Stereo-based path and obstacles detection in multiannual cultivations for autonomous navigation of an agricultural robot

Adrian Orellana[1], David Fernández-Llorca [2], Carlos Soria [1] and Alejandro G. Lorente [2]

***Abstract*— The intelligent agriculture is an area in hard and continuous development. This paper presents a new development carried out over an autonomous robot for agriculture environments to improve stereo vision processing. The automatic calibration system has been developed to obtain the path plane and the path limits without the need of interaction with a human operator for a stereo camera mounted on the robot, with a pitch and roll not precisely defined previously. Then, the camera itself has been previously calibrated in order to be able to give good 3D information referred to the camera coordinates, but not to the world coordinates, depending on the vehicle position and the camera pitch and roll at every instant. The automatic calibration, working with the information the camera collects from the environment, detects the path plane. From that information, generates an elevation map and from it detects the path limits, helping the navigation system to adjust the future position of the robot if necessary.**

I. INTRODUCTION

The main goal of precision agriculture is to optimize field management, increase agricultural efficiency and sustainability and reduce the operating costs and the ecological footprint, by means of new technological developments. Agricultural robots play a key role in the context of precision agriculture applications including autonomous navigation, data collection and automatic actuation capabilities. Data collection can be carried out using aerial or ground level vehicles. Aerial imagery and satellite data usually lack the necessary spatial resolution, and their acquisition systems strongly depends on weather conditions. At the ground level, data collection can be accomplished by sampling on foot, using mobile platforms manually driven or by means of autonomous robots. Sampling on foot is a highly time-consuming task that provides very discrete data. Data gathering using a mobile platform (tractor or vehicle) only requires one human operator and enables continuous sampling. On the other hand, autonomous agricultural robots does not require manual intervention providing data over the entire sample area.

Agricultural robots that perform autonomous navigation tasks require precise knowledge of its environment. In order to move autonomously, a mobile robot needs to be aware of its surroundings. The planning of the local trajectories requires to robustly segment the free space in front of the vehicle to manage both navigation and obstacle avoidance tasks. Opposite to road vehicle conditions where a considerable number of free space computation studies have been proposed working with planar surfaces (the road), field environment remains as one of the most complex and challenging environment due to the irregularities of the ground.

In this paper we present a stereo vision-based free space detection system on board of an autonomous robot, including a calibration method to obtain the relative position of the stereo sensor with respect to the ground plane. The robot is an all-terrain vehicle (ATV) quad that has been automated and equipped with different sensors (see Fig. 1). The ATV is especially designed to accomplish autonomous missions, such as automatic estimation of leaves surface area of plants, fruit ripening estimation and weed detection and control, among others. The type of scenarios where the robot is operating are vineyards and olive tree plantations as can be observed in Figs. 2(a) and 2(b).

Fig. 1. Automated All-Terrain-Vehicle (ATV) quad equipped with different sensors.

In the remainder of the paper, we review free space detection technologies in the context of precision agriculture applications in Section II. An overview of the robot hardware is provided in Section III. Section IV describes both the extrinsic calibration and the free space detection approaches. Some results are provided in Section V. Conclusions and Future works are finally presented in Section VI.

II. STATE OF THE ART

According to [1], the long time wish of endowing agricultural robots with an increasing degree of autonomy is becoming a reality thanks to two crucial factors: the broad diffusion of GPS and the inexorable progress of sensors,

[1]Adrian Orellana and Carlos Soria are with the Faculty of Engineering, Universidad de San Juan, Av. Libertador San Martin(o) 1109, 5400 San Juan, Argentina orellana,csoria@inaut.unsj.edu.ar

[2]David Fernández-Llorca and Alejandro G. Lorente are with the Computer Engineering Department, Polytechnic School, University of Alcalá, Pza. San Diego, s/n, 28801 Alcalá de Henares, Madrid, Spain llorca@aut.uah.es

ISBN: 978-84-695-3472-4

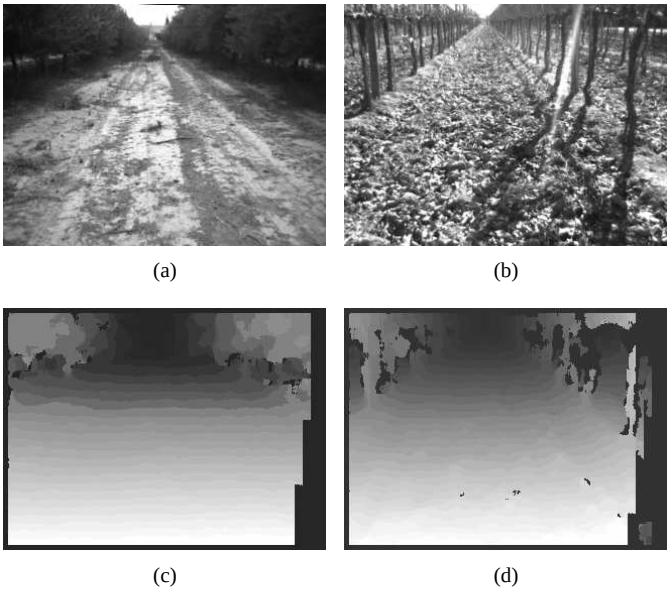


Fig. 2. Typical scenarios and dense disparity images corresponding to (a),(c) olive tree plantations and (b),(d) vineyards.

computers and electronics. The majority of sensors used in agricultural vehicles have been related to autonomous navigation, including mechanical feelers, computer vision cameras, GPS, geomagnetic direction sensors, laser scanners, and ultrasonic range-finders in Japan [2], North America [2] and Europe [3].

Vision systems, and more specifically stereo vision sensors, seem to be excellent candidates for autonomous navigation and obstacle detection [4], [5], [6], [7]. Laser scanners are direct competitors of stereo for this task and they have been widely studied in the literature. Although these sensors provide very accurate range data, they are usually limited to 2D scanning [8], [9]. More sophisticated 3D laser scanners sensors such as Velodyne64 or Riegl VZ-400 have not been still used in the context of precision agriculture applications, mainly due the associated costs. Stereo vision sensors provide a real-time 3D dense representation of the environment at a reasonable cost, with enough accuracy, specially for agricultural robots, since they are low-speed vehicles. Monocular approaches for guidance of agricultural vehicles are mainly based on the Hough transform to detect lines, path boundaries, plant rows, etc. [10], [11], [12]. These approaches are not designed to accomplish either obstacles detection or free-space computation, so they cannot react to unexpected situations.

Stereo sensors have been successfully used to build 3D terrain maps by combining the information provided by the stereo camera with a localization sensor and an inertial measurement unit for precision agriculture applications [13]. In the context of autonomous guidance of agricultural vehicles, stereo vision has been used for crop row detection by searching for optimal navigation points from three-dimensional crop elevation maps [14]. In both cases [13], [14], the extrinsic relationship between the stereo sensor and the ground plane is obtained by a manual measuring process.

Most of the studies concerned with stereo-based free space computation for autonomous guidance of vehicles have been proposed in the context of intelligent road vehicles. In [15], free space problem is addressed as a dynamic programming task building a stochastic occupancy grid. Stereo measurements are integrated over time reducing disparity uncertainty and they are entered into an occupancy grid, taking into account the noise properties of the measurements. This procedure has been integrated with a B-spline road surface modelling method to improve the free space detection in complex road environments [16]. Elevation maps were used for road surface, traffic isle and obstacle detection for intelligent vehicles applications in [17]. However, all these approaches may not be directly applied for free space detection in the context of agricultural vehicles, since the ground surface is not as smooth as the road surface.

III. HARDWARE DESCRIPTION

The system, schematically described in Fig. 3, consists of a pre-calibrated stereo camera (Bumblebee2, PointGrey) and a LIDAR (LMS211, SICK) that are controlled by the High Level CPU of the vehicle, and a DGPS, commanded by the Middle Level CPU. Both CPUs, via the CAN BUS of the vehicle, are interconnected between themselves and connected to the modules: steer control, throttle control, break control, speed sensor, Inertial Unit (IMU) and indeed the bus is open to future expansions. The bus functionality is fully described in [18].

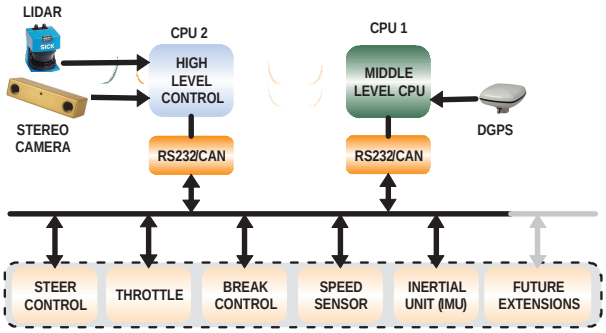


Fig. 3. Global overview of the logic scheme of the autonomous robot.

IV. FREE-SPACE COMPUTATION

A. Unsupervised Extrinsic Calibration

The free space computation module is based on the analysis of digital elevation maps [17]. Elevation maps have to be referenced to the world coordinates (ground plane), so a rigid transformation between the 3D point from the camera to the ground plane has to be performed (see Fig. 4). This transformation includes a rotation and a displacement. We propose an unsupervised method to automatically compute the extrinsic relationship between the camera and the ground plane (plane of the vehicle path).

The proposed method divides the disparity image into a grid of 7×10 cells (see Fig. 5(a) in front of the robot. The

ISBN: 978-84-695-3472-4

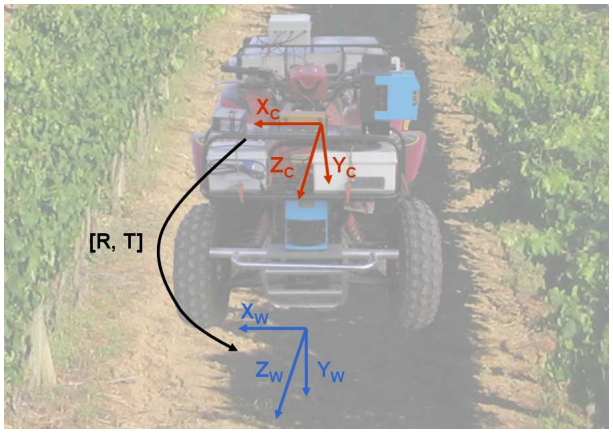


Fig. 4. Rigid transformation between camera and ground plane (world).

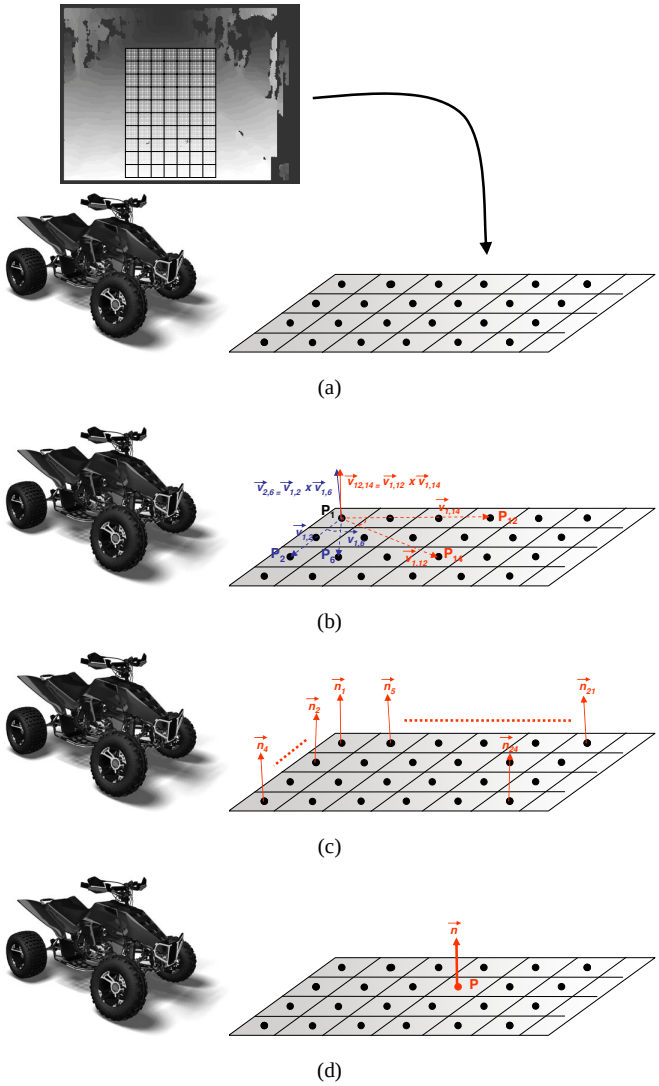


Fig. 5. (a) Grid with 7×10 cells and the 3D representative points of each cell; (b) Vector cloud and cross products; (c) Resulting normal vector for each cell; (d) Final point and director vector of the path plane.

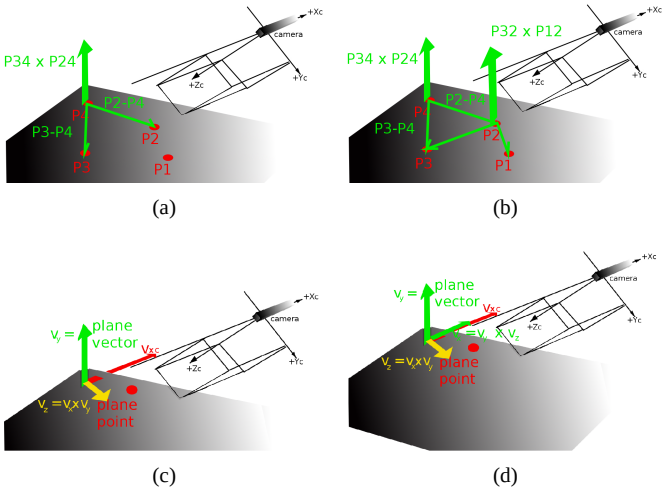


Fig. 6. World axis determination: (a) Cross product to get a plane director candidate; (b) Another candidate computed; (c) The vector $\bar{v}_z$ is obtained as the cross product between the plane director ($\bar{v}_y$) and the x coordinate vector $\bar{v}_{xc}$: $\bar{v}_z = \bar{v}_y \times \bar{v}_{xc}$, belonging $\bar{v}_z$ to the path plane; (d) The vector $\bar{v}_x$ is obtained as the cross product between the plane director ($\bar{v}_y$) and $\bar{v}_x$: $\bar{v}_x = \bar{v}_y \times \bar{v}_z$, also belonging $\bar{v}_x$ to the path plane.

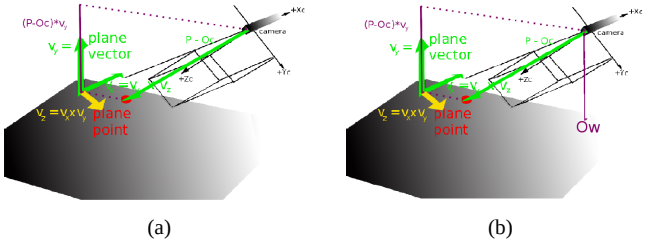


Fig. 7. World origin determination: (a)Projection of the vector representative of the difference between the camera origin and the point of the plane over the vector (v_y); (b) The projection is located at the camera origin position, determining the world origin coordinates at its end.

size of the grid was selected after a trial and error test over a set of 100 images of the working area, to be able to generate the desired information keeping the computational load as low as possible. The 3D points of each cell are computed and sorted according to their Y coordinate. Then, a median filter with respect to Y coordinate is applied obtaining a representative 3D point $\tilde{P}_i$ for each cell of the grid (Fig. 5(a)). A vector cloud is then generated $\bar{v}_{i,j}$, using the mentioned 3D representative points of each cell as vector ends. The use of a median filter is based on the fact that in many cases, the robot can be placed facing toward zones that are not part of the path and, despite more than 50% of the area near the vehicle is actually part of the path, the media of the obtained candidates can not be properly associated with the path plane director vector. A cross product and a further normalization is performed on each pair of those vectors $|\bar{v}_w| = |\bar{v}_j \times \bar{v}_k|$, being $j \neq k$ (see Fig. 5(b)), which gives a set of free normalized vectors $|\bar{n}_w|$ (not necessary associated to a cell). The median with respect to the Y coordinate of the normalized vectors (Fig. 5(c)) is computed using the same method of sorting and choosing the middle vector. The

ISBN: 978-84-695-3472-4

normalized vector is considered the director of the path plane $\bar{n}$ (see Figure 5(d)). Then the selected 3D points are also sorted according to their Y coordinate and the median is selected as the plane point P. With that point P and the director vector $\bar{n}$, the path plane is completely specified for the limited zone near the robot. A detail of the procedure of performing the cross products to obtain the vector director candidates of the plane is depicted in Figs. 6(a) and 6(b) from the cell points P1, P2, P3 and P4. As a consequence of applying this method to get the world coordinates system, the vectors $\bar{v}_x$ and $\bar{v}_z$ are always going to belong to the path plane, despite of pitch and roll of the camera system respect to the path plane. A small difference between the directions of $\bar{v}_{xc}$ and $\bar{v}_x$ is noticed in Fig. 6(b) (the vector $\bar{v}_{xc}$ is entering to the plane).

Once the path plane has been determined, a world reference system is obtained as follows: the unitary axis vectors are: $\bar{v}_y = \overrightarrow{plane.vector}$, the remainder two vectors must be determined belonging to the plane. The vector $\bar{v}_z$, is computed as the cross product of $\bar{v}_y$ already determined and $\bar{v}_{xc}$ of the camera's coordinates because $\bar{v}_z$, computed this way, is orthogonal to both $\bar{v}_y$ and $\bar{v}_{xc}$, then $\bar{v}_z$ belongs to the plane, a further normalization is performed to get a unitary vector $\bar{v}_z = \bar{v}_z/|\bar{v}_z|$. Finally, $\bar{v}_x$ of the world coordinates system is obtained as the cross product between the previously obtained $\bar{v}_y and \bar{v}_z$ vectors as: $\bar{v}_x = \bar{v}_y \times \bar{v}_z$. The origin of coordinates of the world is obtained as the projection of the camera origin on the path plane. Remembering that P belongs to the plane, $proj = \overline{O_c - P} \bullet \bar{v}_y$ (see Fig. 7(a)), is the projection of the vector difference between the camera origin and a point of the plane over the director of the plane. The origin of the world coordinates system is then that projection times the director vector $O_w = proj * \bar{v}_y$ (see Fig. 7(b)).

With this basis, the inverse transformation is evaluated inverting the matrix formed by the unitary vectors $\bar{v}_x, \bar{v}_y$ and $\bar{v}_z$ as rows, and the shift between both references is considered to be the transformation of the camera origin to the world origin. With this transformation defined, the elevation map is computed for every single 3D camera point, roto-translating those 3D points to the world coordinates, being the y_w coordinate the elevation of every 3D point, where O_x, O_y, O_z are the camera coordinates of the world origin, or the world origin referred in the camera coordinates.

$$\begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} \left(\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} - \begin{bmatrix} O_x \\ O_y \\ O_z \end{bmatrix} \right) \quad (1)$$

B. Digital elevation maps and path detection

To build the elevation map, minimum and maximum values of x_w, y_w, z_w are evaluated and the greatest range of data on x_w and z_w determines the grid size in meters for the elevation map. The world 3D points are being allocated on each cells of the elevation map in a way they determine the maximum and minimum value of each cell. The cells are then labeled according to the elevation of their points. We use a color code assigning a red color to the highest elevation cells ($> 1m$), green to the intermediate elevation cells ($< 1m$ and $> 10cm$) and blue to the lower elevation cells ($< 10cm$).

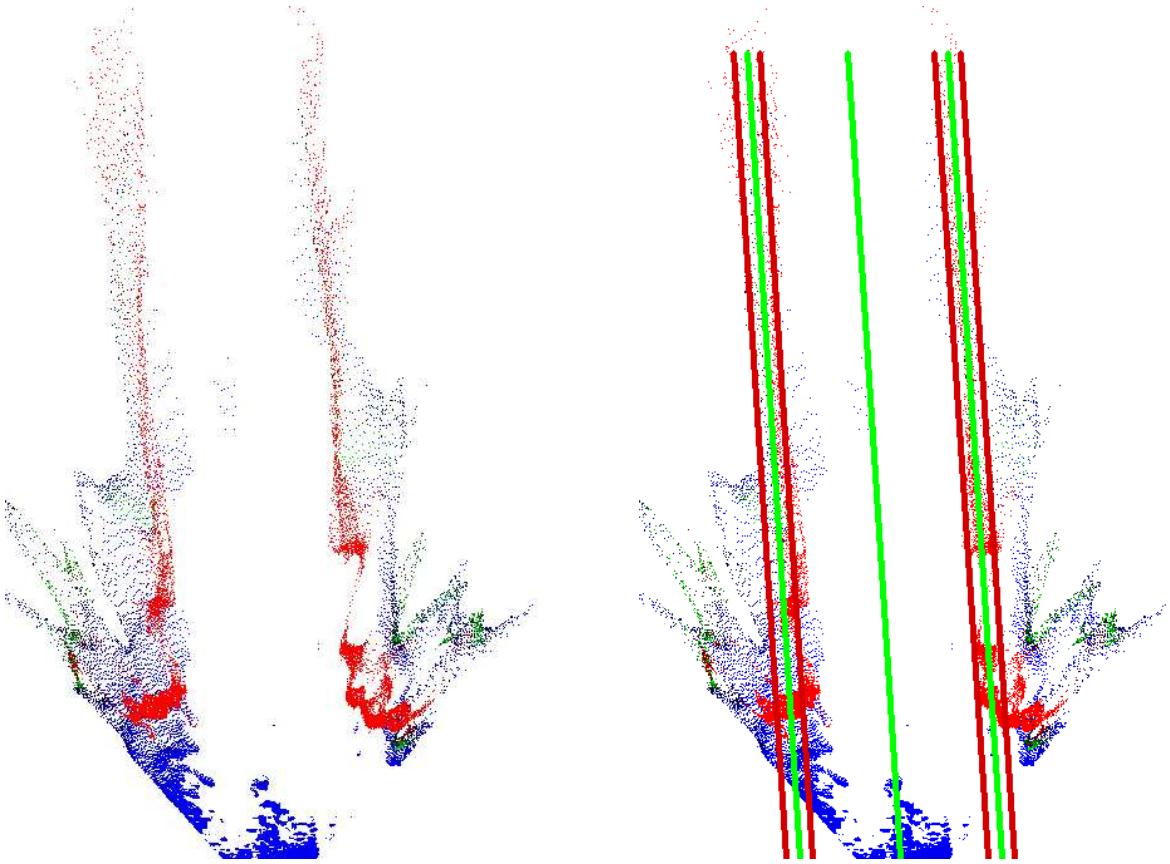
An example of the digital elevation map can be observed in Fig. 8(a). To evaluate the information of the path on the elevation map, lower elevation cells are not marked. To get the limit lines of the path, a sub-sampled elevation map is considered. To search for the limit lines we take two thin rectangles, their orientation and displacement is varied in order to search every possibility of location of the rectangles into the sub-sampled elevation map. The restriction is a minimum and maximum distance between rectangles, remembering that the path limits are being searched. On each position, the quantity of elevation marks are counted. There exits a minimum limit fo each rectangle and for the sum of both of them to be considered valid candidates to be the path limits. The pair with a maximum is considered representing the path limits. After that, the obtained maximum angle and position is translated to the original elevation map, as can be seen in Fig. 8(b). With this information, the center line of the path is obtained as the media between the limit lines and is considered the right position and orientation for the robot. That line is characterized by a point and an unitary director vector: $(P, \bar{l})$.

To get the actual position of the robot, we know that in camera coordinates, the robot is located at $p_{1c} = (0, 0, 0)$. Then, a transformation of those camera coordinates to world coordinates is performed using Eq. 1 to get p_{1w}. The orientation of the vehicle is determined by considering that the vehicle is facing towards $+z_c$ axis. Then, choosing a point as $p_{2c} = (0, 0, 1.0)$ will give that direction, and using again Eq. 1, the world coordinates of that point are obtained p_{2w}. The pose of the robot is then determined by the first point p_{1w} and its direction by the difference vector of the second point minus the first one as $\bar{d} = (\bar{p}_{2w} - \bar{p}_{1w})$, and then, the direction is normalized. Knowing that the center line of the path lies on the path plane, its y coordinate is 0. A unitary vector orthogonal to the line director $\bar{l} = (l_x, 0, l_z)$ is obtained as $\bar{r} = (-l_z, 0, l_x)$, since $\bar{r} \bullet \bar{l} = 0$. The distance of the robot position to the center line of the path is evaluated as $distance = \bar{r} \bullet \overrightarrow{(P - p_{1w})}$, and the correction angle is evaluated as $\phi = cos^{-1}((\bar{l} \bullet \bar{d})/(|\bar{l}| \bullet |\bar{d}|))$. The position and angle correction are used by the trajectory planning module.

V. RESULTS

The proposed approach runs at 1Hz in a double core 2.0 GHz CPU with 320×240 pixel size images, and it is integrated in the vehicle control and data storage systems. The longitudinal vehicle controller is defined to follow a linear speed reference between 1 and 3 m/s (3.6 and 10.8 km/h), so the sensor frequency is enough for autonomous maneuvers. The stereo-based free space and path detection method automatically provides guidance information that is used by the guidance algorithm, including the digital elevation maps, the lateral position error, orientation error, and pitch and roll variations. Specifically, the information of the

ISBN: 978-84-695-3472-4



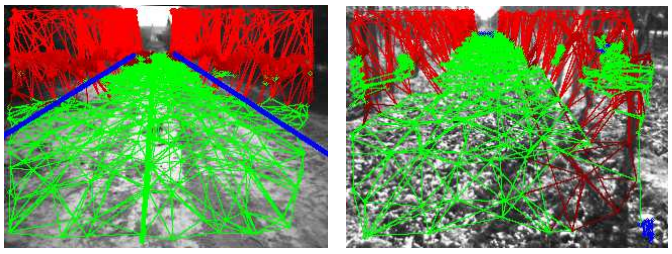
(a) (b)

Fig. 8. (a) Color-coded digital elevation map; (b) Path detection results.

vehicle is maintained accurate enough with an altitude error lower than $0.05m$ within a range of $5m$. This information is renewed over the general elevation map every $5m$ (between 5 and 15 iterations according to our sample frequency). The lateral position error is $\pm 0.25m$ in the x_c direction. The orientation error obtained is ± 0.1 radians.

Projecting the information provided by the digital elevation maps, with the color-code previously described, augmented reality images are generated including the elevation information, the detected path boundaries (blue lines) and the center line of the path (the lateral reference that is passed to the lateral controller) which is drawn in green. Two examples of the augmented reality images are depicted in Figures 9(a) and 9(b) for the olive tree plantation and vineyards respectively. In these pictures, the information of the cells of the elevation map is retro projected over the rectified image of the left camera. On each cell, a dot of the color representative of its elevation is draw. Then, the dots with highest elevation determine the color of segments draw between neighbour dots and generate the augmented reality picture.

The camera/robot orientation with respect to the path center line is showed in Fig. 10 using the digital elevation map representation. In addition, the laser measurements corresponding to the horizontal plane with height around 60cm, are projected onto the same map. As can be observed, the 3D point cloud provided by the laser, which has much better accuracy in terms of depth measurements with respect to the camera, correctly matches with the stereo-based 3D point cloud corresponding to obstacles.



(a) (b)

Fig. 9. Augmented reality image with path boundaries of the (a) olive tree plantations and (b) the vineyards.

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, a stereo-based free space and path detection system has been presented for an agricultural robot (ATV/quad vehicle) in the context of precision agriculture applications. The ATV vehicle was previously equipped with controllers, actuators and trajectory planning modules to

ISBN: 978-84-695-3472-4

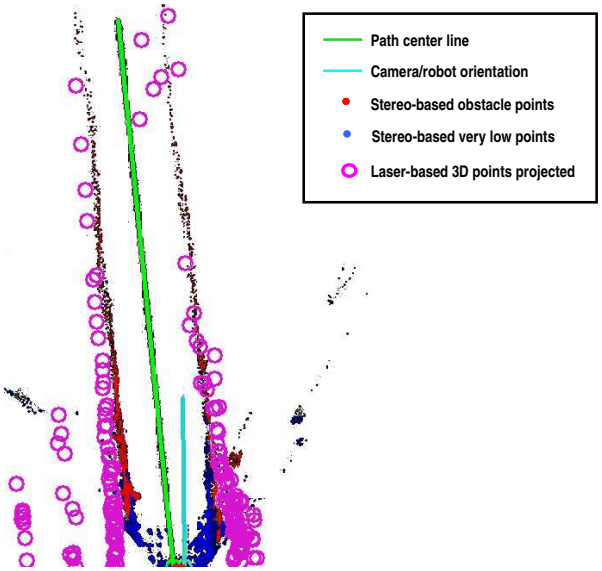


Fig. 10. Digital elevation map, including camera/robot orientation, center line of the detected path and projection of the 3D point cloud corresponding to the laser.

accomplish autonomous navigation maneuvers. It is equipped with several sensors, including a stereo vision system able to provide a dense stereo representation of the environment. An automatic camera/path plane calibration system is proposed to transform the 3D points provided by the camera to the world which is placed at the path plane. Thus, digital elevation maps can be obtained, referencing the Y coordinate of each 3D point with its corresponding altitude. The elevation maps are analyzed to extract the lateral limits of the path, which are usually defined by the plantation, and the center line is finally computed and passed to the lateral controller system as the reference. The stereo system provides the lateral error as well as the orientation error. These variables will be used in further experiments to perform autonomous navigation including trajectory tracking and collision avoidance maneuvers.

The automatic location of the path plane with respect to the camera, includes variations of pitch and roll, which are especially important for agricultural robots due to the nature of the ground plane which is usually far from being a smooth surface. These values will be compared with those provided by the IMU, which is already installed in the ATV, to evaluate the accuracy of the proposed approach. In addition, the ATV is equipped with a laser that provide 3D point cloud corresponding to a horizontal plane. This information is much more accurate than the 3D data given by the stereo system. However, laser data is only related with an horizontal plane. Fusion of both sensor will be a future research topic, including map-matching techniques to obtain accurate measurements of the robot motion. Finally, the 3D point cloud will be integrated and globally referenced to create 3D terrain maps that may be further processed for agriculture applications.

VII. ACKNOWLEDGMENTS

This work was supported by the Universidad de Alcalá (Spain), the Universidad Nacional de San Juan (Argentina) and the SPU of Argentina. This work was partially financed by the Spanish Ministry of Economy and Competitiveness under Research Grant ONDA-FP TRA2011-27712-C02-02.

REFERENCES

[1] F. Rovira-Más, "Sensor architecture and task classification for agricultural vehicles and environments," *Sensors*, vol. 10, pp. 11 226–11 247, 2010.

[2] T. Torii, "Research in autonomous agriculture vehicles in japan," *Computers and Electronics in Agriculture*, vol. 25, pp. 133–153, 2000.

[3] R. Keicher and H. Seufert, "Automatic guidance for agricultural vehicles in europe," *Computers and Electronics in Agriculture*, vol. 25, pp. 169–194, 2000.

[4] C. Caraffi, S. Cattani, and P. Grisleri, "Off-road path and obstacle detection using decision networks and stereo vision." *IEEE Transactions on Intelligent Transportation Systems,*, vol. 8, no. 4, pp. 607–618, 2007.

[5] D. F. Llorca, M. A. Sotelo, I. Parra, M. Ocaña, and L. M. Bergasa, "Error analysis in a stereo vision-based pedestrian detection sensor for collision avoidance applications," *Sensors*, vol. 10, no. 4, pp. 3741–3758, 2010.

[6] D. F. Llorca, V. Milanés, I. P. Alonso, M. Gavilán, I. G. Daza, J. Pérez, and M. A. Sotelo, "Autonomous pedestrian collision avoidance using a fuzzy steering controller." *IEEE Transactions on Intelligent Transportation Systems,*, vol. 12, no. 2, pp. 390–401, 2011.

[7] V. Milanés, D. F. Llorca, J. Villagrá, J. Pérez, C. Fernández, I. Parra, C. González, and M. A. Sotelo, "Intelligent automatic overtaking system using vision for vehicle detection." *Expert Systems with Applications*, vol. 39, no. 3, pp. 3362–3373, 2012.

[8] T. Chateau, C. Debain, F. Collange, L. Trassoudaine, and J. Alizon, "Automatic guidance of agricultural vehicles using a laser sensor," *Computers and Electronics in Agriculture*, vol. 28, pp. 243–257, 2000.

[9] V. Subramanian, T. F. Burks, and A. A. Arroyo, "Development of machine vision and laser radar based autonomous vehicle guidance systems for citrus grove navigation," *Computers and Electronics in Agriculture*, vol. 53, pp. 130–143, 2006.

[10] B. Astrand and A.-J. Baerveldt, "A vision based row-following system for agricultural field machinery," *Mechatronics*, vol. 15, pp. 251–269, 2005.

[11] C. M. Soria, A. Orellana, and R. Carelli, "Sistema de visión estéreo para un vehículo autónomo todo terreno," in *VI Jornadas Argentinas de Robótica JAR2010*, 2010, pp. 1–6.

[12] G. Wu, Y. Zheng, and S. Wang, "Walking goal line detection based on dm6437 on harvesting robot," *Computer and Computing Technologies in Agriculture V. IFIP Advances in Information and Communication Technology*, vol. 370, pp. 351–361, 2012.

[13] F. Rovira-Más, Q. Zhang, and J. F. Reid, "Stereo vision three-dimensional terrain maps for precision agriculture," *Computers and Electronics in Agriculture*, vol. 60, pp. 133–143, 2008.

[14] M. Kise, Q. Zhang, and F. Rovira-Más, "A stereovision-based crop row detection method for tractor-automated guidance," *Biosystems Engineering*, vol. 90, no. 4, pp. 357–367, 2005.

[15] H. Badino, U. Franke, and R. Mester, "Free space computation using stochastic occupancy grids and dynamic programming," in *Programming, Proc. Intl Conf. Computer Vision, Workshop Dynamical Vision*, 2007.

[16] A. Wedel, H. Badino, C. Rabe, H. Loose, U. Franke, and D. Cremers, "B-spline modeling of road surfaces with an application to free-space estimation." *IEEE Transactions on Intelligent Transportation Systems,*, vol. 10, no. 4, pp. 572–583, 2009.

[17] F. Oniga and S. Nedevschi, "Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1172–1182, 2010.

[18] C. M. Soria, M. Muñoz, R. Carelli, and J. M. Sebastián, "Control of an autonomous all terrain vehicle," in *8th Workshop. Robots de Exteriores. Robocity 2030 II*, 2010, pp. 115–126.