Available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/issn/15375110

Research Paper

Design parameters for adjusting the visual field of binocular stereo cameras

Francisco Rovira-Más [a,*], *Qi Wang* [b], *Qin Zhang* [b]

[a] *Departamento de Mecanización y Tecnología Agraria, Polytechnic University of Valencia, Campus Camino de Vera, 46022 Valencia, Spain*
[b] *Department of Agricultural and Biological Engineering. University of Illinois at Urbana-Champaign, 1304 W. Pennsylvania Avenue, Urbana, IL 61801, USA*

ARTICLE INFO

Article history:
Received 19 August 2008
Received in revised form
25 August 2009
Accepted 18 September 2009
Published online 29 October 2009

Stereoscopic cameras are becoming fundamental sensors for providing perception capabilities for automated vehicles; however, they need to be adequately setup to avoid excessive data processing and unreliable outcomes. Combinations of baselines and lens focal lengths were optimised to adjust the field of view of a stereo camera to provide the two fundamental perceptions required for intelligent vehicles: safeguarding distances around 6 m and look-ahead distances up to 20 m for automatic guidance. The main objective was to develop a systematic procedure to find the parameters that best sense the desired field of view. Quantitative indices to estimate perceptive quality, such as relative errors and efficiencies, were defined and applied to particular cases. Experiments, both in the laboratory and outdoor, led to the conclusion that short ranges under 6 m from the vehicle were best acquired with 8 mm lenses and baselines ranging from 100 mm to 150 mm, whereas 200 mm baselines coupled with 12 mm and 8 mm lenses were more suitable for longer look-ahead distances. These experiments also proved the utility of the methodology proposed.

1. Introduction

The automation of agricultural machinery is paving the way for the development of agricultural robotics. A crucial milestone in this process has been the widespread use of Global Navigation Satellite Systems (GNSS) for civilian applications. The availability of global localisation information for agricultural vehicles has resulted in the commercialisation of precision agriculture and auto-steering driving systems. However, field practices have proven that agricultural machinery requires the assistance of local sensors for safe and accurate path tracking. Typical satellite navigation errors such as bias, drift, double path, or signal drop due to dense tree canopies in orchards cause a temporary lack of control which can result in serious accidents. Precise navigation and reliable safeguarding can only be achieved through local perception sensors such as cameras, laser rangefinders and ultrasonic devices. Monocular machine vision has been in use on agricultural vehicles for more than twenty years. It provides a two-dimensional (2D) representation of the target scenes, but it cannot reliably determine the distances at which objects are located within the field of view of the camera. Nodding lasers can supply this information but they usually have a high cost and they demand sophisticated mounting to enable simultaneous scanning in two perpendicular planes. Stereoscopic vision, on the other hand, combines the advantages of both

* *Corresponding author.*
E-mail address: frovira@dmta.upv.es (F. Rovira-Más).
1537-5110/$ – see front matter © 2009 IAgrE. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.biosystemseng.2009.09.013

Nomenclature

B	Stereo camera baseline, mm
f	Lenses focal length, mm
FF	Form factor
h_c	Stereo camera height, mm
R	Stereo range, mm
X_c, Y_c, Z_c	Camera coordinates, mm
X, Y, Z	Ground coordinates, mm
X', Y', Z'	Ground coordinates deduced from the three-dimensional point cloud, mm
ε_x	Relative error in width, %
ε_y	Relative error in range, %
ε_z	Relative error in height, %
η_{2D}	Planar efficiency, %
η_{3D}	Stereo efficiency, %
θ	Camera inclination angle, degrees

systems: a three-dimensional (3D) representation of the scene, thereby including range, at a high updating rate without the need of complex mechanical assemblies. Intelligent vehicles require real-time awareness, and this essential requirement has motivated the increase and development of stereo perception technologies for mobile outdoor applications.

There is a wide variety of agricultural machines that can benefit from stereovision perception for both in-field navigation and localisation applications; from lightweight scouting robots to massive agricultural combines. For example, a corn harvester was automated (Rovira-Más *et al.*, 2006b) with the purpose of following the cut-uncut edge of corn by processing the images acquired with an on-board stereo camera. A compact binocular camera mounted on the cutting header of a harvester was capable of guiding it by detecting the cut–uncut edge of the crop. The difficulties encountered were similar to those found by Thompson and Kagami (2005) with their humanoid robot; high calculation costs and reliability issues caused by low resolution and a limited field of view. The solution to such shortcomings appears to be in the generation of dense depth maps for localisation, navigation and obstacle avoidance. The question is how these maps can be obtained and what particular camera parameters need to be combined to get the best perception for a given application. The investigation reported here attempts to solve some of these issues. Previous research in this field showed how important the longitudinal dimension of the targeted field of view is because it is both vehicle and task dependent. Robust Autonomous Sensor Controlled All-terrain Land Vehicle (RASCAL), an intelligent all-terrain vehicle that participated in DARPA (Defences Advanced Research Projects Agency, USA) *Grand Challenge 2005* operated in a range between 5 m and 25 m, covered by a stereo camera with a baseline of 300 mm and lenses of 8.5 mm focal length to acquire images of 640×480 pixels (Kogler *et al.*, 2006). In spite of aiming at different driving situations, a highway autonomous automobile prepared by Kato *et al.* (1996) used a stereo camera of 300 mm baseline and lenses of 7.5 mm focal length to capture 512×512 pixel images and detected obstacles between 2 m and 20 m from the front of the vehicle. When a robot moves indoors, range requirements usually decrease. The humanoid robot developed by Thompson and Kagami (2005) registered data in a range interval of 1 m–3 m through 320×240 pixel images. Similarly, the mobile platform designed by Herath *et al.* (2006) for localisation and mapping in indoor environments was only required to detect ranges between 1 m and 6 m, employing a stereo system with a 90 mm baseline. In the stereo perception error analysis carried out by Chen and Xu (2005), a 525 mm baseline camera was set to acquire 320×240 pixel images. Errors were estimated in depth and height with ranges between 7 m and 12 m. What makes stereo perception appealing is the wealth of information behind every pair of stereo images. Paradoxically, stereo systems on intelligent vehicles tend to condense information in regular grids due to processing speed constrictions. RASCAL, for example, made use of 2D coarse grids whereas the automobile setup by Kato *et al.* (1996) processed the stereo information through a grid composed of $0.5\ \text{m} \times 2\ \text{m}$ cells. A non-probabilistic method of coping with grids was developed by Rovira-Más *et al.* (2006a) where the concept of 3D density and its practical application with density grids are used to manage stereo data.

The geometric principles of stereoscopic vision have been long known. However, real-time applications have become feasible only over the last ten years as a consequence of the many advances in computer science that led to the advent of compact binocular cameras. Commercial off-the-shelf stereo cameras provide a means of acquiring 3D information at a reasonable cost. Not only do they provide all the hardware needed for such data acquisition, but most of the cameras include software libraries for the integration of the camera in the complete perception system. The task of the camera typically finishes with the registration and transmission of the 3D point cloud, and it remains a mission for the perception or control algorithms to extract the useful information to be employed by the vehicle. While registration of 3D point clouds is easily carried out by commercial cameras, the manipulation and real-time processing of massive arrays of points tends to be the most delicate stage in stereo perception. Electronic noise, stereo mismatches, immense clouds, and other sources of error adversely affect the reliability of a vehicle performance. These difficulties in dealing with the point cloud can be alleviated by selecting an optimum configuration of the camera according to each application developed: an effective combination of camera baseline and lenses determines the set of ranges for which perception is optimal. A reasonable method is, therefore, to find the best configuration of the stereo system, in terms of optimal combination of baseline and focal length, for each particular situation studied before proceeding with the 3D data analysis. An attempt to systemise such choices was initiated by Rovira-Más *et al.* (2007) in a preliminary phase of this research project. In that initial work, a set of experiments were undertaken to study the best configuration of a stereo camera covering ranges between 3 m and 15 m. Combinations for two binocular cameras with

different baselines and lenses led to the conclusion that 9 cm baselines and 8 mm lenses gave the best results for ranges up to 10 m, but when objects laid between 10 m and 15 m from the camera, the 230 mm baseline and 12 mm lenses were more appropriate. In spite of the twelve combinations tried, various intermediate configurations remain untested before more general and conclusive recommendations can be extracted.

The results of the camera parameter search are specifically associated to the scene where the intelligent machine is going to navigate, unlike the methodology developed to find those results, which should be general and versatile. According to Meystel (1991), autonomous mobile robots need to be designed to accomplish two separate tasks: (1) sense obstacles, determining their shape, dimensions or position; and (2) plan a path, outputting steering commands. These two independent challenges can be easily adapted to the particular case of agricultural environments. An agricultural vehicle traversing a field autonomously must be aware of any potential obstacle interfering with its planned course: it needs to look ahead for finding its way while checking around its near surroundings for clearance. There is no fixed rule for establishing the appropriate look-ahead distance at which mobile equipment should be sensing, but, just to provide an estimate, the automatically guided tractor field-tested by Rovira-Más *et al.* (2005) set its look-ahead distance between 10 m and 20 m reaching speeds of 2.5–3.5 ms^{-1}. Given that approximate values for the targeted ranges must be chosen before conducting the search for the optimum parameters of the stereovision camera, the vehicle surroundings were assumed to be safely covered by 6 m ranges, and the guidance distance was set to twice that distance, i.e. 12 m. According to this hypothesis, the stereovision rig was studied to provide its best 3D perception at these two hypothetical range levels. Different conditions will certainly lead to different results in terms of specific camera recommendations but similar procedure and stages can be followed. The main objective of this research is to develop a generic methodology to select the best configuration of a stereovision system based on the quality of 3D perception. The results are presented as suggested configurations for the two perceptive situations studied. In addition, they provide a systematic procedure to evaluate the quality of stereo perception.

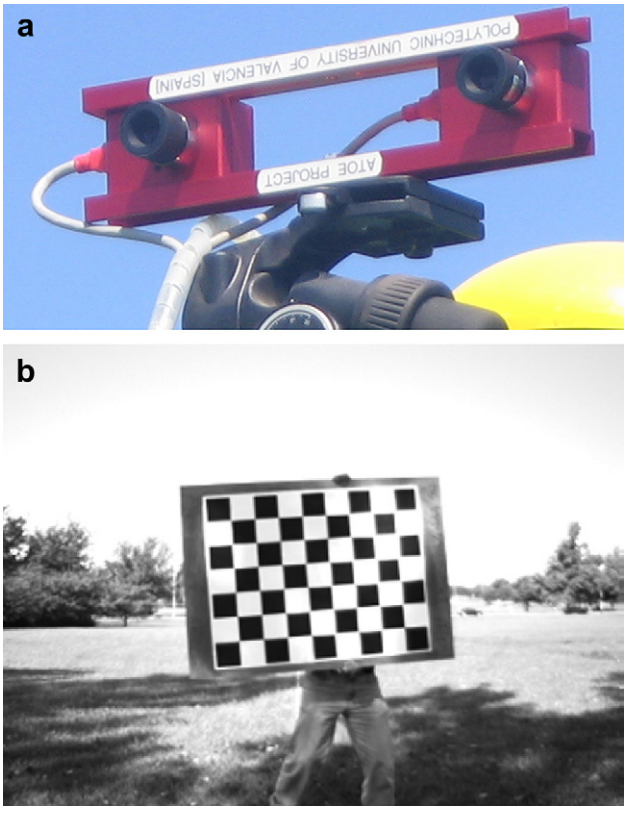

Fig. 1 – Stereo vision camera used in the experiments (a) and calibration board (b).

2. Materials and methodology

2.1. *System architecture for stereo perception*

The flexibility required to modify the distance between the centre of the lenses (baseline) and focal length of the camera lenses was achieved with a compact stereo head (Videre Design, Menlo Park, CA, USA) supporting interchangeable optics, two global shutter cameras controlled *via* the Institute of Electrical and Electronics Engineers (IEEE) 1394 bus, and a continuously variable baseline (Fig. 1a). The baseline ranged from 100 to 200 mm, and the focal lengths of the lenses mounted during the experimental phase were: 2.8 mm, 4 mm, 8 mm, 12 mm, and 16 mm. The stereo camera was operated from a mini-PC with a processor 1.8 GHz Core 2 Duo connected through two IEEE 1394 cables. The computer hosted a customised C++ program specifically designed to acquire original and disparity images, generate 3D point clouds, apply coordinate transformations, and register and display density grids for three views (top, front, and side). The diagram of Fig. 2a shows the system architecture of the stereo perception system, and Fig. 2b is a flowchart of the calculation process. This process begins with the acquisition of stereo images and finalises with the analysis and display of stereo information under two possible formats: (1) the representation of the 3D point cloud; and (2) the 3D space conversion to a discrete density grid where solid objects can be localised. Fig. 2c provides the main control panel of the customised software application especially programmed to investigate 3D stereovision perception in on-board navigation systems. Apart from allowing image and stereo controls, the graphic user interface displays the original images, the disparity map, and the density grid. As depicted in the system architecture schematic diagram of Fig. 2a, the aggregate field of view depends on the focal length of the lenses as well as on the baseline. The shortest range detectable by a camera with a given configuration is shown in Fig. 2a as the minimum range, and depends on the focal length of the lenses and the camera baseline. In contrast, the maximum range is theoretically set at infinity although there is a threshold range after which perception becomes highly unreliable. The purpose of these experiments was to determine the best camera parameters to perceive reliably inside the portion of space critical for the studied application. Every time the baseline was modified or the

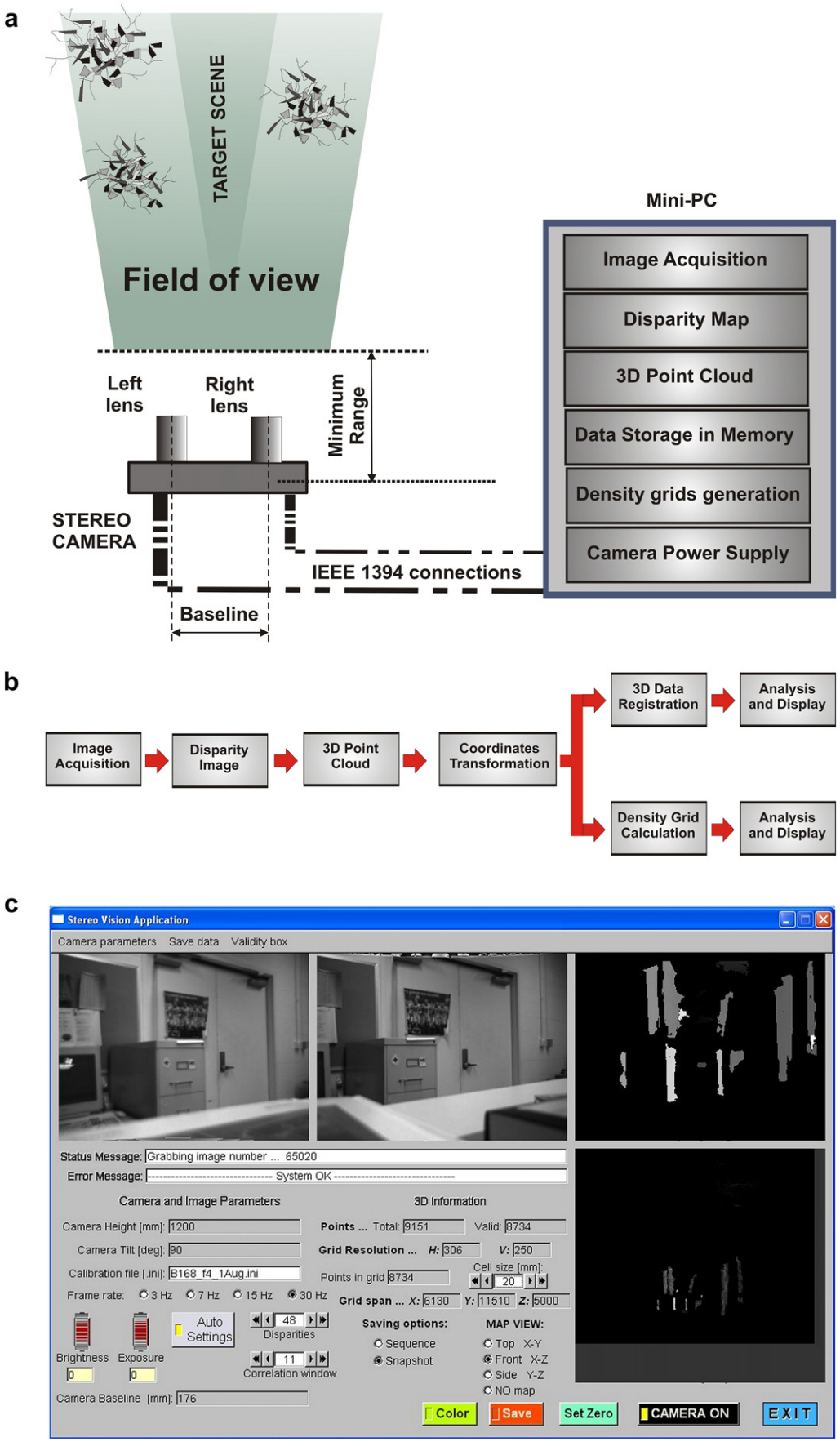

Fig. 2 – Stereo perception system: (a) Architecture; (b) Calculation process flowchart; (c) Software control panel.

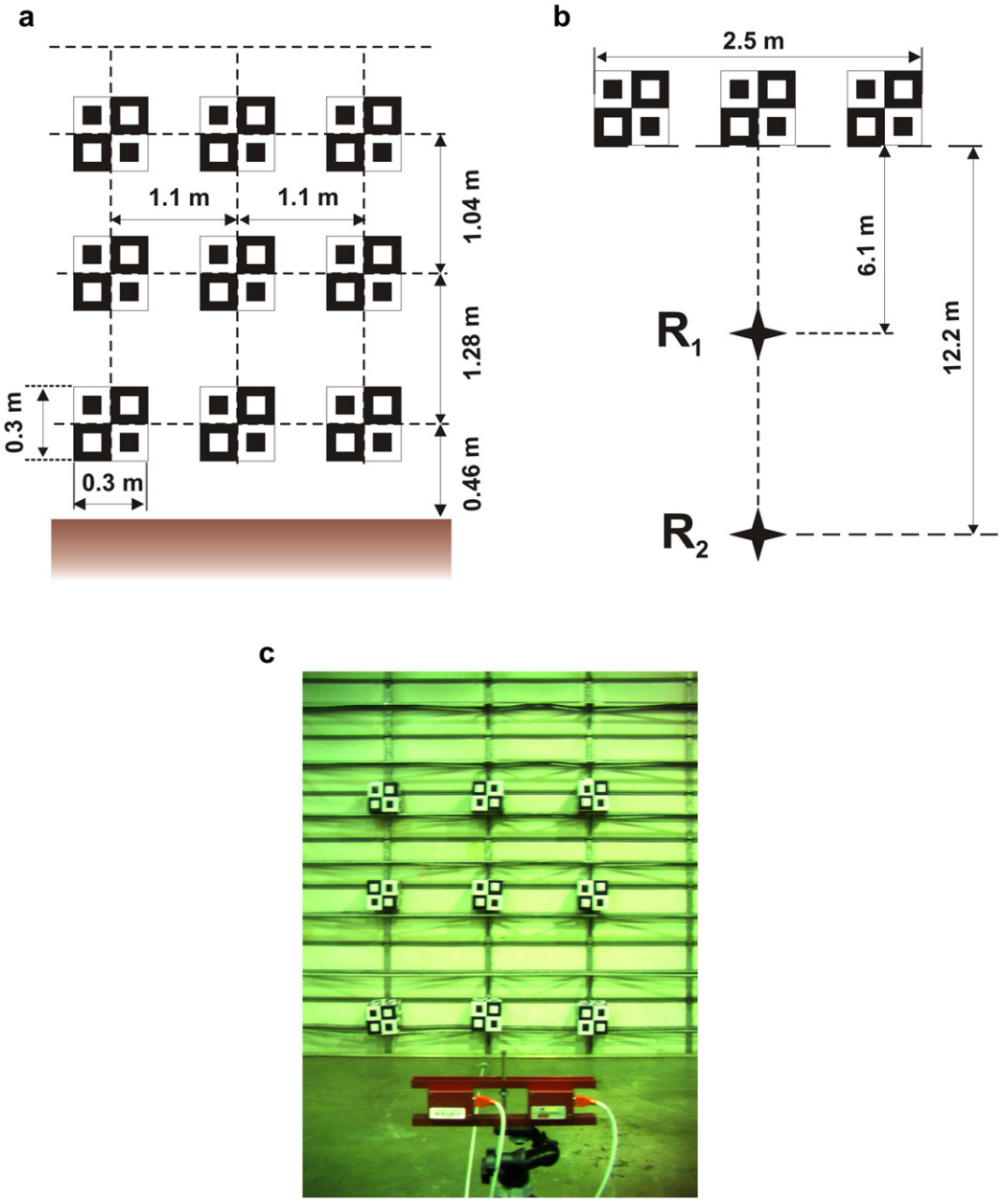

Fig. 3 – Setup for experiment 1: (a) Front view; (b) Top view; (c) Realisation of experiments.

lenses changed, the camera was calibrated. The calibration procedure consisted of taking ten stereo images of the testing board represented in Fig. 1b, and calculating the basic calibration parameters with a customised program provided by the camera manufacturer.

2.2. *Design of experiments*

The search for the best configuration of a stereo camera as the principal perception sensor of an intelligent vehicle requires the study of different parameters and their influence on the output of the camera; range accuracy, reliability in height and width estimates, effects caused by changing illumination and the impact of camera inclination angles. The investigation of these subjects resulted in two sets of tests, designated as experiments 1 and 2. The tests of experiment 1 (E1) were conducted inside the laboratory and experiment 2 (E2) was entirely performed in outdoor environments.

All the tests of E1 followed a similar procedure to the tests described by Rovira-Más *et al.* (2007). Based on this procedure, nine cardboard cubes displaying a chessboard pattern were placed in a plane in front of the camera, and parallel to the imagers of the camera. The inclination angle of the camera was set at 0° to keep the image plane perpendicular to the ground. Fig. 3a represents the exact position of the cubes within the matrix seen from the position of the camera (front

Table 1 – Specifications for experiment 1 tests.

Test	Baseline (mm)	Focal length (mm)	Range (m)	Camera height (m)
1	103	2.8	6.096	1.13
2	103	4	6.096	1.13
3	103	8	6.096	1.13
4	103	12	6.096	1.13
5	150	2.8	6.096	1.13
6	150	4	6.096	1.13
7	150	8	6.096	1.13
8	150	12	6.096	1.13
9	150	4	12.192	1.13
10	150	8	12.192	1.13
11	150	12	12.192	1.13
12	150	16	12.192	1.13
13	194	4	12.192	1.13
14	194	8	12.192	1.13
15	194	12	12.192	1.13
16	194	16	12.192	1.13

view), and Fig. 3b provides a top view of the experimental setup, where the two ranges analysed were determined by positioning the camera at points R_1 and R_2. E1 consisted of a total of 16 test runs. The main parameters involved in this series of tests are listed in Table 1.

The scenes registered inside the laboratory (experiment 1, Fig. 3c) usually presented a challenge in terms of texture and illumination. Outdoors scenes, on the contrary, were texture rich and well illuminated. Perception difficulties outdoors are often caused by a changing illumination pattern rather than the low intensity lighting obtained with artificial illumination in the laboratory. The purpose of E2 was to confirm that the best camera settings identified in E1 also produced good quality point clouds when the camera sensed outdoor scenes. The normal operating environment of an agricultural vehicle is in the field, and therefore verification outdoors is necessary. Table 2 lists the main parameters chosen for the eight tests that comprised E2. Apart from coordinate accuracy, attention was also paid to other features such as point cloud symmetry, position of reference (true) ground, effects of camera inclination angle, texture and disparity richness, noise, and presence of outliers. In order to avoid any confusion in the definition of the inclination angle θ, Fig. 4 provides an explanatory diagram of how the inclination angle and the camera height have been defined in these experiments, seen from a lateral (side view) perspective.

Table 2 – Specifications for experiment 2 tests.

Test	Baseline (mm)	Focal length (mm)	Range (m)	Camera height (m)	Camera inclination angle (°)
1	150	8	9.14	1.24	0
2	150	8	9.14	1.24	20
3	150	16	19.81	1.24	0
4	150	16	19.81	1.24	20
5	194	8	19.81	1.24	0
6	194	8	19.81	1.24	20
7	194	12	19.81	1.24	0
8	194	12	19.81	1.24	20

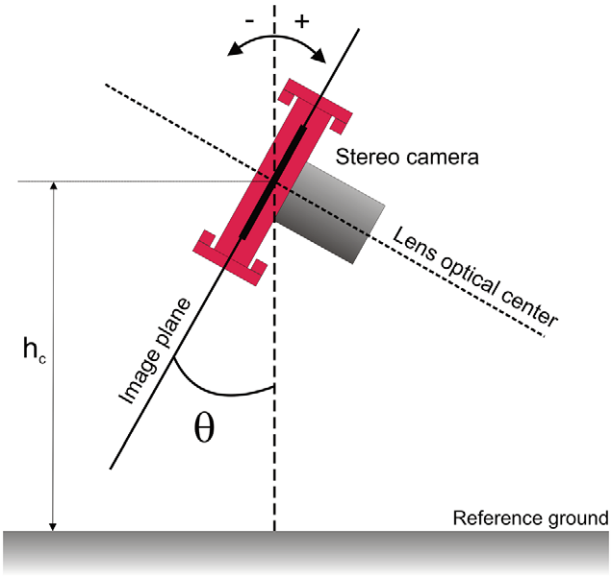

Fig. 4 – Definition of camera inclination angle (θ) and camera height (h_c) in a side view representation.

The test panel illustrated in Fig. 3 was especially built for assessing the perception quality of the camera in the two situations under study; 6 m and 12 m. Because the test panel was the same for every single test of E1, panel size did not influence the results and the best set of parameters were determined for each range level. If the panel is changed, the absolute value assigned to the perceptive quality differs, but in relative terms the best combination will continue to be the best. Assuming that a 1 m length object is to be detected by the camera at a distance of 12 m, after trying several lenses, a poor choice of parameters would result in minimum quality indices whereas adequate optics would perceive the object appropriately. The size of the cardboard boxes has been designed to represent the dimensions of the objects that need to be detected in this application. If a particular baseline-lenses combination is so poor that the test panels are not even registered in the point cloud, the correct conclusion is to reject that combination rather than changing the panel so that quality values will be increased. The ranges of interest (6 m and 12 m) as well as the dimensions of dangerous objects have been initially established. Given a reasonable set of parameters combination, the idea is to quantify their performance and select the most suitable. If the initial conditions are different, the test panel might need to be readjusted, but for off-road intelligent vehicles traversing agricultural fields, the design shown in Fig. 3 has proved to be helpful and efficient.

Every time a lens was mounted on the camera, or the baseline was changed, the stereo camera had to be calibrated and a new calibration file generated. Since it was impossible to try all possible combinations, because the baseline was continuously variable and there were five sets of lenses, a reasonable set of tests was designed. The criterion for pairing baselines and lenses was the following: long baselines and large focal lengths are known to be adequate for long ranges and *vice versa*, consequently, the pairing of short baselines with telephoto lenses was avoided in favour of more

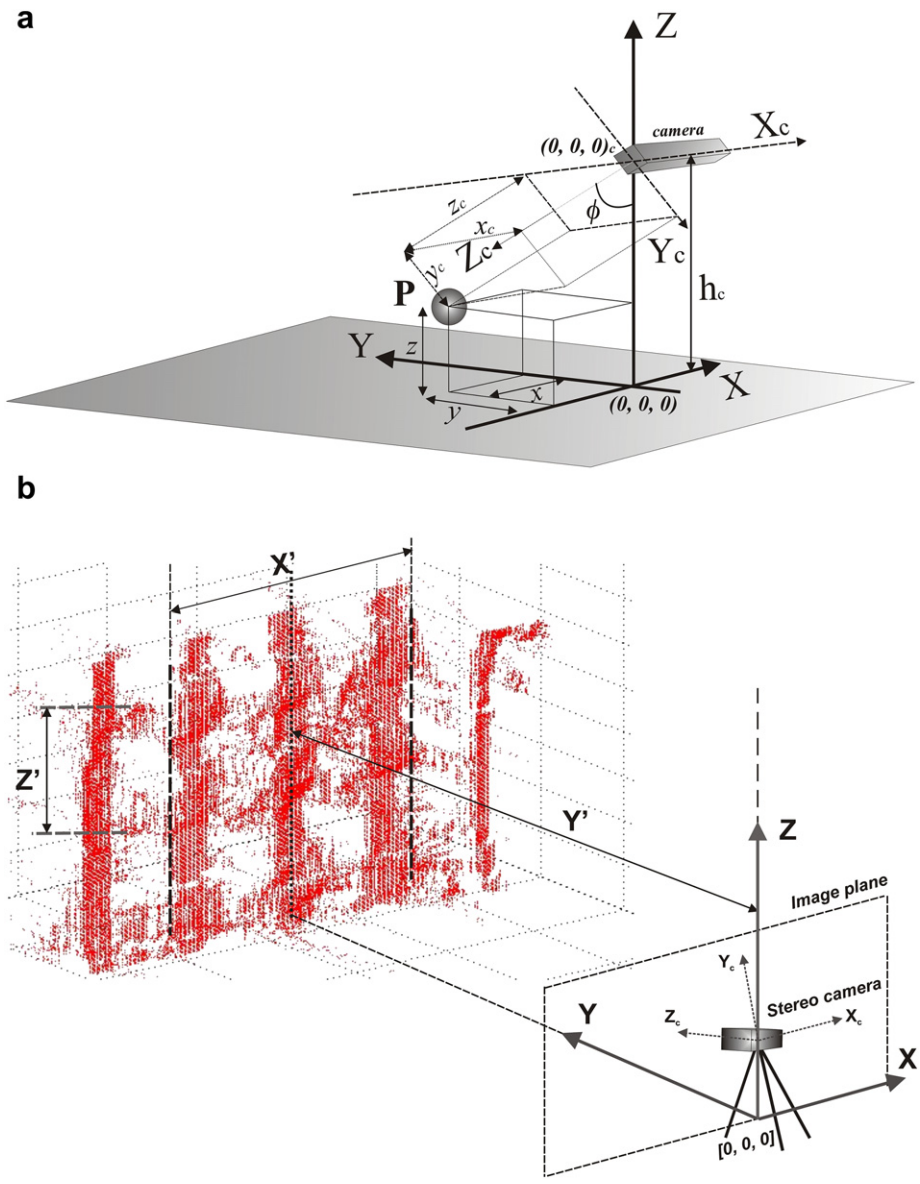

Fig. 5 – Definition of *ground coordinates* for stereo perception: (a) Transformation from camera coordinates to ground coordinates; (b) Application of ground coordinates to 3D point clouds.

logical arrangements. It is evident that modifying the baseline in small intervals will not result in drastic effects, so, for instance, combining a 12 mm lens with a 160 mm baseline and then repeating the experiment at a 150 mm baseline does not appear to be an effective experimental protocol.

2.3. *Definition of performance indicators: relative errors and efficiencies*

The perceptive information carried in every pair of stereo images was registered as a 3D point cloud, where points were localised in the camera coordinates defined by the camera manufacturer (X_cY_c plane coincides with the image plane and Z_c represents the distance object-camera along an axis perpendicular to the image plane) and depicted in Fig. 5a. It is essential to define a system of coordinates that is appropriate to the application devised, and once defined, be consistent in the transformation of coordinates. Such a system for representing the 3D information was denominated *ground coordinates system*, and it is a Cartesian frame whose origin is placed at ground level where a hypothetical vertical line passing through the centre of the camera reference imager intersects the ground, as drawn in Fig. 5a. Fig. 5b shows a point cloud generated with a binocular stereo head and represented in the newly defined system of coordinates. The ground coordinates were used to represent the 3D clouds in all the experiments. The following equation defines the transformation between the camera coordinates (X_c, Y_c, Z_c) and the ground coordinates (X, Y, Z), where h_c and θ are the camera height and inclination angle respectively, as defined in Fig. 4.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\cos\theta & \sin\theta \\ 0 & -\sin\theta & -\cos\theta \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} + h_c \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

Table 3 – Results for experiment 1.

Test	Number of points	ε_x %	ε_y %	ε_z %	η_{2D} %	η_{3D} %
1	15022	17	6	23	64	60
2	21718	4	22	13	84	66
3	29546	0	11	8	92	82
4	9966	25	2	9	68	67
5	26545	17	2	8	76	75
6	26679	11	7	1	88	82
7	16999	4	9	11	85	78
8	8797	20	5	4	77	73
9	22987	17	11	6	79	70
10	38739	11	6	5	85	80
11	19405	17	8	10	75	70
12	21781	0	6	2	98	92
13	23760	19	2	8	74	72
14	29807	7	1	2	90	90
15	13755	0	2	6	94	92
16	16724	5	5	10	86	82

The point cloud must represent the target scene as faithfully as possible, and the better match between virtual and actual scenarios the higher quality of stereo perception. However, it is, generally speaking, quite difficult and subjective to compare the quality of two point clouds based only on visual inspection; therefore, there is a need for a quantitative evaluation of the perceptive quality of a stereo camera. Such an evaluation should quantify how much more accurate one result is over another. The target matrix prepared for E1 (Fig. 3) allows for precise measurements of the three ground coordinates: *X* (horizontal distances among matrix cubes), *Y* (ranges or distances between the cubes and the camera), and *Z* (vertical distances within the matrix). The comparison between the real distances among the nine cubes and the estimated distances extracted from the 3D point clouds, usually through their three views – *XY* (top), *XZ* (front), and YZ (side) –, provided a means of computing numerical errors in the perception of 3D coordinates. The width of the test panel matrix represented in Fig. 3 is 2.5 m. It can be estimated by calculating difference in the *X* coordinate between the external edges of the side cubes, indicating how well the sensor perceived the scene for horizontal magnitudes. In order to avoid the influence of units in the assessment of coordinate accuracy, the set of relative errors defined by Eqs. (2)–(4) were defined, where ε_x represents the relative error in width, ε_y is the relative error in range, and ε_z is the relative error in height.

$$\varepsilon_x = \left|\frac{\min(\Delta X, \Delta X')}{\max(\Delta X, \Delta X')} - 1\right| \cdot 100 \tag{2}$$

$$\varepsilon_y = \left|\frac{\min(\Delta Y, \Delta Y')}{\max(\Delta Y, \Delta Y')} - 1\right| \cdot 100 \tag{3}$$

$$\varepsilon_z = \left|\frac{\min(\Delta Z, \Delta Z')}{\max(\Delta Z, \Delta Z')} - 1\right| \cdot 100 \tag{4}$$

The coordinates deduced from the 3D point cloud, directly read from the partial views of the cloud (top, front, and side), are indicated by $\underline{X}'$, Y', and Z'. The actual distances measured in the field or in the laboratory with a measuring tape are designated by *X*, *Y*, and *Z*. The relative errors defined by Eqs. (2)–(4) are nonnegative and given as a percentage.

Monocular cameras are very useful for detecting features of objects contained in a plane parallel to the image plane, but they cannot reliably estimate the distance of those objects from the camera. In contrast, stereo cameras providing surrounding awareness to intelligent machinery must

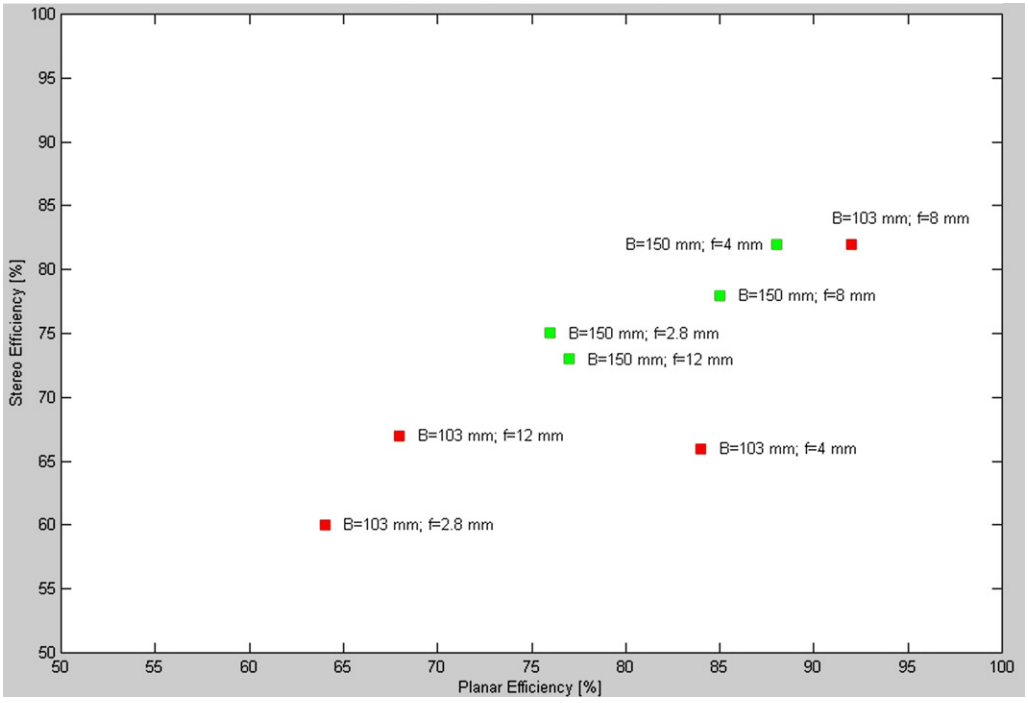

Fig. 6 – Results for short ranges studied in tests 1–8 of experiment 1, where *B* is the baseline set and *f* is the focal length of the lenses mounted on the camera.

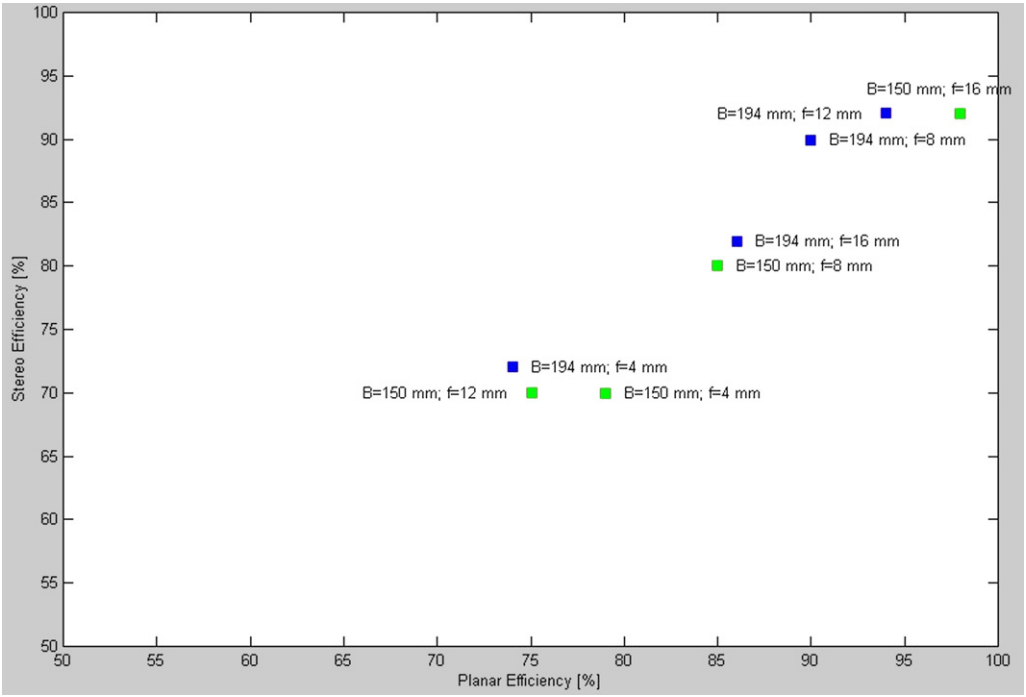

Fig. 7 – Results for medium ranges studied in tests 9–16 of experiment 1, where *B* is the baseline set and *f* is the focal length of the lenses mounted on the camera.

perform acceptably in both situations, that is to say, they must provide vehicles with reliable perception in the XZ plane as well as in the Y coordinate axis. The relative errors defined in the Eq. (2)–(4) can be combined in two equations that give a numerical value to the quality of the perception in the XZ plane and in the 3D space delimited by the field of view of the camera. Such efficiency indicators are designated as the *planar efficiency* η_{2D} and the *stereo efficiency* η_{3D}. Their mathematical formulae are given by Eqs. (5) and (6). Both indices are measured as a percentage.

$$\eta_{2D} = (1 - 0.01 \cdot \varepsilon_X) \cdot (1 - 0.01 \cdot \varepsilon_Z) \cdot 100 \tag{5}$$

$$\eta_{3D} = (1 - 0.01 \cdot \varepsilon_X) \cdot (1 - 0.01 \cdot \varepsilon_Z) \cdot (1 - 0.01 \cdot \varepsilon_Y) \cdot 100 = \eta_{2D} \cdot (1 - 0.01 \cdot \varepsilon_Y) \tag{6}$$

3. Results and data analysis

The three relative errors ε_x, ε_y, ε_z and their corresponding efficiencies η_{2D} and η_{3D} were calculated for the 16 tests listed in Table 1. Two situations for 3D awareness were studied; vehicle surroundings perceived with ranges not further than 6 m, and medium ranges of about 12 m which is considered a normal look-ahead distance for autonomous off-road driving. The objective of E1 was to find the best camera configurations for each of these two situations. Table 3 shows the results obtained from this experiment. For a given range level (6 m–12 m) the combination of baseline and focal length results in a singular stereo pair of images from which 3D information can be extracted. The studied scene is the same for all eight tests, but perception differs according to camera parameters; some combinations were more favourable (highlighted in Table 3), and produced richer 3D clouds. The accuracy of an object location theoretically depends on the relationship among baseline, focal length, range, and size of the pixel in the electronic imaging sensor (imager). However, other factors such as the quality of the lenses, the efficiency of the correlation software, and the level of noise also influence the final outcome of the camera. The accuracy of a 3D reconstructed scene can be practically quantified through the efficiencies defined in Eqs. (5) and (6). The first situation proposed (camera separated 6 m from the test panels) was investigated in tests 1–8, yielding the efficiency chart shown in Fig. 6. According to this chart, the best lens for sensing 6 m distances is 8 mm, with both baselines of 100 mm and 150 mm being adequate for that lens. The combination of 100 mm baseline and 8 mm lenses provided the highest efficiencies. A focal length of 2.8 mm resulted in a field of view that was excessively wide whereas 12 mm lenses captured a too narrow field of view. An overall comparison between the two baselines (100 mm and 150 mm) employed in tests 1–8 showed superior performance for 150 mm baselines, with efficiencies over 70% for all the lenses tried. The distance between the stereo camera and the target matrix was doubled for tests 9 through 16 of E1. Surprisingly, as demonstrated by the efficiency chart of Fig. 7, the longer range of 12 m resulted in higher efficiencies when compared to the results of Fig. 6, which indicates that longer ranges can be more accurate if the camera parameters are appropriate. The intermediate baseline of 150 mm worked well with the 16 mm lenses, and the large baseline of 190 mm was adequate for both lenses of 8 mm and 12 mm. There was a substantial loss in stereo efficiency when the 190 mm baseline was combined with the

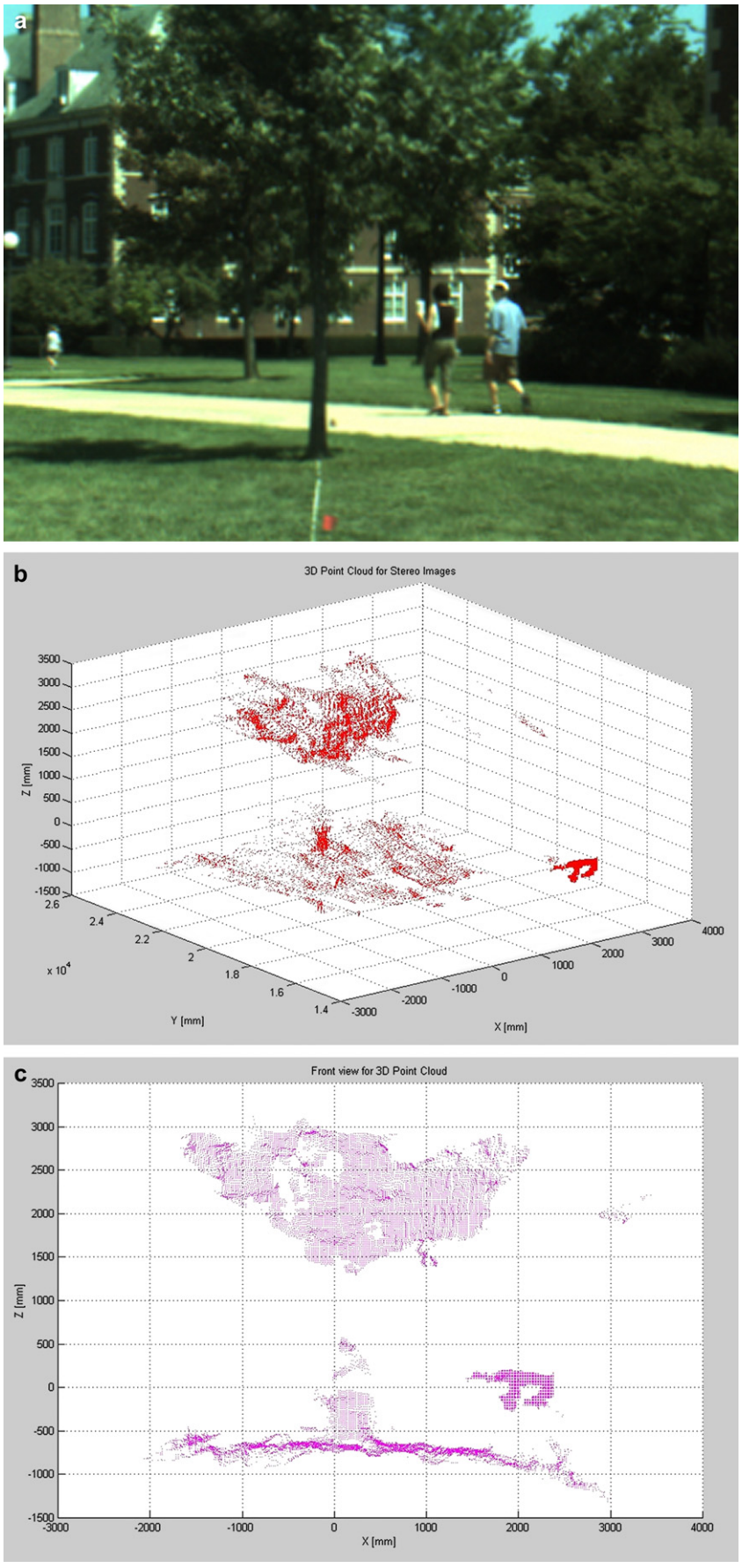

Fig. 8 – Experiment 3, test 7: (a) Target scene; (b) 3D point cloud; (c) Front view.

Table 4 – Results for experiment 2.

Test	Baseline (mm)	Focal length (mm)	ε_y %	Ground offset (m)	Centre offset (m)	Error angle (°) for $\theta = 20°$	FF (mm)
2	150	8	15	0	4	2	1.3
4	150	16	3	0.5	4.6	2.7	1.2
6	194	8	4	0	0.5	0.4	0.8
8	194	12	8	0.7	0.3	4.8	1.2

longest focal length of 16 mm. A similar loss was observed when the medium baseline of 150 mm was matched with 8 mm lenses. The tests demonstrated that a focal length of 4 mm is not suitable for these ranges over 10 m because the field of view is too wide for acquiring images of medium resolution (320×240 pixels) with an acceptable level of accuracy.

The scenes acquired in E1 took place inside a laboratory, and consequently did not possess rich texture and strong illumination. The goal of E2 was to verify that the optimum configurations found in the previous experiment, and shown in Figs. 6 and 7, also provided the most favourable perception when used outdoors. The test scene selected contained a tree surrounded by turf, shown in Fig. 8a. Eight selected tests were conducted according to the parameter combinations of Table 2. The 3D point cloud and its three general views (XY, XZ, and YZ) were generated and saved. From each 3D representation, the following indicators were calculated: relative error in range ε_y, ground offset, centre of coordinates offset, and angular error when the camera was tilted. The 3D point cloud corresponding to Test 7 is shown in Fig. 8b, and its front view XZ is given in Fig. 8c. Table 4 summarises the results of E2 for the four best camera configurations found in E1. Additionally, a form factor (FF) defined by Eq. (7) was computed and added to Tables 4 and 5. The purpose of this factor is to quantify the balance among the key parameters involved in stereo perception as a way of forecasting stereo perceptive performance. The FF is based on the expression of the disparity as a function of the baseline (B), the focal length of the lenses (f), and the range (R). The FF associates a number to any particular combination of the parameters studied; therefore, it can be used to predict the suitability of a set of parameters [B, f, R].

$$\mathrm{FF} = \frac{10 \cdot B\ (\mathrm{mm}) \cdot f\ (\mathrm{mm})}{R\ (\mathrm{mm})} \tag{7}$$

Table 5 shows the FF for all the experiments conducted under experiment E1. There is a correspondence between medium values of FF (1.0–2.0) and the best selection of parameters (highlighted tests in Table 5). Extreme values of FF coincided with unfavourable positions in the efficiency charts of Figs. 6 and 7; Tests 1 and 9 were below 0.5 and Test 13 gave a value of 0.64 whereas the high value of 2.95 corresponded to Test 8. The utility of the FF is to provide a preliminary assessment of the quality of perception before actually assembling the camera; it conveys an expectancy of success, so for example if the FF for a particular combination is comprised between 1.2 and 2.0, there is a good chance of obtaining useful point clouds. However, definite conclusions must be drawn after applying the whole methodology proposed and calculating relative errors and efficiencies.

Table 5 – FF for experiment 1 tests.

Test	Baseline (mm)	Focal length (mm)	Range (m)	FF (mm)
1	103	2.8	6.096	**0.47**
2	103	4	6.096	**0.67**
3	103	8	6.096	**1.35**
4	103	12	6.096	**2.03**
5	150	2.8	6.096	**0.69**
6	150	4	6.096	**0.98**
7	150	8	6.096	**1.97**
8	150	12	6.096	**2.95**
9	150	4	12.192	**0.49**
10	150	8	12.192	**0.98**
11	150	12	12.192	**1.48**
12	150	16	12.192	**1.97**
13	194	4	12.192	**0.64**
14	194	8	12.192	**1.27**
15	194	12	12.192	**1.91**
16	194	16	12.192	**2.55**

4. Conclusions

Frequent changes of camera parameters are not always possible and convenient, and can be difficult when working outdoors. Since different situations require different camera assemblies, the first step in this research was the definition of the situations studied: short ranges of approximately 6 m and medium ranges around 12 m from the front of the vehicle. The first set of tests led to distinct lens-baseline combinations recommended for each type of scene: 6 m ranges are better acquired with 8 mm lenses and baselines ranging from 100 mm to 150 mm, giving stereo efficiencies of 82% and 78%, respectively; 12 m ranges benefit from longer baselines of 200 mm when combined with either 8 mm or 12 mm focal length lenses, resulting in stereo efficiencies of 90% and 92%, respectively. By studying the behaviour of the camera outdoors to obtain the best lenses-baseline combinations, these results were repeated and validated. Once the optimum camera configuration is set for a particular situation, there is no need for further reassembly and calibration. The FF facilitates tentative solutions that can be validated through the planar and stereo efficiencies; FF values between 1 and 2 indicate good expectancy of reaching high efficiencies. Future research, after the hardware architecture has been determined, should focus on the algorithms used for data analysis; how to extract reliable and fast information from the optimally acquired 3D point clouds.

Acknowledgements

The material presented in this paper was based upon work supported partially by the Ministry of Education and Science Funds, Spain (AGL2006-09656/AGR), the United States Department of Agriculture (USDA) Hatch Funds (ILLU-10-352

AE) and Bruce Cowgur Mid-Tech Memorial Funds. Any opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of the University of Illinois, USA; the Ministry of Education and Science, Spain; the USDA, USA; and Midwest Technologies Inc., USA.

REFERENCES

Chen H; Xu Z (2005). Local 3D map building and error analysis based on stereo vision. In Proc. 32nd Annual Conference of IEEE Industrial Electronics Society, pp. 379–382. IEEE.

Herath D C; Kodagoda K R S; Dissanayake G (2006). Modeling errors in small baseline stereo for SLAM. In Proc. 9th International Conference on Control, Automation, Robotics and Vision, pp. 1–6. IEEE.

Kato S; Tomita K; Tsugawa S (1996). Visual navigation along reference lines and collision avoidance for autonomous vehicles. In Proc. Intelligent Vehicles Symposium, pp. 385–390. IEEE.

Kogler J; Hemetsberger H; Alefs B; Kubinger W; Travis W (2006). Embedded stereo vision system for intelligent autonomous vehicles. In Proc. Intelligent Vehicles Symposium, pp. 64–69. IEEE.

Meystel A (1991). Autonomous Mobile Robots. World Scientific Publishing Co., Singapore.

Rovira-Más F; Zhang Q; Reid J F; Will J D (2005). Hough-transform-based vision algorithm for crop row detection of an automated agricultural vehicle. Journal of Automobile Engineering, **219**, 999–1010.

Rovira-Más F; Reid J F; Zhang Q (2006a). Stereovision data processing with 3D density maps for agricultural vehicles. Transactions of the ASAE, **49**(4)), 1213–1222.

Rovira-Más F; Han S; Wei J; Reid J F (2006b). Edge detection for autonomous guidance of a corn harvester using stereo vision. In Proc. ATOE Conference, pp. 171–181. ASABE, St. Joseph, Michigan.

Rovira-Más F; Wang Q; Zhang Q (2007). Design of Stereo Perception Systems for Automation of Off-road Vehicles. ASABE Paper No. 073126. ASABE, St. Joseph, Michigan.

Thompson S; Kagami S (2005). Humanoid robot localisation using stereo vision. In Proc. IEEE-RAS International Conference on Humanoid Robots, pp. 19–25. IEEE.