, the starting point of the data

interpretation is a 3D point cloud given in ground coordinates. The essential question in stereo sensing applications is whether a particular set of points indicates the presence of a solid object, or contrarily comprises a sparse number of noisy points, randomly distributed, in a portion of the scene actually occupied by empty space. The real issue behind stereo perception is being unaware of existing objects, and the pseudo-detection of non-existent objects (false positives). In addition, not only is the detection of objects the ultimate goal, but also the acknowledgement of their dimensions and other properties that might be of interest for a particular application. Since all the 3D information is carried by the cloud of points, a reasonable way to begin is to count the number of points and analyze their distribution in the sensed space. However, an absolute count is not effective because the number of points grows as the space covered by the camera increases. Density is a physical property of objects that relates the mass of a substance to the volume that it occupies. Inspired by the physical density of objects, a property of stereovision 3D clouds can be defined in a way that relates the number of detected points and the volume in space occupied by them. This property is designated 3D density (d3D) and is defined as the number of stereo-matched points per volume unit. The mathematical expression for the 3D density is given by:

$$\mathrm{d\,3D} = \frac{N}{V} \tag{2}$$

where V is the volume of space considered, and N is the total number of stereo-matched points inside V.

A practical way of applying the concept of d3D is by dividing the physical space within the camera field of view into a regular grid and then computing the d3D for every cell. An attribute of the 3D density is that, by definition, it is independent of the cell's size, in the same way the density of a liquid is independent of the size of its container. Grids can, *a priori*, be formed of cells of any size; however, the size of the sampled cell will be vital to detecting certain objects and neglecting others. The cell size will also have a remarkable impact on the processing speed of the algorithm. In general, every application will require a different cell size, but the d3D will be comparable among different grids because it does not depend on the size of the cell. The first step in this methodology is the calculation of the d3D. It is important to emphasize that the d3D is directly related to the number of stereo matches (disparity points) and has nothing to do with the physical density of the object perceived. The maximum number of points that a disparity image can yield depends on the image resolution, which complicates the estimation of a threshold to discriminate objects from vacant space. Such a threshold will be affected by the resolution of the image, but also by other factors, such as uneven illumination since a different light intensity over the same object will render clouds of different d3D. Poor illumination will create zones of low-quality texture, which in turn will generate fewer points in the cloud, and vice versa. This irregular behavior between different images, sometimes consecutive, can be corrected by normalizing the density in the following mode:

$$\left|\mathrm{d\,3D}\right| = \frac{\mathrm{d3D}}{\max.\,\mathrm{d3D\ in\ the\ image}} \tag{3}$$

When the concept of d3D was applied to real data acquired in the field, an additional issue caused by uneven density distribution within each image came up. The normalization proposed in equation 3 equates the density patterns among different images, but it cannot level the density between different areas of the same image. This is a more challenging problem and requires specific measures. The cause for internal variation resides in the loss of resolution as ranges grow, i.e., the space covered by a specific pixel increases as the distance to the camera enlarges. Figure 4 illustrates the problem of internal variation of density in an orchard scene: the color image captured by the left imager is given in 4a, and the d3D (stereo-matched points per cell) is shown in 4b. Notice the non-linearity of the density with respect to the range (Y axis in fig. 4b).

The density distribution pattern portrayed in figure 4 required a specific internal normalization, denoted $\mathrm{d3D}_C(Y)$. A mathematical expression for $\mathrm{d3D}_C(Y)$ is deduced in the following paragraphs. Since the cause of variation is a loss of pixel resolution dependent on the range, a potential source of information may be found in the expression for the range resolution of a stereo image (Rovira-Más, 2003), as given by equation 4:

$$\Delta R = \frac{w}{b \cdot f} \cdot R^2 \cdot \Delta d \tag{4}$$

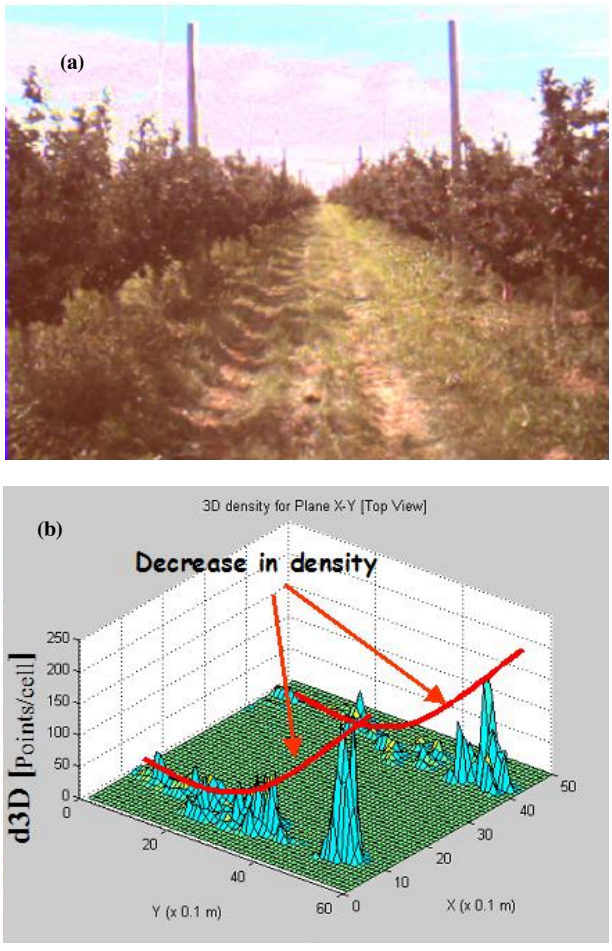

Figure 4. Illustration of internal variation of 3D density in a typical orchard scene.

where ΔR is the range resolution, w is the pixel size, b is the baseline, f is the focal length of the lenses, and Δd is an increment in the disparity.

The quadratic trend of equation 4 was found to be in agreement with experimental results. This expression was introduced in a theoretical expression (eq. 8) that relates range and density. Figure 5b shows the trend of a decreasing d3D as ranges increase. The same plot also includes the best fit for the relationship between the range and the d3D. The sample image acquired in the field is represented in figure 5a. The mathematical expression for the curve fitted to the experimental data plotted in figure 5 is:

$$\text{d3D} = 10^{9} \cdot Y^{-1.7481} \tag{5}$$

where Y is the range (mm), and d3D is the 3D density (points/m^3).

For a given stereo system, the basic parameters that define the properties of the camera, b (baseline), f (focal length), and w (pixel size), are fixed and can be combined in a constant factor K_s, as stated in equation 6:

$$K_s = \frac{b \cdot f}{w} \tag{6}$$

If equation 6 is combined with equation 4, taking into account that the camera constant K_s for the system employed in figures 4 and 5 is 110000, then the range resolution for this stereo system is given by the following expression:

$$\frac{\Delta d}{\Delta R} = 110000 \cdot R^{-2} \tag{7}$$

According to the definition of d3D, there is a close relationship between disparity and d3D. On the other hand, experimental data show that the d3D decreases as the range increases. These two facts lead to the conclusion that there must be a proportional relationship between the range resolution (eqs. 4 and 7) and the d3D. If a proportionality constant of 10^5 is added to equation 7, a theoretical expression that relates the d3D and the range can be found, as given in equation 8:

$$\text{d3D}_{TH} = 11 \cdot 10^{9} \cdot Y^{-2} \tag{8}$$

where d3D_{TH} is the theoretical expression of the d3D as a function of the range Y.

Now the theoretical expression of the d3D can be compared with the actual values of the d3D found in the field.

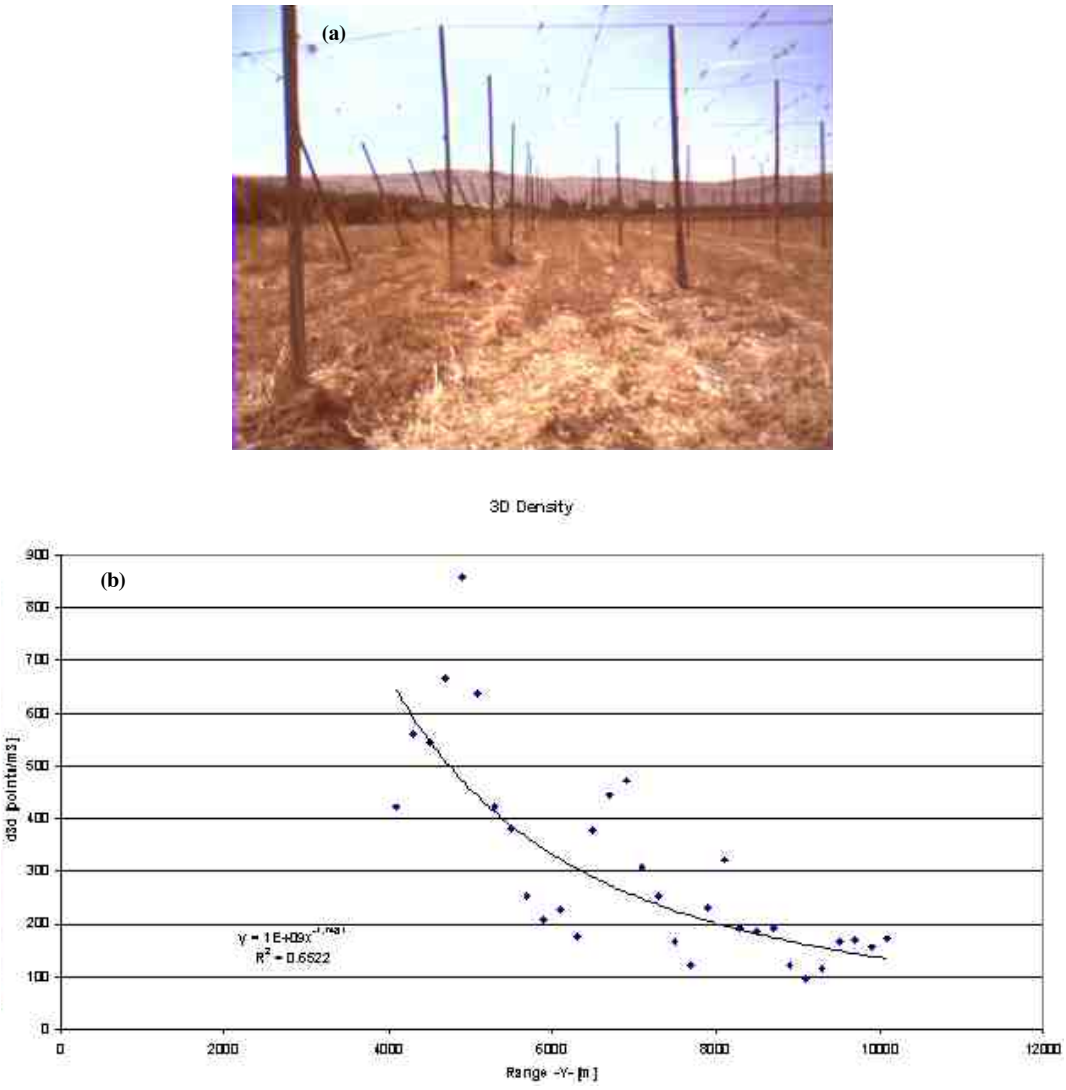

Figure 5. Trend curve for d3D versus range.

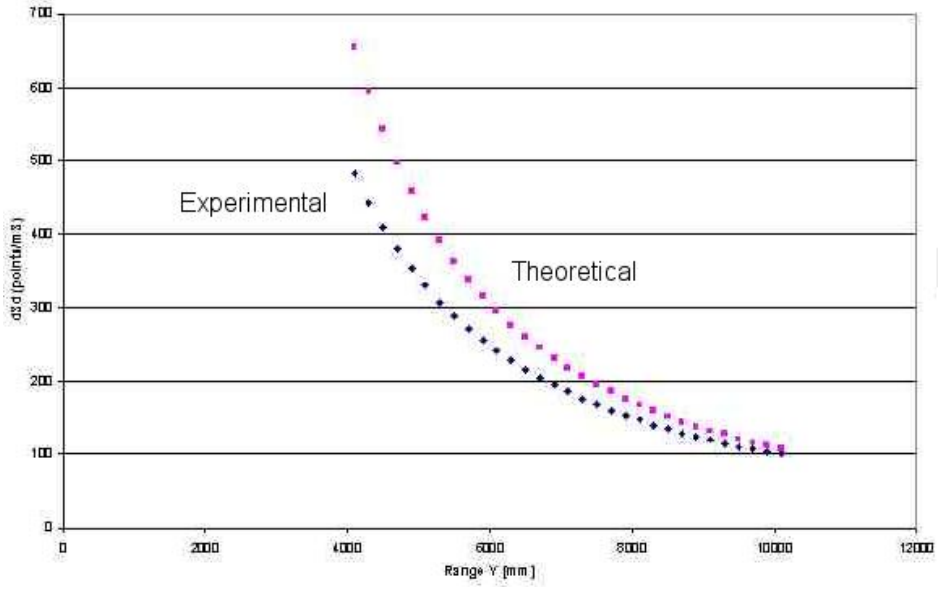

Figure 6. Theoretical and actual relationship between the d3D values and the stereo range.

Figure 6 shows such comparison of experimental and modeled results.

The analysis of various sets of experimental data led to expressions similar to equation 5, with exponents of magnitude close to −2. If a unique threshold is to be applied to the entire image, then some sort of compensation has to be introduced into the formula of the d3D so that densities for large and small ranges can be compared. With this modification, the same threshold can be used to discriminate objects according to their density for any range inside the image. The proposed correction used the theoretical expression of equation 8 to normalize any density to a pre-selected reference range. In the examples shown in figures 4, 5, and 6, the reference range was set at $Y = 4500$ mm. The resulting expression is shown in equation 9:

$$\begin{aligned} \mathrm{d3D}_C(Y) &= \mathrm{d3D}(Y) \cdot \frac{\mathrm{d3D}_{TH}(4500)}{\mathrm{d3D}_{TH}(Y)} \\ &= \mathrm{d3D}(Y) \cdot \frac{11 \cdot 10^9 \cdot 4500^{-2}}{11 \cdot 10^9 \cdot Y^{-2}} \\ &= 4.94 \cdot 10^{-8} \cdot Y^2 \cdot \mathrm{d3D}(Y) \end{aligned} \tag{9}$$

where $\mathrm{d3D}_C(Y)$ is the compensated d3D as a function of the range Y measured in mm.

The new expression presented in equation 9 was applied to the scene presented in figure 4. The compensated d3D is depicted in figure 7, where the d3D does not decay with the increase of the range as quickly as in figure 4b. The effect of the range compensation can be appreciated by comparing figures 4b and 7 as the scene is the same but the d3D presents a more uniform pattern in figure 7. Without this correction, a different threshold is needed for distant objects, given that fewer points represent them. The threshold employed to differentiate objects from empty space was determined by in-field calibration, since every scene requires a particular treatment.

Density Grids

A density grid is defined as an array of cells, either in 3D or in 2D, where the d3D is represented. The most natural way of designing a grid is in a 3D structure, as the data obtained with a stereo camera is given in a 3D cloud where the coordinates x, y, and z are known. However, the excessive computational load demanded by a 3D grid very often results in a simplification to a 2D grid. The specific details of this simplification mainly depend on the application pursued. In general, two configurations have turned out to be efficient for vehicle navigation: a top view arrangement (plane XY) where a look-ahead distance can be set according to navigational needs, and a front view arrangement (plane XZ) where special attention is paid to lateral hazards to the vehicle, such as interfering branches or low-height agricultural structures. Figure 8a depicts a 3D array, and figure 8b shows a simplification of it leading to a two-dimensional top view configuration.

Typically, 3D grids consist of regular cells whose three dimensions are equal (d in fig. 8a), but when downgraded to two dimensions, several slices merge to constitute a unique

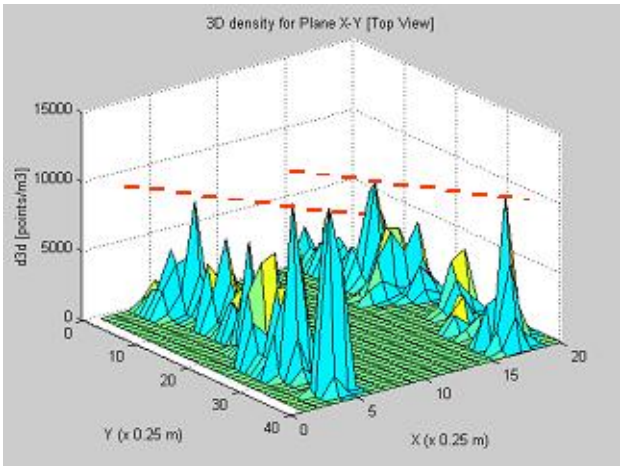

Figure 7. Compensated d3D values for the scene shown in figure 4.

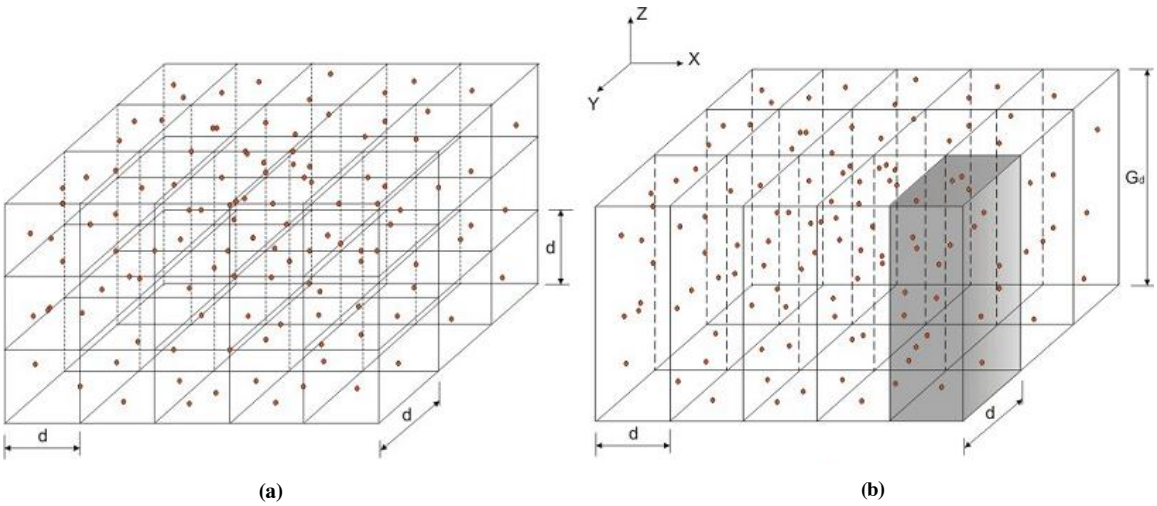

Figure 8. Density grids for stereo perception.

cell in one of the dimensions. Figure 8b represents a horizontal grid (plane *XY*) where the cells have the same dimension (*d*) for *x* and *y*, but the *z* dimension is much larger (G_d), as a result of merging several planes into one unique slice. The magnitude of the *z* dimension of the cell is known as the "depth" of the (two-dimensional) grid (G_d), and the size of the *x* and *y* dimensions is denoted by *d*, or cell size. The number of cells in the grid is the resolution of the grid (5 × 3 in fig. 8b; 5 × 3 × 3 in fig. 8a). The depth of a 2D grid limits the thickness of the space under consideration. If, for example, orchards are to be sensed for autonomous navigation along the rows between the trees, then the portion of the cloud associated with the ground can be neglected because the key information is contained in the trees and not the ground. Likewise, for 3 m tall trees, a slice defined between 0.5 m and 3 m will probably enclose the important information needed for guiding the vehicle between the rows without running over the trees. Neglecting non-essential data increases the speed and efficiency of the process. The cell size (*d*) is directly related to the grid resolution and determines the size of the objects to be detected. Consequently, there is a compromise when selecting the cell size: low-resolution grids (large cells) do not distinguish small objects, but high-resolution grids will result in a heavy computational load. In 2D grids, cell size and grid depth determine the volume of the cells ($V = G_d \cdot d^2$). In 3D regular grids, the cube of the cell size gives the volume of each cell ($V = d^3$). In practical applications, it is desirable to manipulate these parameters with certain flexibility to ensure an optimum performance of the system. In-field parameter tuning is essential to establish the best properties of the grid.

System Architecture

The stereo system developed for this research was implemented in a Gator utility vehicle (Deere & Co., Moline, Ill.) endowed with automatic steering capabilities. The navigation system was comprised of three main functions: a control and localization engine to execute navigational commands, a perception engine to detect obstacles in front of the vehicle, and a path-planner to find an optimum path between the current vehicle's position and a target point, which was arbitrarily selected in the field of view of the stereo camera. The path-planner was based on the A* algorithm (Hart et al., 1968). As a graph search method to find the minimum-cost path, the A* algorithm introduced a heuristic function based on the Euclidean distance to get to the target point. Since the objective of this research is to describe the way stereovision data are processed, the path-planner and the control and localization unit fall outside the scope of this article. The vehicle's operator defined a goal point inside the field of view of the camera, approximately a rectangle of dimensions 8 m (*X*) × 20 m (*Y*). The path-planner found the optimum path avoiding the obstacles present in the field of view. The obstacles were identified by the perception engine, a stereo system featuring the density grid approach. The stereo system consisted of a 22 cm baseline Tyzx stereo camera (Tyzx, Inc., Menlo Park, Cal.) equipped with a DeepSea processing card. The camera was

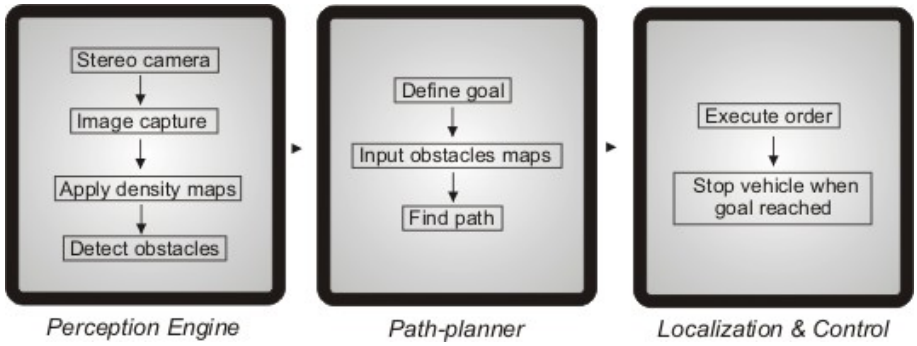

Figure 9. System architecture block diagram.

mounted at the front bumper of the vehicle, approximately 1.5 m above ground level. The navigation controller and the stereo processor were housed in different computers, communicating through an Ethernet cable. The perception computer analyzed the captured image, calculated the position of potential obstacles, and released a path that was sent to the steering controller to execute the command. Figure 9 schematizes the system architecture.

Results and Discussion

To validate the stereo algorithm, several experiments were performed in the courtyard of the Deere & Company Technology Center in Moline, Illinois. A utility vehicle similar to the platform depicted in figure 2 was prepared to embody the architecture described by figure 9. The obstacles employed to challenge the system were plastic cones 1 m high, with the exception of one human-detection test. The camera parameters (baseline, focal length, and image resolution) were adequate for the scenes sensed, where a look-ahead distance between 5 and 20 m was needed for a satisfactory operation. The resolution of the images (400 × 300) and the fine quality of the disparity images resulted in rich data clouds. However, the drop in density was severe for ranges greater than 8 m. The compensation operation for the density yielded positive results, and a general threshold for the entire image was acceptable to discriminate the obstacles from the rest of the objects in the scene. This threshold varied according to the scene, and was adjusted during the tests to reduce the presence of noise in the farthest ranges. The reference plane, set at 4.5 m in the experiments, was chosen based on the experience accumulated during the field tests.

Even though a quadratic range-density relationship seemed a realistic solution in accordance with experimental data, the selection of the reference plane requires further discussion. The camera's basic parameters, mainly baseline and focal length, have a real impact on the determination of the reference plane. Recent experiments performed with a different camera configuration led to reference ranges proportional to the average range of the 3D cloud and to conditional formulas where the range compensation of the 3D density was applied only to those points whose range was greater than a threshold range. The division of the space into a regular grid proved to be a useful operation. The density grid was designed following a two-dimensional top view configuration, whose depth ranged from 300 to 2000 mm for most of the experiments, and the cell size was set to either 125 or 150 mm. The rest of the key parameters for the grid were adjusted to each particular scenario. The grid span was determined by a safety area defined for this application: objects were supposed to appear between 10 and 15 m ahead of the vehicle and less than 3 m from its center plane. These constrictions led to grids with approximately 6 to 8 m × 20 m coverage. The cell size had to be less than half the size of the objects, and the grid depth was designed to avoid useless information (ground) and include substantial information (plastic cones).

Figure 10 shows the obstacle arrangements, the stereo-sensed objects, and the resulting trajectory for three representative experiments. Among them, figure 10a represents a human avoidance case. The principal settings for this case include a grid span of 6 × 20 m with a cell size of 150 mm and a grid depth of 1650 mm, located between z = 350 mm and z = 2000 mm. Considering the camera position as the center of the coordinates for the grid (XY plane), the target point was established behind the standing person at coordinates (0, 10000) mm. The second test, illustrated in figure 10b, created a corridor defined by plastic cones. The goal point was located at the exit of the corridor, so the vehicle had to find its way through the cones without colliding with them to reach the destination point. For this challenge, the grid coverage was 7.5 × 18.75 m, the cell size was 125 mm, and the grid depth was defined by the slice $Z \in$ [200 mm, 2000 mm]. The target point was placed at coordinates (2025, 15937) mm, which gives a distance of approximately 16 m between the camera and the goal. The third case, presented in figure 10c, simulates the typical situation of an obstacle encountered in the middle of the vehicle's path, with

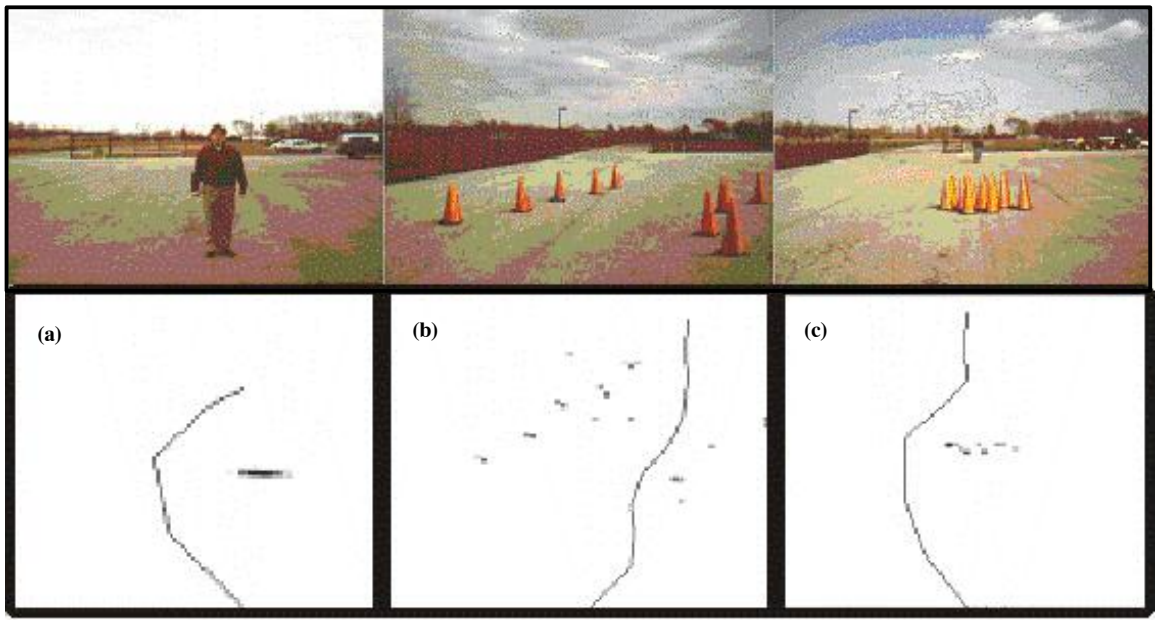

Figure 10. Stereovision obstacle avoidance system based on the concept of d3D and density grids.

the urgent need to detour before resuming the ongoing operation. The basic parameters for this example were the same as in the previous case, except that the target point was located at (0, 17812) mm coordinates.

The vehicle was equipped with numerous emergency stops to halt its motion in the event of problems. In the three cases represented in figure 10, and in the majority of the rest of the experiments, the vehicle reached the target point after a neat sequence of maneuvers. The algorithm was not programmed to re-plan the trajectory; therefore, when the vehicle traveled too close to the objects, it could not autocorrect the error. The selection of cells with sizes under 20 cm ensured the detection of the plastic cones. Noise was discarded by tuning thresholds in the actual scene and ignoring the data points below 30 cm high.

Conclusions

The current technology in stereoscopic vision has led to such a degree of maturity that commercial compact cameras can render reasonable 3D plots of real scenes in real time. The processing of stereo data to extract useful information and make sound decisions in real time habitually entails non-trivial technical difficulties. This article develops the concept of 3D density (d3D), and its practical realization through density maps, to aid in the task of managing stereo data to provide functional information to an autonomous off-road vehicle. The methodology devised in this research was implemented in a utility vehicle with the objective of reaching a target point selected from the stereo camera field of view while avoiding the obstacles encountered in the vehicle's way. The objects were detected by means of a stereo algorithm based on density maps. The trajectory followed by the vehicle was constructed by a path-planner designed following the A* algorithm. Experimental results revealed that the 3D density decreased according to a quadratic curve as the range (y coordinate) increased. Compensation formulas were deduced to correct this tendency, and the concept of 3D density was satisfactorily applied to horizontal two-dimensional grids with look-ahead distances of up to 20 m. The stereo-based navigation system was successfully validated in a series of experiments. The stereo camera turned out to be an efficient sensor for detecting obstacles, and this technology was proven to have great potential in the integration of complex intelligent systems for autonomous vehicles where multiple types of vehicle behavior need to be put into practice for a safe operation.

Acknowledgements

The material presented in this article was based on work supported by Deere & Company. The authors would like to express their thanks to Jeff Puhalla and Tim Scott of Phoenix International, Inc., for their contribution in the development and integration of the path-planner onto the utility vehicle. Appreciation is also conveyed to Christopher Turner and Zach Bonefas of the John Deere Technology Center for their assistance in the preparation of the vehicle and software support, respectively. Any opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of Deere & Company, the Polytechnic University of Valencia, and the University of Illinois.

References

Ahamed, T., T. Takigawa, M. Koike, T. Honma, A. Yoda, H. Hasegawa, P. Junyusen, and Q. Zhang. 2004. Characterization of laser range finder for in-field navigation of autonomous tractor. In *Proc. ATOE Conference*, 120-130. St. Joseph, Mich.: ASAE.

Brünig, M., A. Lee, T. Chen, and H. Schmidt. 2003. Vehicle navigation using 3D visualization. In *Proc. IEEE Intelligent Vehicles Symposium*, 474-478. Piscataway, N.J.: IEEE.

Clark, R. L., and R. Lee. 1998. Development of topographic maps for precision farming with kinematic GPS. *Trans. ASAE* 41(4): 909-916.

Hart, P. E., N. J. Nilsson, and B. Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE Trans. on System Science and Cybernetics*, SCC-4(2): 100-107.

Kim, Y., and H. Kim. 2003. Dense 3D map building for autonomous mobile robots. In *Proc. IEEE Intl. Symposium on Computational Intelligence in Robotics and Automation*, 169-174. Piscataway, N.J.: IEEE.

Martin, M. C., and H. P. Moravec. 1996. Robot evidence grids. Tech. Report CMU-RI-TR-96-06. Pittsburgh, Pa.: Carnegie Mellon University, The Robotics Institute.

McCorduck, P. 2004. *Machines Who Think.* Natick, Mass.: A. K. Peters.

Moravec, H. P. 1996. Robot spatial perception by stereoscopic vision and 3D evidence grids. Tech. Report CMU-RI-TR-96-34. Pittsburgh, Pa.: Carnegie Mellon University, The Robotics Institute.

Murray, D., and C. Jennings. 1997. Stereo vision based mapping and navigation for mobile robots. In *Proc. IEEE Intl. Conference on Robotics and Automation*, 1694-1699. Piscataway, N.J.: IEEE.

Noguchi, N., M. Kise, K. Ishii, and H. Terao. 2002. Field automation using robot tractor. In *Proc. ATOE Conference*, 239-245. St. Joseph, Mich.: ASAE.

Okubo, A., A. Nishikawa, and F. Miyazaki. 1997. Selective reconstruction of a 3-D scene with an active stereo vision system. In *Proc. IEEE Intl. Conference on Robotics and Automation*, 751-758. Piscataway, N.J.: IEEE.

Reid, J. F. 2004. Mobile intelligent equipment for off-road environments. In *Proc. ATOE Conference*, 1-9. St. Joseph, Mich.: ASAE.

Rovira-Más, F. 2003. Applications of stereoscopic vision to agriculture. PhD diss. Urbana, Ill.: University of Illinois at Urbana-Champaign, Department of Agricultural and Biological Engineering.

Rovira-Más, F., Q. Zhang, and J. F. Reid. 2004. Automated agricultural equipment navigation using stereo disparity images. *Trans. ASAE* 47(4): 1289-1300.

Schultz, A. C., and W. Adams. 1998. Continuous localization using evidence grids. In *Proc. IEEE Intl. Conference on Robotics and Automation*, 2833-2839. Piscataway, N.J.: IEEE.

Stopp, A., and T. Riethmüller. 1995. Fast reactive path planning by 2D and 3D multi-layer spatial grids for mobile robot navigation. In *Proc. IEEE Intl. Symposium on Intelligent Control*, 545-550. Piscataway, N.J.: IEEE.

Takahashi, T., S. Zhang, and H. Fukuchi. 2002. Acquisition of 3-D information by binocular stereovision for vehicle navigation through an orchard. In *Proc. ATOE Conference*, 337-346. St. Joseph, Mich.: ASAE.

Van der Mark, W., and J. C. van den Heuvel. 2001. Stereo-based navigation in unstructured environments. In *Proc. IEEE Instrumentation and Measurement Technology Conference*, 2038-2043. Piscataway, N.J.: IEEE.

Wallner, F., R. Graf, and R. Dillmann. 1995. Real-time map refinement by fusing sonar and active stereo-vision. In *Proc. IEEE Intl. Conference on Robotics and Automation*, 2968-2973. Piscataway, N.J.: IEEE.

Yokota, M., A. Mizushima, K. Ishii, and N. Noguchi. 2004. 3-D map generation by a robot tractor equipped with a laser range finder. In *Proc. ATOE Conference*, 374-379. St. Joseph, Mich.: ASAE.