Stereovision system for estimating tractors and agricultural machines transit area under orchards canopy

Corrado Costa[1*], Paolo Febbi[2], Federico Pallottino[1], Massimo Cecchini[2], Simone Figorilli[1], Francesca Antonucci[1], Paolo Menesatti[1]

(1. *Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), Centro di ricerca Ingegneria e Trasformazioni Agroalimentari, Via della Pascolare 16, 00015 Monterotondo scalo, Rome, Italy*;
2. *Department of Agriculture and Forestry Science, Tuscia University, Via S. Camillo de Lellis snc, 01100 Viterbo, Italy*)

Abstract: Managing orchards requires delicate agricultural operations being typically carried out in narrow zones where the operators usually drive machineries under stress that could result in poor performance. In such conditions, the use of technology would help manage the machines to reduce the hazardous work and eventual damage to the plants. To safely drive a tractor, the driver needs to be aware of its surroundings, thus a stereovision system can provide helpful information. Stereo imaging has proven to be an effective three-dimensional vision system. Indeed, the range (or third coordinate) information is useful to detect the obstacle distances. Such distances, when detected during agricultural operations, could be used to assist the operator in driving the tractor at regular or variable working speeds and eventually to provide manufacturers useful indications to model the form of ROPS (roll over protection structure). This study aimed to verify the closeness of agreement between manual and stereo-image measurements, and thus to provide helpful information regarding safety and working purposes. The system used a custom low-cost dual web-camera in combination with an image analysis algorithm in order to automatically extract the information needed. Manual independent measurements were carried out using a metric tape (sensitivity 1 cm). A regular structure was used for the analysis: four rows of ten trees each one. Alternated red and blue paper markers were placed on the hazelnut trees (two per tree) of two couples of rows for enhanced visibility. For each couple of trees (one on the right, the other on the left), the four markers formed a trapezoid that was measured. The results of the analysis demonstrated that the stereo vision provided distance measurements with reasonable accuracy (error <5%) in the range of distances lower than 20 m. The resolution assessed for the developed video system is suitable for obtaining distance information in real scenes. This information could be used to assist drivers to operate agricultural machineries through narrow tree rows during work execution. Moreover, such information could be used for safeguarding decision-making and/or for controlling some tractor functions such as continuing moving, changing driving direction, changing 3-point hitch position, reducing transmission speed, halting the tractor. These functions will be necessary before tractors become fully autonomous. Finally, the measured distances, marking the narrow transitions between the tree rows, could be also used to study the ROPS form, both for working safely and for avoiding possible damage caused to the hazel trees laterally.

Keywords: stereovision, precision agriculture, digital agriculture, hazelnut, canopy, ROPS, orchards
DOI: 10.25165/j.ijabe.20191201.4123

Citation: Costa C, Febbi P, Pallottino F, Cecchini M, Figorilli S, Antonucci F, et al. Stereovision system for estimating tractors and agricultural machines transit area under orchards canopy. Int J Agric & Biol Eng, 2019; 12(1): 1–5.

1 Introduction

To move a tractor safely, the driver needs to be aware of its surroundings. Stereovision systems can nowadays provide helpful information for the purpose. Indeed, stereoscopic vision provides a richer representation of reality than the conventional two-dimensional images. Stereo imaging has already proven to be an effective three-dimensional vision system[1,2].

Cultivation of fruit trees is a delicate agricultural operation because it is typically carried out in narrow zones. The operator usually works under stress and fatigue conditions that can result in poor performance. In this case, a logical use of technology would be helpful in managing the machines since it would relieve the operator from hazardous work. Experience has shown that

Received date: 2018-01-16 **Accepted date:** 2018-09-04

Biographies: Paolo Febbi, PhD, Engineer, research interests: Agricultural mechanization, Email: p.febbi@alice.it; **Federico Pallottino,** PhD, Researcher, research interests: precision farming, digital agriculture, IoT applications, opto-electronic applications, image analysis, Email: federico.pallottino@crea.gov.it; **Massimo Cecchini,** Associate Professor, research interests: agricultural mechanization, ergonomics, health and safety in workplaces, Email: cecchini@unitus.it; **Simone Figorilli,** Bachelor, Technician, research interests: image processing, open source proximal and remote sensing technologies, precision farming, IoT network and LoRa devices, Email: simone.figorilli@crea.gov.it; **Francesca Antonucci,** PhD, Researcher, research interests: rapid and non-invasive techniques (e.g., optoelectronic, imaging, stereovision, thermography), multivariate statistics, precision agriculture, qualitative evaluation of agri-food products, Email: francesca.antonucci@crea.gov.it; **Paolo Menesatti**, PhD, Senior Scientist, research interests: digital engineering technology applications, numerical methodologies in biosystems, agro-food and agro-environmental systems (ICT, imaging, RFiD, spectrometry, proximal sensing, thermography, inferential or predictive multivariate modelling, open source hardware and software), Email: paolo.menesatti@crea.gov.it.

***Corresponding author: Corrado Costa,** PhD, Researcher, research interests: digital agriculture, multivariate statistics, imaging sensors for proximal sensing, IoT networks. Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA), Centro di ricerca Ingegneria e Trasformazioni agroalimentari, Via della Pascolare 16, 00015 Monterotondo scalo, Rome, Italy. Tel: +39 (0)6 906 75214; Fax: +39 (0)6 906 25591, Email: corrado.costa@crea.gov.it.

automatic guidance using Global Navigation Satellite Systems can be difficult to achieve inside the tight arrangement of the tree rows, unless local positioning sensors take part in the localization unit of auto-steered machines[3]. Stereo vision using dual cameras provides extra range information of objects from acquired images that gives the viewer the feeling of depth. It can be used for measurement and analysis of existing objects from acquired field images[4]. The range (or third coordinate) information is especially useful to detect the obstacle distances. With appropriate calibration and registration procedure, range information can be calculated from image pairs with reasonable accuracy[5].

Traditionally, stereo vision uses two cameras disposed horizontally used in order to obtain two views of the same scene reproducing indeed the human vision. Consequently, to the synchronous acquisition, the images of each pair are compared to extract the relative depth information that are consequently expressed through the so-called disparity map. This map ciphers the differences in horizontal coordinates of corresponding image pixels. The resulting values are inversely proportional to the depth of the scene for each pixel[6-8]. The aim of this study was to verify the closeness of agreement between manual and stereo-image measurements in order to provide helpful information regarding safety and working purposes. In this approach, a custom dual web-camera system was used in combination with an image analysis algorithm to automatically extract the information needed. The obstacle distances detected during agricultural operations could be used to assist the operator in driving the tractor at regular or variable working speeds and in addition to provide manufacturers useful indications to model the form of ROPS (roll over protection structure)[9]. Indeed, one of the most important risk factor inherent workers' safety within agricultural frameworks, is represented by tractor rollover that in Italy causes more than 120 death cases per year. In orchards, because of vertical clearance constraints, workers often prefer working with downed ROPS, strongly enhancing the risk of crashing in case of tractor rollover. The use of fixed and compact ROPS, specifically designed and adapted to work under canopy, would considerably reduce the risk of crushing in the event of rollover. The proposed solution could provide, in near real-time, a good degree of detail given by local perception and could be easily integrated on any agricultural machine.

2 Materials and methods

2.1 Experimental setup

Stereo measurements were acquired in an experimental hazelnut field in Soriano nel Cimino (Viterbo, Central Italy; 42°24′58.32″N, 12°15′55.8″). Figure 1 shows an actual scene used to test the video system: a hazelnut grove consisting of hazel trees arranged in equally spaced rows (6 m) separated evenly at known distance (6 m).

Note: Left and right pictures represents images from left and right cameras, respectively.

Figure 1 Actual scene used to test the video system

The tree branches are inclined at certain angles, varying from tree to tree following biological variability and pruning, producing a kind of Δ-shape tunnel. There is, therefore, a priori knowledge about the scene to be acquired by the stereo camera.

A regular structure was considered for the analysis: four rows of ten trees each one. Alternated red and blue 10 cm×10 cm markers, of colored paper, were placed on the hazelnut trees of two couples of rows, in the field of view of the cameras. Two markers were positioned on each tree, for a total equal to 80 markers. For each couple of trees (one on the right, the other on the left), the four markers formed a trapezoid (Figure 2). The extent (*i.e.*, the total length included in the study) was 62×2 m (10 couples of trees × 2 lines).

Manual independent measurements were carried out using a metric tape (sensitivity 1 cm), so that no observed value had any influence on any other. The following distances were acquired: distance between lower markers (major base of the trapezoid, AB); distance between top markers (minor base of the trapezoid, CD); distances between markers of the same tree (oblique sides, BC and DA); distances between top and lower markers of opposite trees (diagonals of the trapezoid, AC and BD); distances between corresponding markers of contiguous trees belonging to the same row (marker couples corresponding to the same vertex of contiguous trapezoids, A_nA_{n+1}, B_nB_{n+1}, etc.). The collected values provided a sequence of observations ordered along a spatial axis. The number of observations in the data series was 100 (in the *XY* plane) + 72 (*Z* coordinate) and the interval (distance) between successive observations was 6 m.

Note: Red and blue trapezoids pictures the target points at the same distance layer (*i.e.*, every 6 m); their vertices represents the measure points.

Figure 2 A hazelnut grove consisting of hazel trees arranged in equally spaced rows (6 m) separated evenly a known distance (6 m)

The stereovision system was positioned between each couple of tree lines. Vegetation, alternate to the wooden structure, limited the camera visibility and produced a variable illumination of the experimental field. Because of shadows, changes in environmental illumination affected the marker visibility. However, the hazelnut grove was well illuminated and disparity images with an extensive coverage were typically produced. Indeed, several couple of trees were visible, all members of the two determined rows and various irregular trapezoids were correctly detected by the cameras in the same scene (see Figure 2). Then, the stereovision system was manually moved forward, along the rows, in steps of 6 m. For each couple of rows, ten couples of images were acquired, moving forward each 6 m, in order to cover a complete visible path for each location and thus be able to correctly acquire measurements.

For each image, four neighboring couples of trees (8 m, 14 m, 20 m and 26 m from the video system) were considered for measurements. Finally, a total of 34×2 trapezoids were obtained and measured. All data were processed to create a database comparable with the manually acquired measurements.

Four couple of hazel trees limited the range (Z coordinate) to a reasonable interval (26 m); this allowed to consider a space of logical placement and to neglect positions outside that space.

A problematic situation arose when branches and leaves blocked part of the field of view and the markers were not always properly detectable in the images causing difficulties of unique identification. In those cases, the meaningless measures were discarded.

Because the trees did not block the view of the neighboring rows, it was not possible to completely avoid the detection of multiple rows. Thus, some markers of the adjacent rows were visible in the images but were not considered in the analysis.

The desired and expected response of the algorithm was an information with the properties of right distances between markers/branches and right spacing for trees (depth information).

Considering the agricultural scenes, the stereo video system was mounted as parallel as possible to the targeted scene (tilt <5°) and positioned 8 metres from the first couple of the considered hazel trees. The image plane results perpendicular to the ground (or forming certain inclination angle with respect to the ground). In this type of images, the effect of perspective is evident.

2.2 Stereo vision measurements

The basic operation in stereovision is the correlation between the left and the right image. For a given scene, the stereo camera captures the stereo pair of images (left and right) and send them out to the computer. Before software analysis, it is necessary to acquire satisfactory stereo images in order to produce acceptable disparity images, which are then used to determine the camera coordinates of the captured points in the 3D scene. The use of stereo-images allows the measurement of the distance between two points using a triangulation of the three-dimensional coordinates[7,11]. The basic principle to extract the distance between the cameras and a target is to estimate the horizontal position difference of the target on the two images acquired with the dual cameras. This difference is defined as the disparity of the target in a stereo image pair and is used to calculate the depth information of the target[4]. The origin of the camera system (of coordinates X, Y, Z) is set at the optical center of the left lens. The X-Y plane is coincident with the image plane; the Y coordinate gives the height of the selected point; the Z coordinate represents the range or camera-object distance.

The used stereo vision system was developed and described by Menesatti et al.[7] Digital stereo-images for the dual video analysis were acquired using two low cost WebCam Logitech C920 hdpro (mounting Carl Zeiss lens and acquiring images till 15 megapixel and video in Full HD resolution of 1920×1080 pixels). The two cameras were connected, synchronized and powered via USB cable with a laptop and managed in Matlab 7.1 environment by a specific software engineered for this kind of application. The delay between synchronous images acquisition was around 1500 ms. The cameras were mounted on a 150 cm aluminium bar, equipped with bubble level, mounted on a tripod at a relative distance of 119 cm (camera separation). The cameras were perfectly vertically aligned to avoid the use of image rectification algorithm. Focal length angle of the cameras and convergence angle was measured with a geometric procedure and resulted to be 68.843° (1400.9 pixel at the optical resolution) and –1.277° respectively. A more acute convergence would decrease the useable field of view, but improve the measurement precision[12]. The stereo video system, at a distance lower than 10 m showed mean errors always lower than 2%[7].

A Partial Least Square algorithm was used to post-process distance data from triangulation procedure. Given the images positioning of the two points (x_{1left}, y_{1left}, x_{2left}, y_{2left} and x_{1right}, y_{1right}, x_{2right}, y_{2right}), space positioning data (x_1, y_1, z_1 and x_2, y_2, z_2) and the calculated Euclidean distance, the PLS algorithm was trained to return corrected distances. PLS is a soft modelling method for constructing predictive models with many and highly collinear factors[13-15]. All the triangulation and PLS correction has been implemented in Matlab 7.1 environment by another specific software engineered for this kind of application. The camera coordinates are then transformed into conventional length units.

2.3 Error analysis

The error (%) was estimated as:

$$\text{Error (\%)} = \frac{|\text{Stereo measurement} - \text{Real measurement}|}{\text{Real measurement}} \times 100$$

Shapiro-Wilk normality tests were conducted on the different repetitions of distances and x, y, z measurements. Non-parametric Kruskal-Wallis tests were conducted to test the null hypothesis of equal medians among different repetitions of distances and x, y, z measurements.

3 Results

Table 1 shows the comparison among the mean (±SD) positions (x, y, z) and distances. Shapiro-Wilk normality tests conducted on the different repetitions of distances and x, y, z measurements rejected the H_0 null hypothesis of normality (p<0.05). For this reason, non-parametric Kruskal-Wallis tests were conducted to test the null hypothesis of equal medians different repetitions of distances and x, y, z measurements. Results of the tests are showed in Table 1. On the x axis the mean percentage error significantly decreases with increasing distance and are always lower than 3.4%. On the y and z axes the percentage error significantly increase with increasing distance after 14 m; below 14 m mean percentage errors are always lower than 3.8%, below 26 m mean percentage errors are always lower than 7%. At 8 m and 14 m non-significant differences were shown among positions. At 20 m and 26 m significant differences were shown among x and y-z positions.

Table 1 Comparison of mean (±SD) percentage errors at different positions (*x*, *y*, *z*) and distances

m

Distances	Positions			
	x	*y*	*z*	Average
8	$3.4\pm2.5^{a}_{a}$	$3.1\pm2.9^{a}_{a}$	$3.1\pm3.4^{a}_{a}$	$3.2\pm2.9_{a}$
14	$2.7\pm1.4^{a}_{ab}$	$3.8\pm3.3^{a}_{ab}$	$3.8\pm3.8^{a}_{ab}$	$3.4\pm3.0_{a}$
20	$2.2\pm1.6^{a}_{bc}$	$4.8\pm4.5^{b}_{b}$	$7.0\pm6.0^{b}_{b}$	$4.6\pm4.8_{a}$
26	$2\pm1.6^{a}_{c}$	$6.7\pm7.7^{b}_{b}$	$2.4\pm6.6^{b}_{b}$	$3.7\pm5.3_{a}$
Average	2.6 ± 1.9^{a}	4.4 ± 4.7^{b}	4.4 ± 4.7^{b}	3.7 ± 3.9

Note: The subscript letters indicate the Kruskal-Wallis tests significance ($p<0.05$) results. Different letters represent statistically significant differences according to the Kruskal-Wallis tests for values of $p\leq0.05$; subscript letters represent significance among distances, superscript letters represent significance among positions.

4 Discussion

The quality of the results depends to a large extent on the marker visibility and cursor positioning. The images representing the targeted scene usually yielded the maximum resolution that a particular stereovision system can produce. Nevertheless, for some applications less resolution is preferred, for example, to get a higher processing speed, even though that implies a less detailed representation of reality[1]. A vision system requires to record position and dimension of obstacles to be avoided and does not need to include very small objects. However, the resolution of an image (number of pixels) can be a critical factor because the measured values depended on the cursor position during distance analysis. Indeed, even if the markers were clearly identified, distance values between them changed moving the cursor from one pixel to the next. Likewise, the distance dual camera-markers was important because pixels of far markers, with a lower resolution capability and/or erroneous stereo information, led to coarse measurements. The error estimations were reported in Table 1. The series exhibit irregular and unpredictable fluctuations (noise) which are due to non-permanent perturbation factors.

The result analysis demonstrated that stereo vision provided distance measurements with reasonable accuracy (error <5%) in the range at distances lower than 20 m, defined in Table 1. The result shows as the developed video system is suitable for obtaining distance information in real scenes. This information could be used to assist agricultural machine drivers in the monotonous task of guiding a tractor through rows of trees for the execution of agricultural works, such as ground milling, grass shredding, nut harvesting, harrowing, etc.

In the case of agricultural fields where people and heavy machines coexist, safety is of primary concern. The registration of branch locations, their heights and tree positions are helpful to obtain information on agricultural fields, for safety agricultural applications, because they add details with respect to possible lateral hazards to the tractor. Stereo images can provide the distances from a hypothetical vehicle to possible obstacles; so, objects can be detected within a certain boundary around the tractor, in near real time. These information could be used for safeguarding decision-making and/or for controlling some tractor functions such as continue moving, change driving direction, change 3-point hitch position, reduce transmission speed, halt the tractor. These functions will be necessary before tractors become fully autonomous[16].

Furthermore, considering the position of the tree branches and the spacing between rows, it is possible to estimate the offset (lateral misalignment) of a tractor. The offset was defined[3] as the lateral distance between the actual position of the vehicle and its optimal position, determined by the location of operation points. The computed offset could be used to delimit a potential trajectory through the path, avoiding the obstacles located on both sides.

The ROPS shape and dimensions are determinant for the offset estimation. The measured distances, marking the narrow transitions between the tree rows, could be also used to study the ROPS form, both for working safely and for avoiding possible damage caused to the hazel trees laterally.

5 Conclusions

The study of the interested area under orchards canopy enables the design of non-removable and compact protective structures (ROPS) specifically designed for orchard tractors, in accordance with a suitable “safety volume” around the operator. These structures should allow the transit of the machine under canopy without damaging branches and trees, while, at the same time, safeguarding the operator. This goal can be reached adopting protective structures with profiles designed to facilitate the sliding of twigs and branches during tractor transit under the canopy.

The elaborated methodology was based on the detection of markers, signaling the overhanging branches of hazel trees, using stereoscopic vision as perception method. Next step is to recognize the branch splitting without adding markers or modifying the current structures of the field, even if more tests are necessary to optimize an algorithm that can work in variable conditions for a safer guidance of tractors.

Acknowledgment

Part of this study was funded by the project “ALForLab” (PON03PE_00024_1) co-funded by the Italian National Operational Programme for Research and Competitiveness (PON R&C) 2007-2013, through the European Regional Development Fund (ERDF) and the national resource Revolving Fund - Cohesion Action Plan (CAP) MIUR and project AGROENER (D.D. n. 26329) funded by Italian Ministry of Agriculture, Food and Forestry Policies (MiPAAF). Authors would like to thank Gianluca Febbi for helping in the field operations and manual measurements. All the authors equally contribute to the writing of the paper and to its content.

[References]

[1] Rovira-Más F, Reid J F, Zhang Q. Stereovision data processing with 3d density maps for agricultural vehicles. Trans. ASABE, 2006; 49(4): 1213–1222.

[2] Rovira-Más F, Zhang Q Reid J F. Stereo vision three-dimensional terrain maps for precision agriculture. Comp. El. Agr., 2008; 60: 133–143.

[3] Rovira-Más F, Han S, Wei J, Reid J F. Autonomous guidance of a corn harvester using stereo vision. Agricultural Engineering International: the CIGR Ejournal, 2007; 9: ATOE 07 013.

[4] Lin T T, Hsiung, Y K, Hong G L, Chang H K, Lu F M. Development of a virtual reality GIS using stereo vision. Comp. El. Agr., 2008; 63: 38–48.

[5] Shapiro L G, Stockman G C, 2001. Computer Vision. Upper Saddle River (NJ): Prentice-Hall, Inc.

[6] Costa C, Scardi M, Vitalini V, Cataudella S. A dual camera system for counting and sizing Northern Bluefin Tuna (*Thunnus thynnus*; Linnaeus, 1758) stock, during transfer to aquaculture cages, with a semi automatic Artificial Neural Network tool. Aquaculture, 2009; 291(3-4): 161–167.

[7] Menesatti P, Costa C, Antonucci F, Steri R, Pallottino F, Catillo G. A low-cost stereovision system to estimate size and weight of live sheep. Comp. El. Agr., 2014; 103: 33–38.

[8] Pallottino F, Steri R, Menesatti P, Antonucci F, Costa C, Figorilli S, et al. Comparison between manual and stereovision body traits measurements of Lipizzan horses. Comp. El. Agr., 2015; 118: 408–413.

[9] Cecchini M, Monarca D, Colantoni A, di Giacinto S, Menghini G, Longo L. Study on the possibility of application of a compact roll over protective structure for agricultural wheeled narrow track tractors. J. Agr. Eng., 2013; 441S.

[10] INAIL, Open data. Banca dati statistica. http://dati.inail.it/opendata/default/Qualidati/index.html (last access July 2017).

[11] Wu J, Tillett R, McFarlane N, Ju X, Siebert J P, Schofield P. Extracting the three-dimensional shape of live pigs using stereo photogrammetry. Comp. El. Agr., 2004; 44(3): 203–222.

[12] Harvey E, Shortis M. A system for stereo-video measurement of sub-tidal organisms. Mar. Technol. Soc. J., 1995; 29(4): 10–22.

[13] Wold H. Estimation of principal components and related models by iterative least squares. Multivar. Anal., 1966; 1: 391–420.

[14] Costa C, Antonucci F, Boglione C, Menesatti P, Vandeputte M, Chatain B. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquac. Eng., 2013; 52: 58–64.

[15] Febbi P, Menesatti P, Costa C, Pari L, Cecchini M. Automated determination of poplar chip size distribution based on combined image and multivariate analyses. Biomass & Bioenergy, 2015; 73: 1–10.

[16] Han S, Zhang Q. Map-based control functions for autonomous tractors. ASAE 2001; 011191. St. Joseph, MI.