IEEE Sensors Council

Augmented Perception for Agricultural Robots Navigation

Francisco Rovira-Más, Verónica Saiz-Rubio, and Andrés Cuenca-Cuenca

Abstract—Producing food in a sustainable way is becoming very challenging today due to the lack of skilled labor, the unaffordable costs of labor when available, and the limited returns for growers as a result of low produce prices demanded by big supermarket chains in contrast to ever-increasing costs of inputs such as fuel, chemicals, seeds, or water. Robotics emerges as a technological advance that can counterweight some of these challenges, mainly in industrialized countries. However, the deployment of autonomous machines in open environments exposed to uncertainty and harsh ambient conditions poses an important defiance to reliability and safety. Consequently, a deep parametrization of the working environment in real time is necessary to achieve autonomous navigation. This article proposes a navigation strategy for guiding a robot along vineyard rows for field monitoring. Given that global positioning cannot be granted permanently in any vineyard, the strategy is based on local perception, and results from fusing three complementary technologies: 3D vision, lidar, and ultrasonics. Several perception-based navigation algorithms were developed between 2015 and 2019. After their comparison in real environments and conditions, results showed that the augmented perception derived from combining these three technologies provides a consistent basis for outlining the intelligent behavior of agricultural robots operating within orchards.

Index Terms—3D Vision, field robotics, autonomous navigation, digital farming, local perception, sensor fusion.
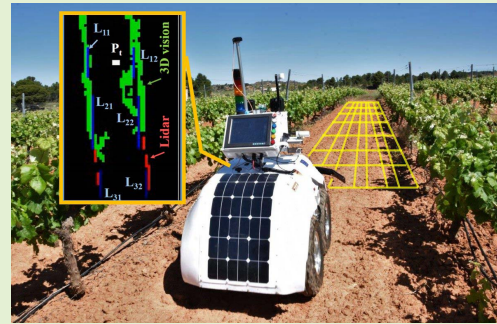
I. INTRODUCTION

THE turn of the 21st century coincided with the appearance of off-the-shelf commercial stereo cameras, which made 3D perception accessible to many on-vehicle and outdoors applications due to their compactness, easy connectivity, and reasonably fast correlation algorithms that solved the stereo matching in real time. Previous attempts [1] showed the great potential of 3D perception in general, and stereo vision in particular, but had required bulky rigs where physically keeping the stereo geometry of binocular assemblies, developing their own matching algorithms, and finding capable computers to make calculations fast, practically discouraged any chance to work outdoors from moving vehicles. The advent of compact stereo cameras, in combination to the availability of more powerful processors, however, changed

such landscape. The Census algorithm [2], for example, offered a reliable correlation software that generated 3D point clouds for images of common resolution (320×240) in real time from any standard laptop. Solutions like this opened a wide range of applications for agriculture, beginning with pioneering experiences on autonomous navigation of tractors for fields structured in crop rows, back in 2004, by analyzing the morphology of disparity images [3], the detection of obstacles for safeguarding in 2005 [4], and the creation of 3D crop maps from terrain [5] and aerial [6] vehicles. This innovative work led to the concept of 3D density for the real-time analysis of three-dimensional point clouds obtained with compact stereoscopic cameras [7]. This stereo analysis based on density grids made of regular cells has been the core algorithm of the safety system developed for an autonomous rice harvester. The machine was guided with a multi-GNSS receiver and a GPS compass, but the 3D perception algorithm for detecting obstacles, in particular people standing in paddy fields, was proved to be efficient, except for a blind zone in the close vicinity of the camera where stereo matching was not possible [8]. Time of flight (TOF) sensors offer a promising alternative to stereo vision, where a matrix of infrared beams produce 3D point clouds of a scene. In comparison, stereo vision provides –at present– more resolution (number of pixels or points), but TOF sensors are active sensor with their own

illumination source, and therefore they are independent of natural illumination, working both during the day and at night. In an experiment conducted in a laboratory setting, a 3D Kinect sensor was used to operate a robotic manipulator to sample leaves, using a 3D occupancy grid to find collision-free paths [9]. Even though 3D perception provides a faithful reconstruction of the surrounding environment as a result of a point cloud, where each point can be well determined by its three Cartesian coordinates (x, y, z), monocular vision is also capable of retrieving the visual cues needed to assist an autonomous vehicle. Such an assistance was implemented to control deviations from pre-planned paths for a tractor guided by combining an RTK-GPS and monocular vision. The camera produced fine-tuning corrections to achieve a more precise steering, in addition to contribute to obstacle detection for safeguarding by using color bands and texture analysis [10].

The development of GPS-based autonomous guidance for agricultural equipment operating within commodity crop fields, typically tractors, harvesters and self-propelled sprayers, was intense after year 2000 when GPS selective availability was cancelled by the US Department of Defense for free civilian use. However, the situation for orchards and groves was the opposite due to the uncertain signal visibility caused by dense –sometimes tall– canopies. For this situation, the aid supplied by local perception sensors is crucial. An alternative to machine vision, mostly when stereo vision still resulted computationally expensive, was represented by laser rangefinders, as the row-following algorithm based on the Hough Transform to guide a utility vehicle equipped with two laser rangefinders located at its front corners [11]. As important as the sensing devices becomes the processing algorithms to convert point sets into meaningful steering commands. A single 2D laser scanner was mounted on a commercial robotic platform to compare navigation algorithms in an apple orchard. The experiment showed that a particle filter produced better results than the Kalman filter [12]. Even though commercial laser rangefinders, also known as lidars, are quite accurate in their range measurements, the fact that a unique beam needs to sweep the space ahead of the vehicle poses some challenges for off-road conditions where shocks and vibrations are permanent. This problem has been circumvented by using a set of several rangefinders, which makes the solution bulkier and currently too costly for orchard equipment. A row following system based on fixed laser scanners and wheel encoders was implemented in various robotic vehicles under the Comprehensive Automation for Specialty Crops project (CASC), which evidenced the actual benefits inherent to this sensing technology [13]. The vehicle that won the DARPA Grand Challenge in 2005, the robot Stanley, featured five lidars and six processor computing platforms, which offered a solution for an off-road environment very different from orchard settings, where predefined structures exist as trees follow ordered rows, and the traveling speeds are low in comparison to the average velocity reached by Stanley, about 33 km/h [14].

Both lidar and machine vision solutions mentioned above do not suffice, independently by themselves, to constitute a general framework that solves the problem of autonomous navigation inside orchards arranged by equidistant rows. Therefore, a combination of diverse technologies actuating synergically leads to results with higher robustness, something that was detected early on [15] and still continues being applicable [16]. Although an orderly array of quasi-parallel rows seems, a priori, an affordable task, the challenge is immense. There are no two equal rows, and even repeating the same rows is usually different depending on the wind, soil conditions, and the always-changing illumination pattern. As a result, redundancy and sensor fusion need being the norm for orchard navigation with robots. In addition to imaging sensors and lidars, ultrasonic devices also provide ranging information for short-range distances, as those measured in the tight surroundings encountered when performing headland turns [17]. The benefits brought by local perception sensors to the automation of agricultural vehicles has made equipment manufacturers start considering them for commercial solutions. Claas, for instance, has introduced a color stereo camera for steering implements in fields with structures that can be identified by colorimetric, textual, or height information [18]. Similarly, John Deere presented a concept tractor equipped with a stereovision binocular camera in the tradeshow Agritechnica in 2019. The robot Bakus, on the other hand, integrates eight time-of-flight sensors that cover the entire vicinity of the vineyard robot during the day and at night [19]. The interest of industry in robotics for agriculture, and the fact that perception sensors are being considered in various assemblies and solutions to grant stability in vehicle automation, indicate the relevance of finding reliable perception solutions for automating agricultural equipment.

II. 3D PERCEPTION FOR OPEN ENVIRONMENTS: CHALLENGES AND SOLUTIONS IN AGRICULTURE

At the dawn of artificial intelligence, it was soon made evident that developing a General Problem Solver was not the way to go [20], as even though a computer program was capable of simulating human behavior in a first approximation, it did it successfully only in a narrow domain. Well-determined problems, such as chess or checkers, were attainable with highly focused algorithms, but a complete understanding of the problem to solve was indispensable. Likewise, agricultural environments are too diverse and complex for attempting a common solution within robotics. However, a set of requirements unambiguously defined is essential for deploying an autonomous robot in open environments. Requirements for covering many vineyards in Europe can be a row spacing between 1.5 m and 3 m, canopies structured in vertical trellises, and slopes below a certain angle, say 15°. For the case of the vineyard robots considered for this study, the problem of navigation is split into two independent modes: inside-row guidance and headland turning. Given that the scope of this research considers perception-based navigation rather than solutions based on GNSS positioning, the structure of the environment is key. Unlike planetary and military rovers that also traverse off-road terrains, agricultural vehicles are subjected to regular structures except for the case of farming barren fields. However, the diversity of environments requires dealing with the specificity of each particular situation, such

Fig. 1. Layouts for crop row structures (a) and orchard row structures (b).

Fig. 2. Effects of ambient temperature on a stereo camera: (a) Temperature below 30°; (b) Temperature above 40° C causing the loss of the red color component in the RGB CMOS sensor.

that a successful solution begins by successfully understanding and characterizing the surrounding environment. In general, the morphology of agricultural fields can be classified as crop rows (Fig. 1-a) or orchard rows (Fig. 1-b). Although they may seem equivalent from an aerial perspective, the challenges posed for autonomous navigation of ground vehicles are totally different, as well as their perceptive features. Several crop rows are typically tracked simultaneously, whereas only the two bounding (left-right) rows are visible for orchard guidance. In addition, orchards create higher risks as vehicles tend to be squeezed between thick canopies, with little room for steering corrections, and with likely chances of damaging valuable assets. Crop row guidance greatly benefits from GNSS solutions, as large equipment usually places antennas several meters above crops. In orchard layouts, however, satellite signals are many times blocked or reflected (multi-path errors) by large canopies over medium-size machinery; for these cases, perception-based navigation and safeguarding is essential. Fig. 1 illustrates the structural differences between crop rows (soybeans) and orchard rows (cherry trees).

A fundamental precept for perception-based navigation is the presence of features from which extract guidance and safety commands. Furthermore, not only the presence but also the properties of surrounding objects have a strong influence on the performance of these navigation systems. Stereoscopic vision, for instance, strongly relies on the texture of surrounding canopies for the right execution of the correlation algorithm, whereas the reflective properties of vegetation are fundamental for lidars and sonar rangefinders. Apart from the influence exerted by the specific properties of vegetation on the behavior of perception sensors and their capacity to sense the surrounding environment, the fact that agricultural robots must work outdoors in harsh environments, conditions their long-term performance as well as their cost-efficiency opportunities. The temperature and humidity conditions found in many farms are usually ruinous for electronic boards and components, which aggravates by vibration and shocks induced by rough terrain and a not always advantageous suspension system. Military-certified components with protection indices above IP-65 are convenient for these environmental conditions, but unfortunately, they are out of reach for most agronomical solutions in which cost, market competition, and a favorable return of investment are decisive factors. Fig. 2 shows the effects of ambient temperature on a stereoscopic camera used in a Portuguese vineyard that exceeded 40° C. The vision sensors overheated and the red color of the RGB CMOS imager was momentarily lost.

This article focuses on *inside-row guidance* of robots operating in vineyards. A preliminary study of headland turning strategies for guidance in vineyards is available in [17], and the safeguarding algorithm for obstacle detection implemented in the developed robots falls outside the scope of this article, as it operates according to a different logic even though it utilizes the same sensor suite. Each guidance mode – inside-row and headland turning– presents its own challenges for stable navigation and reliable safeguarding. The reader should never be misled by the apparent geometrical simplicity of navigating inside well-determined vertical walls made of leaves and branches. Just the fact that we are dealing with live organisms introduces high doses of uncertainty and the occurrence of special situations. One such case is the presence of large gaps within canopies caused by dead or severely damaged vines. Large gaps create complex situations for navigation algorithms relying on the features of the scenes ahead of the vehicle. When large gaps coincide in both sides simultaneously, the robot might get confused and engage the headland turning routine. Another control challenge may come from the unpredictable response of the suspension system to the terrain, which may depend on the status of the soil (farmed or untilled), its moisture content, and even the tractive capacity for a given battery power. Boundary rows and irregular row ends, where one side is significantly longer than the other, also require additional capacities from the navigation system. A commercial vineyard consists of many rows and the conditions of all of them are normally unknown. There is an abyss between a 10-minute demonstration and a solution that must work for hours, where the simple fact of losing battery power with time typically affects the vehicle dynamics and its navigation accuracy. The following sections describe the strategies for guiding a ground robot within vineyard rows after facing all the challenges mentioned above.

III. AUGMENTED PERCEPTION FOR LOCAL-BASED NAVIGATION

The navigation algorithm for *inside-row guidance mode* relies on the Augmented Perception two-dimensional (2D) Obstacle Map (APOM), which roots in two principles:

a) The APOM is a discrete division of the 3D space surrounding the vehicle and consists of square cells, which may be filled by a set of perception sensors of three diverse working natures and covering ranges.

b) The APOM populating procedure is not expected to be random; occupied cells are supposed to align around two high-density nuclei representing the rows ahead of the vehicle providing the guidance features. Further geometrical derivations will depart from this assumption.

The rationale behind the way augmented perception has been physically articulated in the APOM obeys to two facts: 1) sensor redundancy is necessary because electronic devices are prone to fail or perform under expectation when used repeatedly over long periods of time, mainly outdoors; and 2) perception sensors typically excel for a determined limited area, but lose consistency as targets move away from it; therefore, the complementarity of field of views adds robustness to the solution. Three range levels, in particular, are defined within the APOM:

a) Long ranges zone: pointing ahead between 4 m and 8 m from the sensing head, this zone is instrumental to calculate the target point towards which the robot is directed. It provides mild corrections and smooth navigation.

b) Short ranges zone: pointing ahead below 4 m from the sensing head. This zone produces corrections that are more reactive but it is key to keep the robot at a safety distance from canopies.

c) Close vicinity zone: covers a 2 m ring around the robot, it is highly reactive, and basically exerts safeguarding corrections to re-center the robot or stop it in the presence of interfering obstacles.

The first stage for the implementation of the inside-row guidance algorithm is the creation of the APOM, being the final output the position (x_t, y_t) of the target point P_t in it. With this position, the onboard navigation control system calculates the front-wheel Ackerman angle θ that needs to be steered for the robot to reach P_t. The APOM will be populated with the measurements retrieved from the 11-beam lidar sensor and the 3D stereo camera. The definition and origin of coordinates for both sensor must be the same. Notice that the lidar produces flat coordinates as it senses in a plane, whereas the 3D camera provides the three Cartesian dimensions. The origin of coordinates was located in the symmetrical plane of the robot at ground level, as indicated in Fig. 3, which also illustrates the definition of the Cartesian frame [X, Y, Z]. Let $C_L = \{(x^L,y^L)_1, \ldots, (x^L,y^L)_k\}$ be the set of coordinates retrieved from the lidar sensor, and let $C_V = \{(x^V,y^V,z^V)_1, \ldots, (x^V,y^V,z^V)_m\}$ be the set of point cloud coordinates output by the stereoscopic camera. These sets of coordinates are bounded by the number of beams in the lidar, and by the image resolution set in the camera for every sample obtained at a given cycle time of the central computer running the perception engine. For the particular case of the robot shown in Fig. 3, the number of lidar beams is 11 ($k \leq 11$) and the image resolution of the stereo camera is 640 pixels in the horizontal dimension by 480 pixels in the vertical dimension ($m \leq 640 \times 480$). Fig. 3 also shows the stereo camera located at 1.05 m from the ground and tilted 10° downwards, whereas the lidar beams scan a horizontal plane parallel to the ground at a height of 0.9 m. The perceptual capacity of the sensors
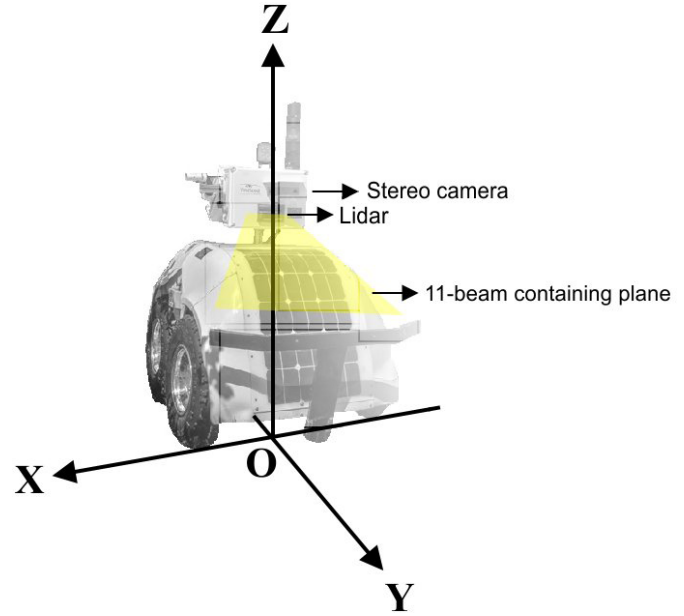

Fig. 3. Definition of coordinate system used in the robotic platform.

depends on their respective specifications, but the calculation of coordinates is always prone to errors. To avoid severe outliers, the elements of C_L and C_V were limited to logic values through the concept of the Validity Box [5], which establishes the logical limits for both types of coordinates and for a given agricultural environment according to (1):

$$P_i (x, y, z) \in V_{BOX} \Leftrightarrow \begin{cases} X_{min} & \leq x \leq X_{max} \\ 0 & \leq y \leq Y_{max} \\ Z_{min} & \leq z \leq Z_{max} \end{cases} \quad (1)$$

Although the APOM has been defined to cover the entire surroundings of the vehicle, the geometrical calculation of the target point P_t for inside-row guidance only uses the perception information of the lidar rangefinder and the 3D stereo camera, both of which can only sense ahead of the vehicle, and therefore cannot yield negative values for the Y axis as stated in (1). The four sonar sensors covering the close vicinity of the vehicle, however, provide emergency corrections and safeguarding commands but do not populate the APOM for the calculation of P_t. The active APOM, therefore, will only consider the positive side of the Y axis. If $\underline{c}$ is the size of the square cell (Fig. 4) in the active grid that represents long and short ranges determined by the lidar and the camera, the dimensions for the grid are bounded by (2). For the vehicle shown in Fig. 3, the active grid has dimensions 50 cells $\times$ 80 cells, being $\underline{c} = 0.1$ m, which implies an area of 5 m $\times$ 8 m = 40 m^2 covered ahead of the robot in the forward direction, as depicted in the schematic of Fig. 4.

$$dim\,(active\,APOM) \begin{cases} X_G = \left\| (X_{max} - X_{min})\,/\underline{c} \right\| \\ Y_G = \left\| Y_{max}/\underline{c} \right\|; \end{cases}$$
$$X_G, Y_G \in \mathbb{N} \quad (2)$$

To fill the active APOM grid, all the coordinates included in the sets C_L and C_V that fall inside the validity box have
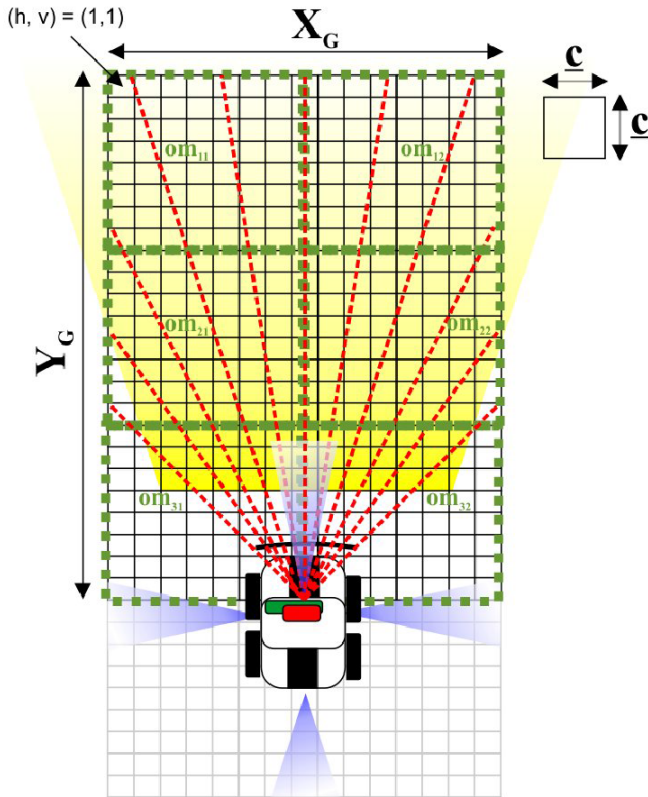
Fig. 4. Layout for the APOM and its active grid zone.

to be discretized. As $dim(C_L) + dim(C_V)$ is usually greater than $(X_G \cdot Y_G)$, it is common to have cells containing various points. The function γ^L is defined to hold the content of a given cell based on lidar readings; when two locations coincide in the same cell, γ^L increases its value in one. The procedure to discretize the coordinate positions of C_L is similar to the calculations of (2), as detailed in (3):

$$\forall \left(x^L, y^L\right)_i \in C_L \rightarrow \begin{cases} L_V = \left\| y^L / \underline{c} \right\| \\ L_H = \left\| (x^L - X_{min}) / \underline{c} \right\|; \end{cases}$$
$$L_V, L_H \in \mathbb{N} \quad (3)$$

The filling of the grid with the data coming from the 3D point cloud retrieved by the stereoscopic camera requires several intermediate steps. The purpose of these steps is the normalization of the content of the cells such that the function γ^V is uniform regardless of the position of the cell, given that closer objects are represented by a larger number of pixels in stereo cameras, due to a horizontal field of view (43° in this case) that expands as the distance from the camera grows [7]. The way to process the 3D point cloud was based on the concept of 3D density [7], by which all the stereo-correlated points of the cloud within the validity box were fit into the cells of the active APOM grid. Notice that 3D points have a z component, and therefore, all points for which $Z_{min} \le z \le Z_{max}$ were enclosed in the same cell according to the following

discretization (4):

$$\forall \left(x^V, y^V, z^V\right)_i \in C_V \rightarrow \begin{cases} V_V = \left\| y^V / \underline{c} \right\| \\ V_H = \left\| (x^V - X_{min}) / \underline{c} \right\|; \end{cases}$$
$$V_V, V_H \in \mathbb{N} \quad (4)$$

The number of points that fall inside a given cell of coordinates (V_H, V_V) is denominated its 3D density, and it is represented as $D(V_H, V_V) \ge 0$. The normalized density D_N that compensates for the loss of resolution in far ranges is given in (5), as proved in [7], and it corrects the density for ranges farther than 3 m while leaving unchanged those cells closer than 3 m from the camera:

$$D_N (V_H, V_V) = \begin{cases} D(V_H, V_V) \cdot \left(\dfrac{V_V \cdot \underline{c}}{3}\right)^2 & V_V \cdot \underline{c} > 3 \\ D(V_H, V_V) & \text{Otherwise} \end{cases} \quad (5)$$

Not all the cells with $D_N > 0$ are representative of occupancy and therefore pointing at actual plant rows, as point clouds are typically corrupted with a small number of noisy outliers. The definition of the γ function for the stereo camera, namely γ^V (6), accounts for scattering noise through the application of a threshold TH that assures that only cells with high density pass to the final composition of the augmented obstacle map APOM. If TH is the threshold to discriminate obstacles from empty space, the definition of γ^V is:

$$\gamma^V (V_H, V_V) = \begin{cases} 1 & D_N(V_H, V_V) > TH \cdot max(D_N) \, ; \, 0 \le TH \le 1 \\ 0 & \text{Otherwise} \end{cases}$$
$$(6)$$

Once the functions γ^L and γ^V have been defined for both sources of perception information, their content can be merged in a unique map that will populate the active grid of the APOM under the fusion function $\Gamma(h, v)$ described in (7). However, for the merging function and augmented map to be coherent, the distance units of $\underline{c}$, validity box boundaries (1), and coordinates of sets C_L and C_V have to be necessarily the same, for example meters; in such case, (L_H, L_V) points at the same cell as (V_H, V_V), and therefore coordinates may be simplified to the notation (h, v) of (7). Let $n = X_G \cdot Y_G$ be the resolution of the active grid, let $\mathcal{L} = \{\gamma^L (1,1), \gamma^L (1,2), \ldots, \gamma^L (X_G, Y_G)\}_n$ indicate the set of cells activated by the lidar rangefinder, let $\mathcal{V} = \{\gamma^V (1,1), \gamma^V (1,2), \ldots, \gamma^V (X_G, Y_G)\}_n$ indicate the set of cells activated by the stereo camera. The active grid holding the information needed for the calculation of the target point P_t is the union of both sets, i. e., APOM $\equiv \mathcal{L} \cup \mathcal{V}$. The diverse nature of the sensors advises for a weighted union in the merging of both sets to yield Γ. In particular, the lidar produces more accurate readings than the stereo camera, but the fact that just a limited number of beams is readily available each cycle, results in an unbalanced filling of the grid, where lidar readings are scarce but very reliable and stereo-based points are numerous but prone to noise. In order to make lidar perception more consistent in the augmented grid, each cell activated by the lidar automatically activated the two cells immediately above and below, as can be seen in the red cells of the grids depicted in Fig. 5. The higher

accuracy and reliability of lidar, however, was mathematically conveyed to the final grid by the introduction of weighting constants K_L and K_V, as shown in the formal definition of Γ given in (7). In the final version of the navigation system, which yielded the grids of Fig. 5, $K_L = 3$ and $K_V = 1$. It is important to keep in mind that the function Γ, that determines the content of the cells in the grid, can only admit nonnegative integers, i. e., natural numbers. In (7), the horizontal position h coincides with V_H and L_H, as well as the vertical position v in the grid is equivalent to positions V_V and L_V calculated in (3) and (4).

$$\Gamma(h, v) = K_V \cdot \gamma^V(V_H, V_V) + K_L \cdot \gamma^L(L_H, L_V);$$
$$\forall \, h, v \subset APOM/h \leq X_G, v \leq Y_G; \Gamma(h, v) \epsilon \mathbb{N} \quad (7)$$

At this point, the active grid of APOM represented in Fig. 4 is populated according to the Γ function of (7). The following operations have the purpose of analyzing the grid to identify the guiding rows as the perceptual features to determine the best position for the target point to which the robot will be guided. The first stage consists of subdividing the grid into six equal operational zones, as outlined in Fig. 4 and mathematically defined in (8) through the occupancy matrix OM. Notice that indices h and v in (8) must be positive integers, and therefore the limits of the summations in (8) have been ideally set at a fraction of grid limits X_G and Y_G, such as $X_G/2$ and $Y_G/3$. However, a limit ideally set at $Y_G/3$, in practice means that one summation will end at $||Y_G/3||$ and the consecutive series will initiate at $||Y_G/3|| + 1$. The components of the occupancy matrix om_{ij} are basically the counting of the occupied cells within each operational zone.

$$OM \overset{\text{def}}{=} \left\{ \begin{array}{l} om_{11} = \sum_{h=1}^{\frac{X_G}{2}} \sum_{v=1}^{\frac{Y_G}{3}} \Gamma(h, v) \\ om_{12} = \sum_{h=\frac{X_G}{2}}^{X_G} \sum_{v=1}^{\frac{Y_G}{3}} \Gamma(h, v) \\ om_{21} = \sum_{h=1}^{\frac{X_G}{2}} \sum_{v=\frac{Y_G}{3}}^{2 \cdot \frac{Y_G}{3}} \Gamma(h, v) \\ om_{22} = \sum_{h=\frac{X_G}{2}}^{X_G} \sum_{v=\frac{Y_G}{3}}^{2 \cdot \frac{Y_G}{3}} \Gamma(h, v) \\ om_{31} = \sum_{h=1}^{\frac{X_G}{2}} \sum_{v=2 \cdot \frac{Y_G}{3}}^{Y_G} \Gamma(h, v) \\ om_{32} = \sum_{h=\frac{X_G}{2}}^{X_G} \sum_{v=2 \cdot \frac{Y_G}{3}}^{Y_G} \Gamma(h, v) \end{array} \right\} om_{ij} \in \mathbb{N} \quad (8)$$

After the six regions of the grid have been mathematically defined by (8), the subsequent geometrical parameters are calculated for each specific region om_{ij}. The first such parameter is the cumulative profile $CUM_{ij}(h)$ defined in (9):

$$CUM_{11}(h) = \sum_{v=1}^{\left\| \frac{Y_G}{3} \right\|} \Gamma(h, v); \quad h = 1, \ldots, \|X_G/2\|$$

$$CUM_{12}(h) = \sum_{v=1}^{\left\| \frac{Y_G}{3} \right\|} \Gamma(h, v); \quad h = \|X_G/2\| + 1, \ldots, X_G$$

$$CUM_{21}(h) = \sum_{v=\left\| \frac{Y_G}{3} \right\|+1}^{2 \cdot \left\| \frac{Y_G}{3} \right\|} \Gamma(h, v); \quad h = 1, \ldots, \|X_G/2\|$$

$$CUM_{22}(h) = \sum_{v=\left\| \frac{Y_G}{3} \right\|+1}^{2 \cdot \left\| \frac{Y_G}{3} \right\|} \Gamma(h, v); \quad h = \|X_G/2\| + 1, \ldots, X_G$$

$$CUM_{31}(h) = \sum_{v=2 \cdot \left\| \frac{Y_G}{3} \right\|+1}^{Y_G} \Gamma(h, v); \quad h = 1, \ldots, \|X_G/2\|$$

TABLE I
DEFINITION OF PERCEPTION SITUATIONS

Sit.	Description	Outcome
0	No features: perception error	Halt robot and send warning
1	Both rows are detected	Calculate target point P_t
2	Left row is detected	Calculate target point P_t
3	Right row is detected	Calculate target point P_t
10	Too close to left row	Severe correction to right
11	Too close to right row	Severe correction to left

$$CUM_{32}(h) = \sum_{v=2 \cdot \left\| \frac{Y_G}{3} \right\|+1}^{Y_G} \Gamma(h, v); \quad h = \|X_G/2\| + 1, \ldots, X_G \quad (9)$$

The cumulative profile of (9) was used to calculate the moment M_{ij}, whose expression for the first zone M_{11} is given in (10), being the rest of moments M_{ij} for the rest of the zones easily deducible using the same procedure as in (9) and (10). Associated with the moment, and following the same philosophy, the summation of the cumulative profile SUM_{11} was calculated according to (11). Notice that the summation of the cumulative profile for a given operational zone is equivalent to the amount of cells in that zone, which implies that $om_{ij} = SUM_{ij}$.

$$M_{11} = \sum_{h=1}^{\|X_G/2\|} ((\|X_G/2\| - h) \cdot CUM_{11}(h)) \quad (10)$$

$$SUM_{11} = \sum_{h=1}^{\|X_G/2\|} CUM_{11}(h) \quad (11)$$

The objective of calculating moments is detecting the highest likelihood, within operational zones om_{ij}, of locating vegetation rows based on perceptual evidences. Taking into account that the algorithm expects parallel rows ahead of the vehicle, each zone resulted in one expected placement for the section of the row given by function L and defined in (12) for both the left and right side of the field of view. The alignment of L_{ij} in well-populated zones was an indication of reliable perception and thus led to stable estimations of the target point. Misalignments and scarce filling of the grid anticipated complex navigation scenarios. Some examples of the calculation process until the estimation of the position of the rows given by function L is shown in Fig. 5. The morphology of occupancy matrix OM resulted in the definition of the six perception situations enunciated in Table I.

$$L_{ij} \overset{\text{def}}{=} \left\{ \begin{array}{l} L_{i1} = \left\| \frac{X_G}{2} - \frac{M_{i1}}{SUM_{i1}} \right\| \\ L_{i2} = \left\| \frac{X_G}{2} + \frac{M_{i2}}{SUM_{i2}} \right\|; \end{array} \right.$$
$$i = 1, 2, 3; \; j = 1, 2; \; SUM_{ij} \neq 0 \quad (12)$$

The specific results derived from the calculation of the occupancy matrix OM in (8) provide the evidence of the perception reality ahead of the robot, and therefore are determinant to choose one expected situation from the list given in Table I. In particular, two activation modes were defined according to (14) and (15): *high activation* and *low activation*. In physical terms, high activation represents a strong evidence of feature

TABLE II
ACTIVATION OF SITUATION 2

	C 1	C 2	C 3	C 4	C 5	C 6	C 7
om$_{11}$	δ^L_{11}		δ^H_{11}			δ^H_{11}	
om$_{21}$	δ^H_{21}	δ^H_{21}	δ^L_{21}	δ^L_{21}			δ^L_{21}
om$_{31}$				δ^L_{31}	δ^L_{31}		
om$_{12}$		δ^L_{12}					

TABLE III
ACTIVATION OF SITUATION 3

	C 1	C 2	C 3	C 4	C 5	C 6	C 7
om$_{11}$		δ^L_{11}					
om$_{12}$	δ^L_{12}		δ^H_{12}			δ^H_{12}	
om$_{22}$	δ^H_{22}	δ^H_{22}	δ^L_{22}	δ^L_{22}			δ^L_{22}
om$_{32}$				δ^L_{32}	δ^L_{32}		

existence (canopy and supporting structure), whereas low activation implies a weaker evidence and thus major uncertainty in the identification of guidance patterns.

$$\overline{OM} = \max_{\substack{1 \le i \le 3 \\ 1 \le j \le 2}} om_{ij} \tag{13}$$

$$\textit{High activation } \delta^H_{ij} \Leftrightarrow om_{ij} \ge \left\| 0.4 \cdot \overline{OM} \right\| \tag{14}$$

$$\textit{Low activation } \delta^L_{ij} \Leftrightarrow om_{ij} \ge \left\| 0.2 \cdot \overline{OM} \right\| \tag{15}$$

The preceding set of equations, in particular (8), (14), and (15), enunciate the conditions that place the robot in one of the situations defined in Table I. Only situations 1, 2, and 3 will lead to the calculation of the target point, which will be computed with the application of (9) to (12). Situation 0, to begin with, represents the absence of features, which is an indicator of a failure in the perception system (all sensors failing) or the possibility of the robot getting out of the field by mistake, or even large unexpected gaps at both side rows. In any case, the situation is unstable and requires stopping the robot motion. From a practical standpoint, we can consider that the occupancy matrix OM is empty (situation 0), allowing for occasional momentary noise, with less than a value of one ($om_{ij} = 1$) per operational zone in average, i. e., less than a sum of six for the entire grid, as mathematically defined in (16).

$$\sum_{i=1}^{3} \sum_{j=1}^{2} om_{ij} \le 6 \tag{16}$$

The logic conditions for meeting situation 2 make use of the activations states δ^H_{ij} and δ^L_{ij} defined in (14) and (15), according to the combinations stated in Table II. Any of the seven conditions of the table will activate situation 2, which according to Table I indicates that the left row has been detected by the perception system, and therefore is eligible for the calculation of the target point P_T.

In a similar fashion to the logic rationale for activating situation 2, Table III provides the combinations to fire situation 3, which indicates a correct detection of the right row in Table I.

Situation 1, which is the most desired in terms of stability, occurs when both situations 2 and 3 are simultaneously activated, as determined by Tables II and III. For between-rows guidance purposes, only situations 1, 2, and 3 are valid, as are the ones used to calculate the target point, which in turns determines the steering angle sent to the front wheels. Situations 10 and 11, by contrast, indicate a potential collision

risk whose anticipation requires a sharp reaction with no need of knowing the ideal position for the target point, obviously without value in such circumstances. The activation of situations 10 and 11 combines information from the occupancy matrix –just like the operations firing situations 2 and 3– with lidar and sonar specific constrains.

A. Calculation of the Target Point P_t

The position with the highest likelihood for the left and right vegetation rows that serve as guidelines is given by (8) and (12) according to the rate of occupancy found in OM. Situations 2 and 3 (Table I) only perceive one of the guiding rows, and therefore are forced to position the target point displaced half the row spacing from the detected row boundary given by function L (12). Situation 1, on the contrary, locates both guiding lines within the APOM, and P_t will reside in the geometrical locus that is equidistant from both estimated row boundaries (12). Each operational zone with the proper filling of its cells will lead to an estimated line position L_{ij}, as graphically represented in Fig. 5. The particular section of the line L_{ij} that intervenes in the calculation of P_t (x_t, y_t) depends on the activation of specific om_{ij}. Situation 1, for instance, made use of 16 logic propositions to determine the horizontal position (h_t) of P_t in the APOM. The vertical position (y_t), was calculated as the sum of the look-ahead distance of 5 m and the wheel base (0.6 m in the robot of Fig. 3). Fig. 5 shows diverse real scenarios in which the APOM has been populated with a 3D stereoscopic camera (green cells) and a multi-beam lidar (red cells). Each grid also depicts (in blue) the six lines L_{ij}, given by (12), pointing at the best estimates for the row boundaries, together with the resulting position for P_t in the grid. In the two scenes portrayed in Fig. 5, vine variety, soil conditions, and row spacing were different (left vineyard in Portugal; right vineyard in Spain). Notice on the right scene that when the 3D perception (canopy represented with green pixels) was lost for the right row, the lidar marked the position of that row for the estimation of L_{22} and L_{32}, proving the value of augmented perception for a reliable solution. In this case, the proper position of P_t was determined without the participation of L_{12} in the calculation of P_t because om_{12} was empty.

IV. METHODS FOR ANALYTICAL COMPARISON

The evaluation of a perception system for autonomous navigation must be based on the real capacity of the auto-steered
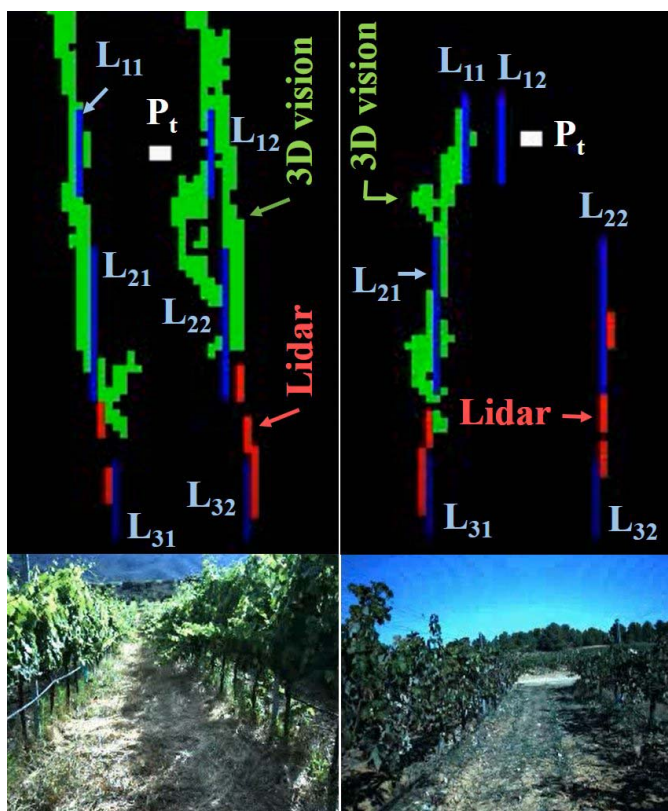
Fig. 5. Calculation of the target point in the APOM for real scenarios in commercial vineyards.

Fig. 6. Evaluation of guidance performance with lime-marked trajectories.

vehicle to navigate safely in relevant environments while executing a task efficiently. This is easy to check by visual inspection over a limited period of time, but very complex –if not inviable– to quantitatively assess for any possible environment (complying with the design requirements) encountered during the life span of the vehicle. In order to validate the proposed multi-perception strategy, we will use five different evaluation procedures, none of which perfect by itself, but in conjunction providing a useful assessment of the performance of the navigation algorithm embedded in the robot. To begin with, the observable results are the coupled effect of the perception algorithm and the control system executing the steering commands. Therefore, a tight maneuver might be the result of a poor calculation of the steering command, or the inadequate actuation of the steering mechanics, or very likely both in certain degree of participation.

In addition to the coupling effect of the control system actuating on the front wheels, there is another handicap affecting the evaluation of navigation performance: the lack of a well-determined reference allowing a quantitative and objective assessment of deviations without ambiguity. It is obvious that an autonomous vehicle must circulate between adjacent rows without touching them, but after traversing several rows without crashing, the question to answer is which one had a better performance, and what made it be the optimal. GPS-based navigation has been evaluated in especially designed tracks, where key parameters are under control and the track is very accurately geographically referenced. However, this solution makes no sense for perception-based navigation where the

surrounding environment is continuously changing. Even the same rows when the canopies grow, there is wind, or the soil conditions vary offer a completely different situation. As a result, a dynamic evaluation, which would be very difficult to replicate in a reference track, is necessary. A dynamic method to evaluate the performance of auto-guidance in the field without the need of fixed reference tracks was developed in [21], but it relies on the accurate recording of the reference trajectory, which requires a very precise GNSS receiver with sub-inch errors (RTK) and the assistance of a very skillful driver, both of which difficult to assure in the field in a regular basis. The methods described below bring together complementary views of the principal goal of evaluating navigation performance, and although the following sections will analyze some particular runs of various fields and conditions, the strength of the methodology resides in the capacity of analyzing many rows under all kind of conditions; only by conducting massive analytical comparisons will we be able to conclude on the superiority of one algorithm over the rest.

The first evaluating method is very straightforward, and consists of dropping lime powder as the robot moves forward to draw the actual trajectory followed in autonomous mode. Once the trajectory was drawn over the ground, deviations from a geometrical centerline that is equidistant to the polylines defined by the vine trunks at both sides were manually measured with a tape. Fig. 6 illustrates the procedure and Fig. 7 plots the results of applying this procedure. The second method focuses on the monitoring and analysis of the perception situations defined in Table I. As the objective is the evaluation of inside-row guidance, those situations leading to the calculation of the target point P_t will be favorable, i.e., 1 to 3, whereas the rest will indicate the activation of warning signs and correcting commands. The third approach takes advantage of the lateral distance measured by the side sonars. The more centered the vehicle is, the closer these two distances will be among them. If L_S is the distance to the canopy measured by the left sonar (cm), and R_S (cm) is the corresponding distance measured by the right-side sonar, we can define the left-right offset ratio ρ by expression (17), where the most stable situation from the navigation stand
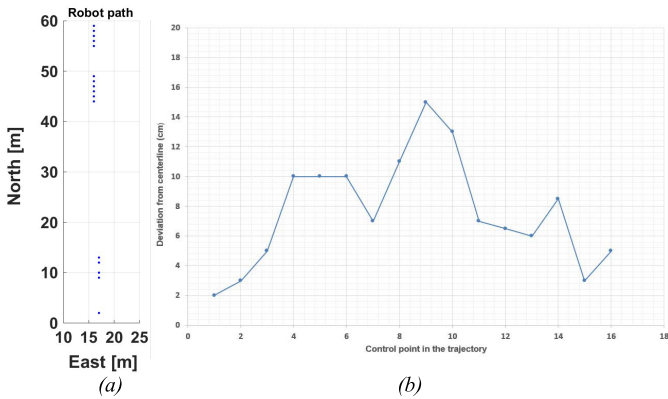
Fig. 7. Analysis of deviations from virtual centerline: (a) GPS trajectory of full row; (b) Deviations measured with lime dust.

point will occur when $\rho \to 0$. Notice that L_S and R_S are bounded in (17) by the physical limitations of the specific ultrasonic sensors used. The fourth method focuses on the precision in the execution of steering commands, and therefore also accounts for the performance of the control system and steering design, which as mentioned before, are coupled with the behavior of the perception system. It is based on the comparison of the profile of commanded steering angles and the actual orientation of the front wheels measured by a linear potentiometer (PC67, Gefran spa, Provaglio d'Iseo, Italy). Finally, the last comparison method envisioned makes use of the onboard electronic compass to assess heading stability, under the hypothesis that smooth rides will be associated to slight yaw fluctuations around a virtual centerline that is equidistant from left and side rows.

$$\rho = \left| 1 - \frac{R_S}{L_S} \right| \quad R_S \in [8, 200]; \quad L_S \in [8, 200] \quad (17)$$

V. RESULTS IN REAL ENVIRONMENTS

A. Selection of Representative Runs and Coupling Effects

The navigation strategy based on augmented perception presented in this article was developed along two research projects lasting seven years, and the results obtained come from field experiments conducted between 2015 and 2019 in commercial vineyards of France, Spain and Portugal. The ideal experiment would be that conducted in an invariant vineyard where conditions are fully controlled and numerous test would succeed over and over under permanent challenges. Unfortunately, that is not possible in practice. The algorithms and sensing capabilities of the robots have been improving along the time, such that many vineyards under very different conditions were tested. The specific characteristics of the testing runs along the vineyard plots used in the upcoming analysis is included in Table IV. Three different robotic platforms have been used with diverse configurations for the perception engine. In this section, we will compare such perceptive configurations for vineyard rows that approximately present similar challenges in terms of canopy structure, soil conditions, robot forward velocity, and environmental hardships; and contrarily, we will

Data Series	Total Data points	Points inside rows	Total runtime (min)	Average speed (km/h)	Total run length (m)	Rows traversed
AA	2203	1577	25.1	1.7	694	5
B	12623	10914	78.7	1.4	1885	15
C	4604	3649	23.3	1.8	712	6
E	13465	11176	46.7	1.4	1114	9

analyze how a particular perception configuration reacts to scenarios posing challenges of different nature. The comparison methods applied have been described in Section IV, and in addition to the fact that diverse perceptive solutions were compared using different vineyard rows in different moments, results should be interpreted taking into account the coupling effect of the steering control system on the final behavior of the robot, which is the observable outcome upon which measurements and comparisons may be carried out.

B. Analysis of Deviations From Virtual Centerline of Trajectories Marked With Lime

Fig. 6 provides an overview of this experimental setup, where a lime container with a vibrating implement was affixed to the rear of a robot, placed in its geometrical center. When the robot was engaged in autonomous mode, the lime container dropped the white dust on the ground as the vehicle moved forward, tracing the trajectory of the robot until the container ran out of lime. Fig. 7a plots the full run where only the first 16 m were measured with the lime trace. Fig. 7b provides the detailed deviations for the 16 points measured, with an approximate distance between adjacent points of 1 m. The average deviation from the centerline was 7.6 cm. The perception engine at the time of the tests (21 June 2016) only used the forward-looking stereoscopic camera for inside-row guidance.

C. Quantifying Inside-Row Navigation Complexity Through the Comparison of Perception Situations

The perception situations enunciated in Table I provide a quantitative means of tracking the events of sub-optimal actuation, when the robots get too close to the vines for a safe navigation. Situations 1 to 3, in particular, are considered reliable outputs of the perception system, whereas 10 and 11 indicate a risk of collision, and although the vehicle usually got out of these situations and recovered the centerline, the maneuver did not convey the stability desired in autonomous guidance, while increasing the chances of an accident. Showing stability for a demonstration run of a few minutes is not a big problem, but what results interesting for developing an autonomous vehicle is tracking stability and behavior in the long run, where batteries are not fully charged, computers may become overheated, sensors are exposed to strong sun radiation, and new areas of the field never traversed before may pose unforeseen challenges due to careless canopies, unseen damage, unknown weeds, or rough terrain. The objective of this analysis, and those in the remaining of Section V, is twofold:
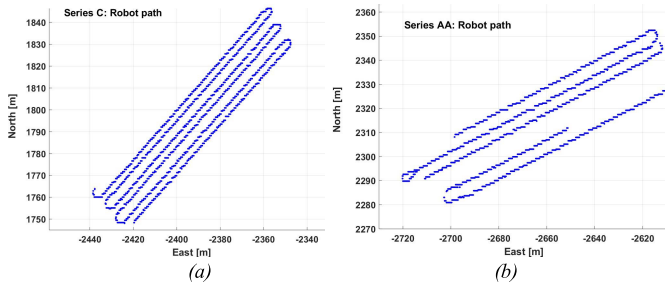
Fig. 8. Trajectory followed by the robot in autonomous mode: (a) Series C; (b) Series AA.

1. Demonstrate the advantageous performance of augmented perception for autonomous navigation inside orchard rows.

2. Develop a methodology to quantify the behavior of local perception systems, especially those that combine sensors working under diverse physical principles such as vision, lidar, and sonar. This methodology should be applicable in any relevant environment (unknown beforehand) rather than pre-defined testing tracks.

Aligned with the second objective stated above, the first challenge encountered in the field was to quantify the complexity that any given orchard scenario poses to a determined perception engine. The comparison of perception situations (Table I) was conducted with the same robotic prototype in two vineyards of the same region and cooperative (Data series AA and C). Data were acquired with the robotic prototype of Fig. 6, whose perception engine included a binocular 3D stereoscopic camera (Bumblebee 2, FLIR Systems, Inc., Wilsonville, OR, USA) for inside-row guidance and six low-cost ultrasonic sensors (Ping, Parallax, Rocklin, CA, USA), three of them facing forward, two looking sideways, and the last one in the rear for reverse maneuvers over the headlands. The classification of perception situations was basically driven by the 3D stereo camera and the lateral sonars facing the canopies. Both data series were recorded in Buzet-sur-Baïse, France. Series C (5 September 2016) represents an ideal vineyard, where canopies were carefully trimmed, weeds were incipient or removed, and there was no slope or vegetation gaps. Series AA (23 June 2016), in contrast, was acquired in a complicated vineyard with certain slope and soft terrain where the robot found tractive problems. Fig. 8a plots the six rows of Series C that the robot traversed in autonomous mode without any incident, and Fig. 8b represents the Series AA trajectory followed by the same robot where in the middle of the five rows the operator had to intervene once.

According to Table I, except for the case of situation 0 when there is no perception, the ideal perception situation is 1, situations 2 and 3 are weaker than 1, but still allow for the calculation of the target point, and alerting situations are penalized with scores 10 and 11. As a result, the summation of perception situations with time will grow faster as more unstable commands take place along the rows. To be able to compare plots and series of different size, this summation must be normalized by the number of data points, and the situations related to the headland turns have to be removed
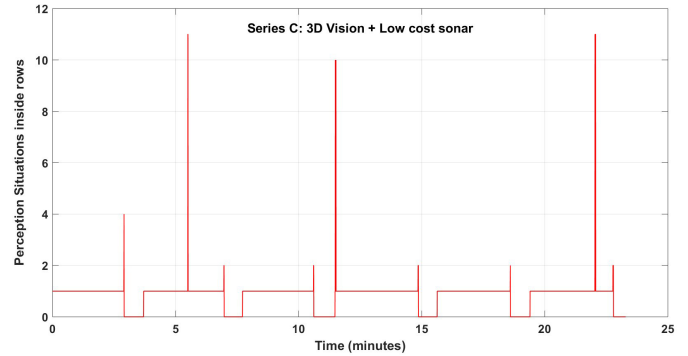

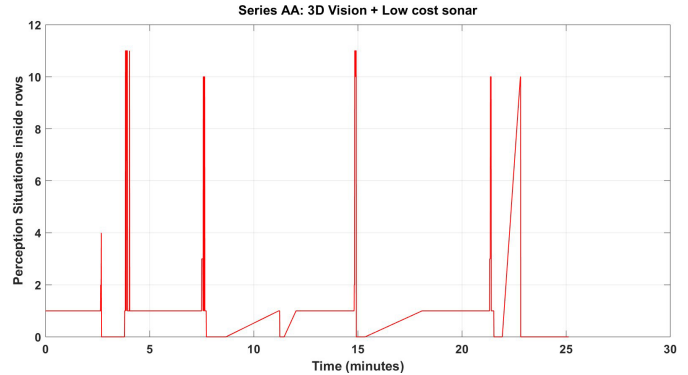Fig. 9. Perception situations for Series C.

Fig. 10. Perception situations for Series AA.

from the series for the analysis, as we are analyzing inside-row navigation. This way of charging over risky conditions from the perception standpoint is similar to the idea of "weighing evidences" proposed by Marvin Minsky [22], where weights increase fast with unexpected orientations of the robot (heading) and their consequent dynamic instability. Fig. 9 depicts the perception situations detected in Series C after removing the data at the headland turns, and Fig. 10 shows the same plot for Series AA. Notice that situation 0 was assigned to the headland turns in Figs. 9 and 10, as such perception situation (sit $0 \equiv$ no features detected) was never detected during runtime in any data series. Likewise, the earlier version of the algorithm also included situation 4 to point the beginning of the first headland turn, as it appears in both figures. Taking the size of the data series from Table IV, the summation of situations for Series C was 3742/3649 points = 1.02 whereas for Series AA was 1943/1577 points = 1.23. The number of situations 10-11 for Series AA was 35 after removing the headlands, and for Series C was 8. If we calculate the percentage of risky situations out of the number of points recorded for inside-row guidance, the results are given in (18):

$$\left.\begin{array}{l} S_{AA} = \dfrac{35 \text{ situations } 10-11}{1577 \text{ points}} \cdot 100 = 2.2\% \\[2mm] S_{C} = \dfrac{8 \text{ situations } 10-11}{3649 \text{ points}} \cdot 100 = 0.22\% \end{array}\right\} \quad (18)$$

D. Navigation Stability From the Study of Left-Right Offset Ratio ρ

The situations of Section V-C and Table I provide an estimate of the overall performance of the perception algorithm in
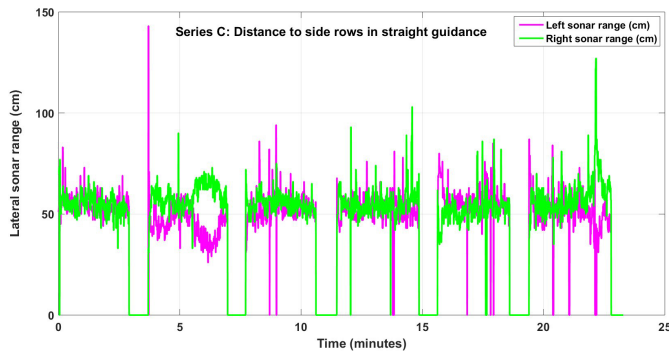
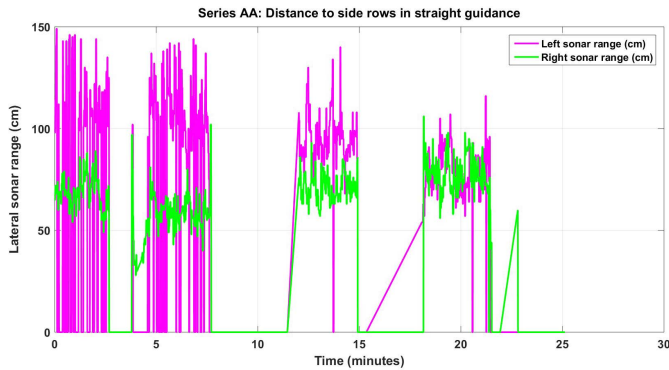Fig. 11. Distance to side rows in straight guidance for Series C.

Fig. 13. Left-right offset ratio ρ for Series C.

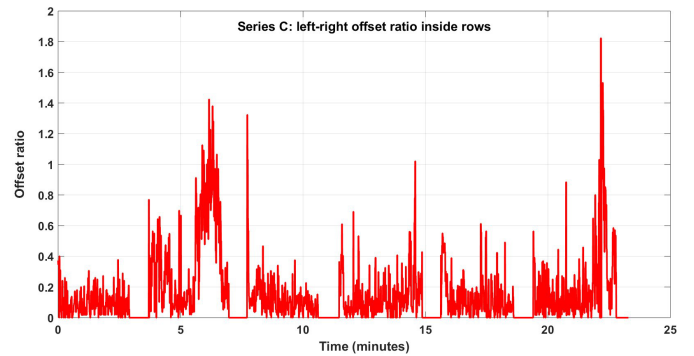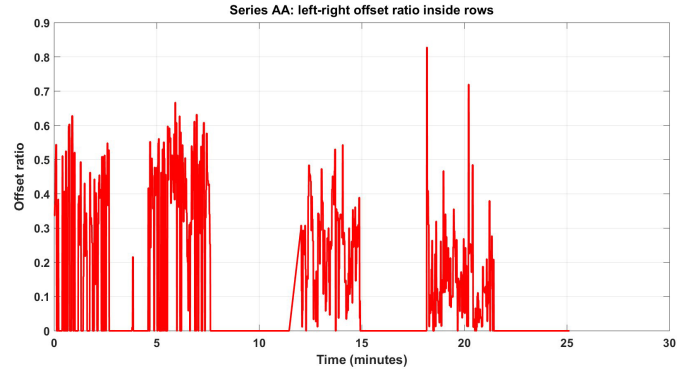Fig. 12. Distance to side rows in straight guidance for Series AA.

Fig. 14. Left-right offset ratio ρ for Series AA.

the detection of guiding rows, but do not account for the actual position of the robots related to the surrounding canopies. However, the instantaneous measurement of the distance from the robot sides to the bounding canopies provides an accurate account of the capacity of the perception algorithm to keep the robot close to the virtual centerline that is equidistant to adjacent guiding rows. This account, therefore, should reflect the intrinsic difficulty of any tested run, as well as allow the numerical quantification of such difficulty. Figs. 11 and 12 plot these measurements for Series C and AA, respectively. For the former, the average left separation was 52 cm whereas the average right separation was 56 cm. For the Series AA, the average left separation was 76 cm and the average right separation was 67 cm. The smaller differences between left and right distances, the more centered the robot navigates. The left-right offset ratio ρ of (17) avoids negative numbers when the vehicle shifts from the left to the right side of the centerline, and yields 0 when the robot is centered in the row. The stability of runs, therefore, can be estimated by tracking the summation of ρ normalized by the series size. Fig. 13 plots the ratio ρ for the C Series, and Fig. 14 shows the same profile for Series AA. The summation of the offset ρ for Series C was 699/3649 points $= 0.19$, whereas for Series AA was 350.8/1577 points $= 0.22$. Stability increases as the summation of ρ approaches to 0. After removing the points associated with the headlands, there were still invalid values for (17) that were coded as 0 in Figs. 13 and 14. These null values altered the calculation of the real average offset. After their removal, the final average offset ratio R for each series was obtained

in (19).

$$R_{AA} = \frac{\sum \rho}{\sum \rho \neq 0} = \frac{350.8}{1207 \text{ points}} = 0.29$$

$$R_{C} = \frac{\sum \rho}{\sum \rho \neq 0} = \frac{699}{3405 \text{ points}} = 0.2$$

(19)

The results of (19) yield the average offset ρ, but higher divergences are found when the median is calculated, as the median for Series C is $\text{Med}(\rho_C) = 0.12$ whereas for Series AA is $\text{Med}(\rho_{AA}) = 0.29$.

E. Analysis of Steering Performance to Detect Mechanical Limitations

The perception situations and offset ratio previously analyzed allow the assessment of navigation performance, but mask the effect of the steering system, whose accurate actuation is essential for automatic steering systems. Very slight misalignments in the steering mechanism result in asymmetrical performance of the front wheels for Ackerman steering. Even when the robot mechanics are carefully built, the recurrent exposure to uneven terrains and rough ground ends up loosing linkages and reducing the precision of commanded angles. For manned vehicles, this problem is not usually a hazard because the operator somehow corrects for the misalignments and sends the vehicle to the maintenance shop when necessary. For unmanned vehicles, by contrast, steering misalignments and loose fittings can have severe consequences whose origin are sometimes hard to identify. A continuous –or
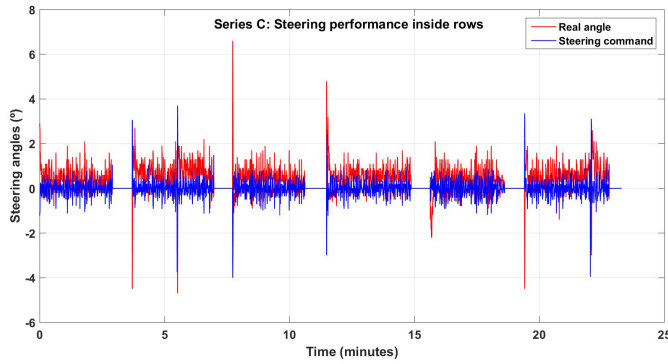
Fig. 15. Analysis of steering stability for Series C.

Fig. 16. Analysis of steering stability for Series E.

periodic– self-checking procedure for assessing steering consistency can result instrumental for long-term use of intelligent equipment endowed with automated navigation.

The perception system –sensor suite and algorithm– calculates steering angles based on the instantaneous position and orientation of the robot with respect to the surrounding environment. The profile of the calculated steering angles gives an idea of the stability of the steering actuation, such that sudden large commands are an indication of navigation alerts leading to situations 10 and 11. The response of the steering system, however, is traceable by tracking the actual angles of the front wheels measured by a linear potentiometer. We cannot assume that all commands sent by the algorithm are precisely materialized in due time, as there are always delays in the system response and inertias to overcome caused by the interaction of the tire with the terrain. Fig. 15 plots the comparison of calculated angular commands and real angles for Series C. As desired, the average angle commanded by the algorithm is 0, which is a clear indication that the robot travelled along the centerline between the adjacent vineyard rows. However, while the robot moved straightforwardly, the actual angles measured at the front wheels had an average of 0.4°, which reveals that the steering linkage had a slight misalignment of +0.4°. The plot also shows that angles are slightly larger at the entrance maneuver of each new row when the wheels have to recover the alignment after the 180° turn at the headlands, except for two occasions in the 2nd and 6th rows where an angular correction of 4° was necessary in the middle of the run.

The limitations introduced by the mechanical embodiment of the steering system not only affect permanent misalignments caused by systematic offsets, but also a lack of symmetry in the execution of turns, which is more acute at the sharp angles required for changing rows. Fig. 16 plots the steering angles (actual and commanded) for Series E, including the nine turns at the headlands. In addition to show a misalignment when the robot moves inside the row, with an average real angle of −0.6° when the commanded angles are 0° in average, the turning capacity of the robot was asymmetrical and limiting. The graph reveals that the maximum right angles (positive angles) were around 20°, which were quickly reached with commanded angles of 15°. However, left angles were physically constrained by −15°, and
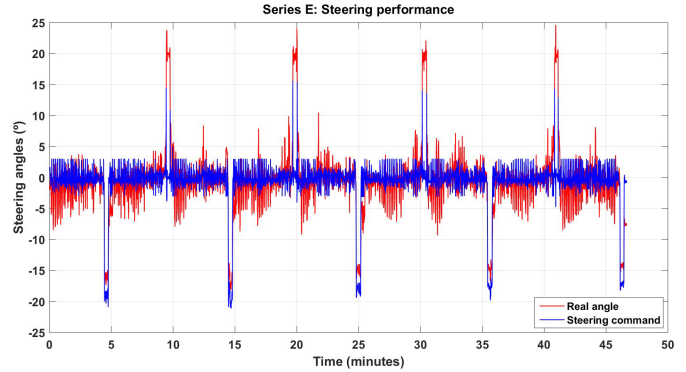
could never exceed this limit in spite of recurrent commands of −20°. Although headland turns (where extreme turning is necessary) fall outside the scope of this article, it is worth mentioning the capacity of this evaluation method to locate important sources of problems and malfunctioning, especially in automated equipment where constant supervision is no longer present.

F. Self-Assessment of Augmented Perception

After developing a methodology to evaluate the performance of a perception system for guiding a vehicle inside rows, and taking advantage of the fact that this methodology can be applied to any vehicle operating in any relevant environment (specialty crops under vertical trellises), the goal of this section is to apply such methodology to different configurations of the perception system for a quantitative evaluation. The first case is represented by Series E (Table IV), which features a perception system based on the 3D stereoscopic camera used in series C and AA but augmented with a forward looking 11-beam 2D LIDAR (OMD 8000-R2100-R2-2V15, Pepperl + Fuchs, Manheim, Germany). Series E was recorded in Portugal on 16 July 2018, and portrays a vineyard with the typical challenges of commercial plots: canopy gaps, long shoots, mild slope (up to 10°), and a terrain of varying conditions. The specifications of the series are included in Table IV, and Fig. 17 plots the trajectory followed by the robot in autonomous mode.

As there were no lateral sonars in the robot when Series E was recorded, the evaluation based on the left-right offset ratio ρ cannot be carried out. As a result, navigation stability will be assessed by the analysis of perception situations, as outlined in Section V-C. Fig. 18 depicts the perception situations monitored over the nine straight paths of Fig. 17, in which the summation of situations was 11476/11176 points = 1.02. The number of situations 10-11 for Series E was 30, and the percentage of alerts recorded for inside row guidance was 0.27 % as detailed in (20).

$$S_E = \frac{30 \text{ situations } 10-11}{11176 \text{ points}} \cdot 100 = 0.27\% \qquad (20)$$

The Series B was recorded with the prototype of Fig. 3, and featured the full-perception augmented approach consisting of the 3D stereo camera already used in Series E, amplified
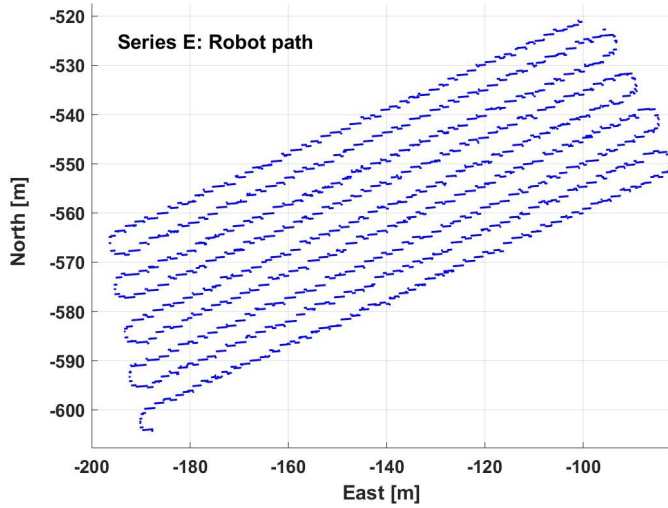
Fig. 17. Trajectory followed by the robot in Series E.
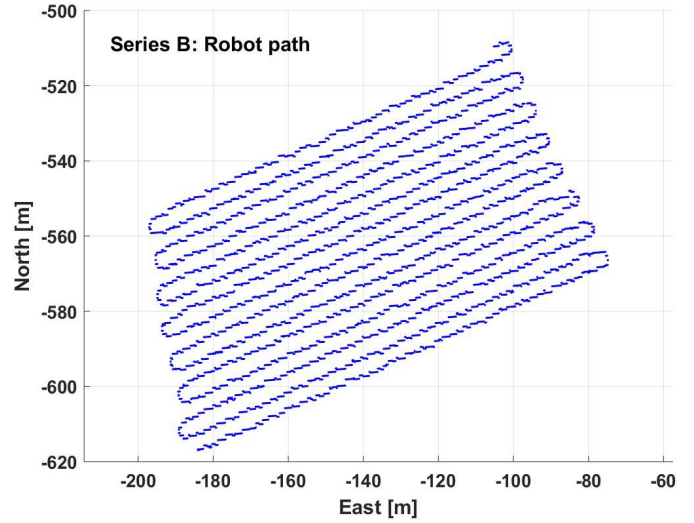

Fig. 19. Trajectory followed by the robot in Series B.
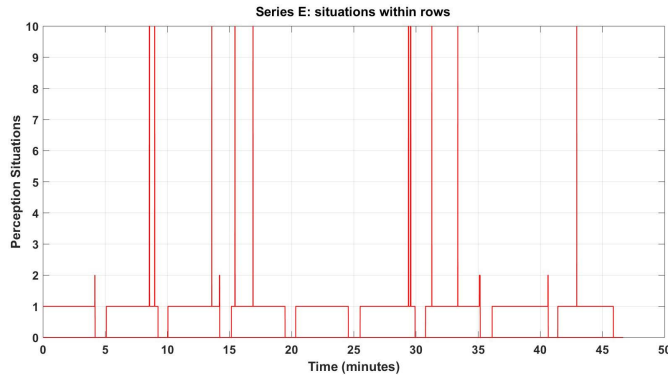

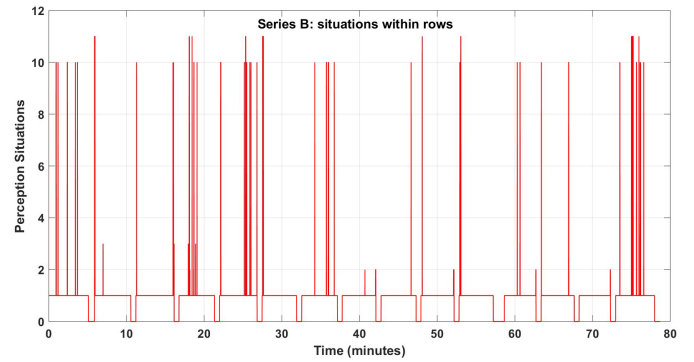Fig. 18. Perception situations for Series E.

Fig. 20. Perception situations for Series B.

with three ultrasonic sensors (UC2000-30GM-IUR2 V15, Pepperl + Fuchs, Manheim, Germany) and the forward looking 11-beam 2D LIDAR (OMD 8000-R2100-R2-2V15, Pepperl + Fuchs, Manheim, Germany). The forward-looking sonar was only used for obstacle detection and headland turning, therefore only the lateral sonars (left and right) were actually used for inside row guidance. Fig. 4 indicates the position and range of the ultrasonic sensors. Series B data were recorded in Quinta do Ataíde, Portugal, on 5 September 2019, which is the same vineyard used for Series E. The specifications of Series B are included in Table IV, and Fig. 19 plots the trajectory followed by the robot in autonomous mode.

The first analysis centers on the perception situations tracked along the 15 rows plotted in Fig. 19. The summation of situations for this series, derived from Fig. 20, was 12070/10914 = 1.1, whereas the number of situations 10-11 was 119, leading to a percentage of alerts for inside row guidance of 1.1 % as detailed in (21).

$$S_B = \frac{119 \text{ situations } 10-11}{10914 \text{ points}} \cdot 100 = 1.1\% \quad (21)$$

The analysis of the navigation stability based on the offset ratio ρ results from the lateral distances of Fig. 21 and the subsequent ratio of Fig. 22. Specifically, the average left

separation measured with the lateral sonar was 67 cm whereas the average right separation was 73 cm. The summation of the offset ρ for Series B was 3636/10914 points = 0.33. As usual, stability increases as the summation of ρ approaches to 0. After removing the points associated with the headlands and other null values, the average offset ratio R_B for this series was obtained as detailed in (22). The median of offset ratio ρ for Series B was $Med(\rho_B) = 0.25$.

$$R_B = \frac{\sum \rho}{\sum \rho \neq 0} = \frac{3636}{10700 \text{ points}} = 0.34 \quad (22)$$

The steering performance for Series B can be deduced from the profiles plotted in Fig. 23. The average real angle measured by the potentiometer was $-0.2°$ whereas the average commanded angle was $0.1°$. Both are close, and close to 0, but the profile of real angles in Fig. 23 shows larger corrections than in Fig. 15, which implies a less stable navigation performance. Mathematically, these fluctuations around the centerline can be estimated through the standard deviation, being $2.6°$ for the real angles and $1.3°$ for the angular commands sent by the controller to the steering motor. Fig. 23 also indicates that the front wheels turned sharper to the right than to the left.

The last comparison uses an onboard electronic compass (SEC385, Bewis Sensing Technology LLC, Wuxi City, China)
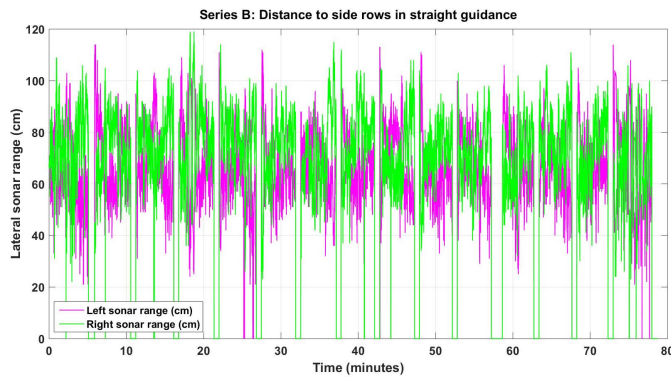
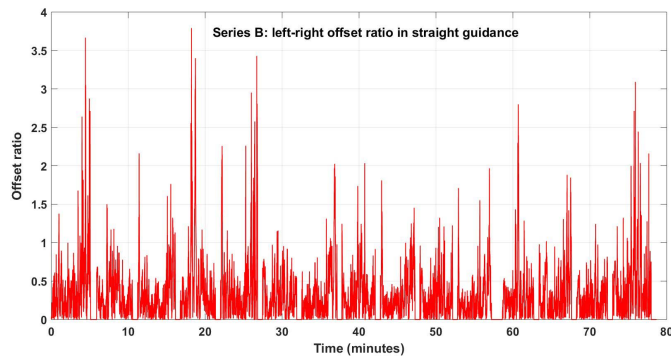Fig. 21. Distance to side rows in straight guidance for Series B.

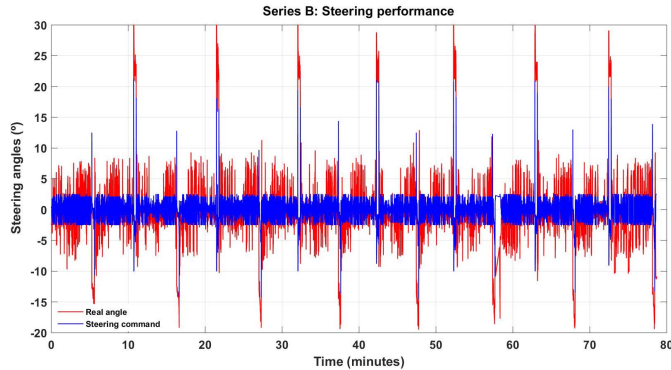Fig. 22. Left-right offset ratio ρ for Series B.

Fig. 23. Analysis of steering stability for Series B.

Fig. 24. Analysis of heading stability for Series B.

to assess heading stability, by associating smooth rides to small yaw fluctuations around the average heading of each row. Fig. 24 overlays the instantaneous heading measured by the electronic compass and, at the same time, by the onboard GPS. As shown in the plot, the GPS estimates have so much variability due to noise that they do not reflect the behavior of the robot, and therefore cannot be used in the stability analysis. The standard deviation registered for the rows whose heading averaged 72° was $\sigma_{72} = 3.7°$, and the dispersion for the rows of average heading 245° was $\sigma_{245} = 4.7°$.

G. Discussion

The procedure illustrated in Fig. 6 to evaluate the navigation performance of an autonomous vehicle by tracing its trajectory
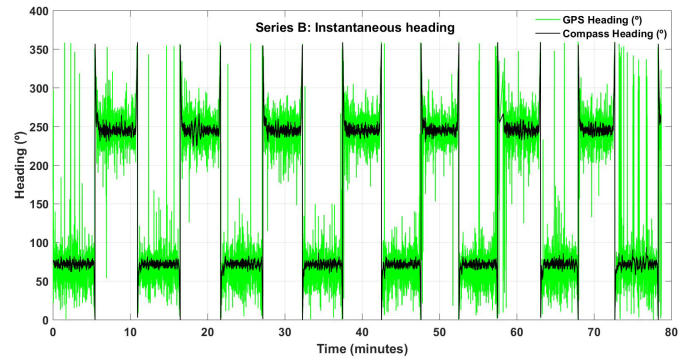
with lime turned out to be cumbersome, impractical, and inaccurate. Lime dust had to be under constant stirring to avoid clogging, and a manageable tank size only reached for 16 m. The measurement of deviations from the centerline resulted time-consuming, physically-demanding, and prone to error as it was not always clear to identify the precise boundaries defining the deviations with a lime line several cm wide. The average deviation of 7.6 cm, however, is comparable with other estimates based on the analysis of the lateral distances to the canopies. Thus, for example, the average deviation for Series C was 2 cm, for Series AA was 4.5 cm, and for Series B was 3 cm. With a longer run, one complete row at least, the deviations measured with lime would probably have been smaller once the vehicle reached its regular traveling velocity, but this method is not applicable in a regular basis, and therefore cannot be considered an option for evaluating navigation performance.

The analysis of the situations defined in Table I provided a convenient self-assessment tool to evaluate navigation stability; it was not the summation of all situations but the normalized counting of situations 10-11 what resulted useful. The summation of situations, even after normalization with the number of measurements, was biased by population size, that is, the length of the series. The reason is that the majority of situations are 1, and if the series is long, the occasional summation of 10 or 11, even if these situations are weighted ten times, will have an overall mild effect. The counting of alerting situations, by contrast, allows the grading of navigation stability. Section V-C provides a reference range, where the same robot was tested under mild (Series C) and challenging (Series AA) environments. The most favorable outcome was $S_C = 0.22$ % (18) whereas the more complex scenario was quantified with $S_{AA} = 2.2$ % (18). Series E and B represent an intermediate case, i. e., a commercial vineyard with typical challenges, but with a robot endowed with augmented perception capabilities. $S_E = 0.27$ % (20) indicates a very stable navigation, but $S_B = 1.1$ % (21), even running the robot in the same plot as Series E, reveals a more oscillating behavior. The reasons for this may be various; one could be the fact that Series B was recorded after 6 pm, where batteries were at low charge. Other reason could be the vines or the terrain being in different conditions (E was recorded in July and B in September). Overall, even though it resulted quite complex to quantify the behavior of a vehicle in real

time before a changing environment, according to the field results it seems reasonable to expect good performance when S is below or around 1 %.

The evaluation method based on the study of left-right offset ratio ρ has a key advantage over the previous method: *simplicity*. The definition and verification of the perception situations of Table I is elaborate, including multiple conditions to meet that differ according to the perception sensors on board. The measurement of lateral distances and its corresponding calculation of ρ through (17), on the contrary, is straightforward, and only requires the side sonars to estimate R_S and L_S. The average values of R_S and L_S provide an estimate of the deviations from the centerlines, as shown in the discussion of the lime-based evaluation. As for ρ, the normalized summation suffers from the same disadvantage detected in the summation of perception situations; a masking effect as the population size grows. Therefore, even though the monitoring of ρ is quite simple, the normalized summation is not very helpful. However, the most significant parameter in relation to the offset ratio was the median. The same divergence found for Series C and AA regarding the perception situations was also found for the analysis of the offset ratio, specifically yielding $\text{Med}(\rho_C) = 0.12$ and $\text{Med}(\rho_{AA}) = 0.29$. For Series B, as expected, the outcome fell within that interval. A median above 0.3 would recommend a deeper examination of the navigation stability before proceeding further.

The analysis of steering performance was also revealing, which makes it attractive due to its simplicity. Only the real-time measurement of the front wheels Ackerman angle suffices to track what is occurring at the steering mechanism. With only basic inference statistics it was possible to detect slight misalignments –under $1°$– of the steering linkage, asymmetrical performance of the front wheels, and an oscillating-steady execution of guiding commands. More advanced analysis tools applied to the profiles of Figs. 15, 16 and 23 may bring a more complete picture of how efficiently the autonomous vehicle is executing automatic steering.

Finally, the assessment of heading stability with the onboard electronic compass showed potential, but could not be extensively analyzed with the detail of all other methods because it was implemented in the robots in 2019, and heading data was available only for Series B out of the rest of series cited in Table IV. A conclusion was straightforward, though: GPS instantaneous heading extracted from VTG NMEA messages resulted so unstable that it was actually useless. However, the standard deviations of compass-determined heading for two rows at $3.7°$ and $4.7°$ show what seems to be a stable behavior, although more data series will be needed for a consistent comparison.

VI. Conclusion and Future Steps

Autonomous navigation in the open environments of vineyards is a challenging feat, and its performance evaluation is equally demanding. To face the former, navigation was split into two mutually exclusive tasks: inside-row guidance and headland turning, with this article focusing on the first task. To confront the latter, a set of complementary methods were enunciated and demonstrated for a variety of scenarios. These methods are based on the measurements exerted by local perception sensors, an electronic compass, and a potentiometer to estimate the front wheels angle. The measurement of vehicle deviations with a lime dust dispenser makes no sense for common field extensions, but the permanent monitoring of the row-matching perception algorithm, the continuous logging of the lateral distance of the vehicle to the surrounding canopies, and the profile of the steering actuation allow for the calculation of various quality indices to assess the behavior of an autonomous vehicle. It is clear from this study that self-assessment is the most practical option, and to do so, we only need to use the sensors already onboard for navigation.

The goal of this work was not to come up with an evaluating method. As a matter of fact, this article mainly deals with perception algorithms for in-field navigation, but we cannot address navigation without a reliable way to assess guidance results for any possible situation unknown beforehand. Training the system always for the same row is unrealistic and misleading. Once a general methodology to evaluate results in real environments was established, various perception configurations were tried. The result was as expected; by fusing different –but complementary– technologies, the outcomes are more consistent and fail-safe, as shown in Fig. 5. For the strategy proposed in this research, the suite of sensors selected was 3D stereovision, multi-beam lidar, and sonar, chosen on the grounds of reliability under harsh environments and cost-efficiency (key in agricultural applications). For agricultural robots, and for specialty crops in particular, it is important to rely on on-vehicle perception solutions for navigation that offer stability and safety with independence to the availability of global positioning signals.

Future steps on this research topic will focus on three directions: first, the improvement of the augmented perception system with better devices and algorithms; second, the elaboration of a solid framework for headland turning, which is even more challenging than inside-row guidance; and third, keep developing the self-assessment methodology until it becomes independent of the vineyard size and configuration, as well as the environmental conditions. With no doubt, there is still a long way to go, but the only way to keep moving forward will certainly be by testing and developing in real scenarios, many times, many different geographical regions, and many hours under the sun, because the natural environments of agricultural robots are the fields rather than the labs.

REFERENCES

[1] H. P. Moravec, "Robot spatial perception by stereoscopic vision and 3D evidence grids," Carnegie Mellon Univ., Robot. Inst., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-96-34, 1996.

[2] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. 3rd Eur. Conf. Comput. Vis.*, 1994, pp. 151–158.

[3] F. RoviraMás, Q. Zhang, and J. F. Reid, "Automated agricultural equipment navigation using stereo disparity images," *Trans. ASAE*, vol. 47, no. 4, pp. 1289–1300, 2004.

[4] F. Rovira-Mas, J. F. Reid, and S. Han, "Obstacle detection using stereo vision to enhance safety of autonomous machines," *Trans. ASAE*, vol. 48, no. 6, pp. 2389–2397, 2005.

[5] F. Rovira-Más, Q. Zhang, and J. F. Reid, "Stereo vision three-dimensional terrain maps for precision agriculture," *Comput. Electron. Agricult.*, vol. 60, no. 2, pp. 133–143, Mar. 2008.

[6] F. Rovira-Más, Q. Zhang, and J. F. Reid, "Creation of three-dimensional crop maps based on aerial stereoimages," *Biosyst. Eng.*, vol. 90, no. 3, pp. 251–259, Mar. 2005.

[7] F. Rovira-Más, J. F. Reid, and Q. Zhang, "Stereovision data processing with 3d density maps for agricultural vehicles," *Trans. ASABE*, vol. 49, no. 4, pp. 1213–1222, 2006.

[8] M. Iida, R. Asada, Y. Zhang, M. Suguri, and R. Masuda, "Obstacle detection and avoidance for autonomous combine harvester," in *Proc. ASABE Annu. Int. Meeting*, 2018, pp. 1–4, Paper 1800737.

[9] Y. Bao, D. S. Shah, and L. Tang, "3D perception-based collision-free robotic leaf probing for automated indoor plant phenotyping," *Trans. ASABE*, vol. 61, no. 3, pp. 859–872, 2018.

[10] G. Pajares *et al.*, "Machine-vision systems selection for agricultural vehicles: A guide," *J. Imag.*, vol. 2, no. 4, p. 34, 2016.

[11] B. Hamner, S. Singh, and M. Bergerman, "Improving orchard efficiency with autonomous utility vehicles," in *Proc. ASABE Annu. Int. Meeting*, 2010, pp. 1–9, Paper 1009415.

[12] P. M. Blok, K. van Boheemen, F. K. van Evert, J. IJsselmuiden, and G.-H. Kim, "Robot navigation in orchards with localization based on particle filter and Kalman filter," *Comput. Electron. Agricult.*, vol. 157, pp. 261–269, Feb. 2019.

[13] D. Kohanbash, M. Bergerman, K. M. Lewis, and S. J. Moorehead, "A safety architecture for autonomous agricultural vehicles," in *Proc. ASABE Annu. Int. Meeting*, 2012, Paper 12-1337110.

[14] S. Thrun *et al.*, "Stanley: The robot that won the DARPA grand challenge," *J. Field Robot.*, vol. 23, no. 9, pp. 661–692, 2006.

[15] M. Dickson, N. Noguchi, Q. Zhang, J. F. Reid, and J. D. Will, "Sensor-fusion navigator for automated guidance of off-road vehicles," U.S. Patent 6 445 983 B1, Sep. 3, 2002.

[16] B. Benet and R. Lenain, "Multi-sensor fusion method for crop row tracking and traversability operations," in *Proc. Conf. AXEMA-EurAgEng*, 2017, pp. 1–11.

[17] F. Rovira-Más, V. Saiz-Rubio, and C. Millot, "Sonar-based aid for the execution of headland turns by a vineyard robot," in *Proc. ASABE Annu. Int. Meeting*, 2016, pp. 1–9, Paper 162456431.

[18] G. L. Andersen and T. Kuennen, "CULTI CAM HD-efficient weed control in row crops using active implement steering and stereo camera," in *Proc. Land Technik AgEng Conf.*, 2019, pp. 455–461, Paper 2361.

[19] Vitibot. (2018). *Bakus, the Robot That Solves the Problems of Sustainable Viticulture*. Vitibot Corp., Reims, France. [Online]. Available: http://vitibot.fr/en/bakus

[20] P. McCorduck, *Machines Who Think*. Natick, MA, USA: A. K. Peters, Ltd., 2004, pp. 243–276.

[21] F. Rovira-Más, S. Han, and J. F. Reid, "Trajectory-based algorithm for the quantitative evaluation of automatically steered vehicles," *Trans. ASABE*, vol. 51, no. 5, pp. 1545–1552, 2008.

[22] M. Minsky, *The Society of Mind*. New York, NY, USA: Simon & Schuster, Inc., 1988, pp. 202–203.

Francisco Rovira-Más graduated in agricultural engineering from the Universitat Politècnica de Valencia, Spain, in 1996. He received the Ph.D. degree in agricultural engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 2003.

He has been a member of the Intelligent Vehicles System Group, John Deere Technology Center, Moline, IL, USA. He is currently the Director of the Agricultural Robotics Laboratory and a Professor with the School of Agricultural and Biological Engineering, Universitat Politècnica de València, Spain. His work for industry resulted in six U.S. patents on machine vision, stereoscopic 3-D perception, and the fusion of optical and GPS-based field data. He is the leading author of the book *Mechatronics and Intelligent Systems for Off-Road Vehicles* (London, U.K., Springer-Verlag, 2010). His research interests include mechatronics, agricultural robotics, precision agriculture and machine vision, and stereoscopic vision.

Prof. Rovira-Más is a member of ASABE and the Chair of the Section III of CIGR (International Commission of Agricultural Engineering). He has received four ASABE paper awards. He is the Coordinator of the International Research Project VineScout funded by the European Union.

Verónica Saiz-Rubio received the B.Sc. degree in agricultural engineering and the Ph.D. degree in automation, robotics and industrial computing from the Universitat Politècnica de València, in 2008 and 2013, respectively, and the B.Sc. degree in mechanical engineering from Linnaeus University, Sweden, in 2009.

Her main research interests include digital farming, field monitoring, and Agriculture 5.0. She is a member of ASABE. She was awarded the Edmund Optics Research Award in 2011, the CIGR Armand Blanc Prize for young researchers in 2012, the European Farming by Satellite Prize in 2013, and the ASABE Meeting Paper Award in 2015. Since 2016, she has been the Project Manager of EU-Funded Project VineScout.

Andrés Cuenca-Cuenca received the B.Sc. degree in mechanical engineering from the Universitat Politècnica de València in 2012.

He has published several papers on agriculture automation, and holds a national patent (P201531507) for a spraying machine adapted to orchards, and two utility models on smart spraying. He is in charge of mechatronics design and sensing architecture for the VineScout Project. His research interests include mechatronics, robotics, agricultural equipment automation, sensors, and precision farming.