available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/compag

Stereo vision three-dimensional terrain maps for precision agriculture

Francisco Rovira-Más[a,*], *Qin Zhang*[b], *John F. Reid*[c]

[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, Urbana, IL, USA*
[c] *John Deere Technology Center, Moline, IL, USA*

ARTICLE INFO

Article history:
Received 5 October 2006
Received in revised form
24 July 2007
Accepted 26 July 2007

Keywords:
Autonomous navigation
Precision agriculture
Stereo vision
3D map
GPS
IMU
Terrain mapping

ABSTRACT

The combined interest in precision agriculture, information technology, and autonomous navigation has led to a growing interest in the generation of 3D maps of mobile equipment surroundings. This article proposes a method to create 3D terrain maps by combining the information captured with a stereo camera, a localization sensor, and an inertial measurement unit, all installed on a mobile equipment platform. The perception engine comprises a compact stereo camera that captures field scenes and generates 3D point clouds, which are transformed to geodetic coordinates and assembled in a global field map. The results showed that stereo perception can provide the level of detail and accuracy needed in the construction of 3D field maps for precision agriculture and field robotics applications.

1. Introduction

The technological development of disciplines such as agricultural robotics and precision agriculture requires the knowledge of certain spatial information as, for example, the position of a vehicle or an approximate description of its environment. In order to move autonomously, a mobile platform needs to be aware of its surroundings, and a real-time generated map can provide such information. In the same fashion, precision agriculture systems require the registration of crop location before achieving site-specific operation management. The purpose of this paper is to develop a conceptual framework to obtain 3D terrain maps of agricultural fields for precision agriculture applications. According to Han and Zhang (2001), the automatic control of tractor functions and implement operations such as engine throttle, transmission speed, and 3-point hitch position will be necessary before tractors become fully autonomous. To automatically implement site-specific operations, each control variable is associated with a field location at any time, and the time sequence of a control variable can be viewed as a 2D map called a field operation map. Such maps are often created off-line via GIS, and then loaded into the navigation computer. The benefits and uses of terrain maps of agricultural fields are numerous, as stated above, but selecting the most appropriate sensors to generate them remains a research subject.

Among the sensors currently available, vision systems seem to be excellent candidates for map generation,

[*] *Corresponding author*. Tel.: +34 96 387 7291; fax: +34 96 387 7299.
E-mail address: frovira@dmta.upv.es (F. Rovira-Más).

0168-1699/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2007.07.007

apparently with higher potential than other perception sensors such as radar, sonar or laser. The last two decades have witnessed great advancements in computer vision, mostly image processing techniques for monocular CCD cameras. It is generally agreed that stereoscopic vision provides a much richer representation of reality than the conventional two-dimensional (2D) images. The processing speed of present-day computers allows stereo engines to perform in real-time. Stereo imaging has proven to be an effective three-dimensional (3D) vision system in remote missions, such as NASA's Mars Pathfinder. It has been used in a wide variety of environments, including such extreme conditions as the ones found in planetary missions and underwater vehicles (Weast et al., 1999). A particular situation where ranges detected by a Martian rover reached a few kilometers was solved by constructing a wide-baseline stereo vision system with only one camera, as reported by Olson et al. (2003). The extended time necessary to generate the point cloud together with epipolar complexities make this option difficult to develop for precision agriculture applications. This technique of employing one single camera to produce the stereo effect by locating it in pre-established stations has also been applied to habitual situations as for instance mobile robot surveillance (McBride et al., 2005); however, the advantages given by compact binocular cameras make single camera stereo a secondary choice for terrain mapping. Stereo cameras have become a favorite sensor for real-time modeling of a robot's environment for manipulative tasks, as shown by Lee et al. (2005): objects can be detected within a certain boundary around the robot, but it is necessary to know the global position of the robot before estimating the global position of the objects. This is the principal problem of terrain mapping. One way of dealing with the difficulties of global localization is by placing landmarks in the field of view of the camera as suggested by Wang et al. (2005), where both the vehicle and targets remained stationary. The placement of landmarks in the field is not convenient for agricultural applications unless they are installed over firm and very stable platforms, which would result in an increase in cost for the farmer as well as a handicap to operate vehicles around them, especially if they move autonomously. The need for merging global and local sensor information is, in fact, a restrictive condition which makes many published solutions unviable, as in the interesting case reported by Hrabar et al. (2005), where a tractor was autonomously guided using optic-flow combined with stereo vision, but no global reference was attained.

Laser scanners are direct competitors of stereo for achieving 3D maps. A global 3D map requires building models from multiple views. As addressed by Huber et al. (2000), the range data acquired using a laser range finder could be used to create maps of both indoor environments and large-scale terrains. A laser 3D mapping requires an accurate synchronism between the laser and the attitude-position sensors. A procedure to generate a 3D map from aerial images was described by Miller and Amidi (1998). A laser scanner was installed on an autonomous helicopter, and laser measurements were converted into 3D coordinates in an Earth-fixed reference frame.

A crucial issue related to the fusion of local maps for generating a global map is the alignment of multiple 3D submaps. A solution to this problem was provided by Se et al. (2002), where a trinocular stereo system was used to track natural landmarks and to build a 3D map. However, due to problems when large slippages or long-term drifts occur, the alignment of multiple submaps had to use the close-the-loop constraint to obtain a consistent global map, that is, a technique to correct all the submap alignments knowing that the final point is the same as the initial point and therefore the coordinates should be the same. A sensor combination of GPS and stereo vision was proposed by Zhao and Aggarwal (2000) to reconstruct urban scenes. The algorithm matches feature points, and must be regarded as a proof of concept since the authors did not apply it to real images to generate a 3D map. A benefit of building a map of a vehicle's surroundings is the possibility of computing its absolute location. This challenge has been dealt with the simultaneous localization and map building (SLAM) problem, where a mobile platform can start in an unknown location in an unknown environment, and, using relative observations only, build a map of the world (Dissanayake et al., 2001). SLAM applications have been growing in number and popularity in the last 5 years. The ideas raised by SLAM problems are attractive to agricultural mapping, but their actual implementation often results in overwhelming limitations such as excessive time per iteration, the need of *a priori* information about the environment, or the requisite to close the loop.

This paper describes a methodology to create 3D field maps of agricultural scenes using stereo vision. The basic requirements for the searched solution are (1) global reference with the degree of detail given by local perception, including color awareness; (2) real-time capabilities and on-line map generation; (3) portability to any agricultural or off-road platform; (4) functionality without adding or modifying the current structures of the field; and (5) applicability to precision agriculture needs. The solutions found in the bibliographical review contribute to some of the five requirements mentioned above, but, on the other hand, they also introduce certain constraints that disable them for becoming the optimal approach for three-dimensional (3D) terrain mapping in agriculture. The following sections of the article present the architecture of the developed system, the devised approach, and the results obtained in orchard scenes recorded in the State of Washington during 2002 and 2003 as well as other general scenes taken in the University of Illinois at Urbana-Champaign in 2004.

2. System architecture

In order to acquire the necessary information for the generation of 3D terrain maps, three different sensors have to be synchronized: a stereo camera, a localization sensor, and an inertial measurement unit. This equipment needs to be carried in a vehicle together with a computer, so that the 3D map can be generated as the vehicle travels along the field. The following paragraphs describe in detail the choice of sensors and the architecture devised.

2.1. *Stereoscopic cameras*

Two different sensors were selected to acquire stereoscopic images:

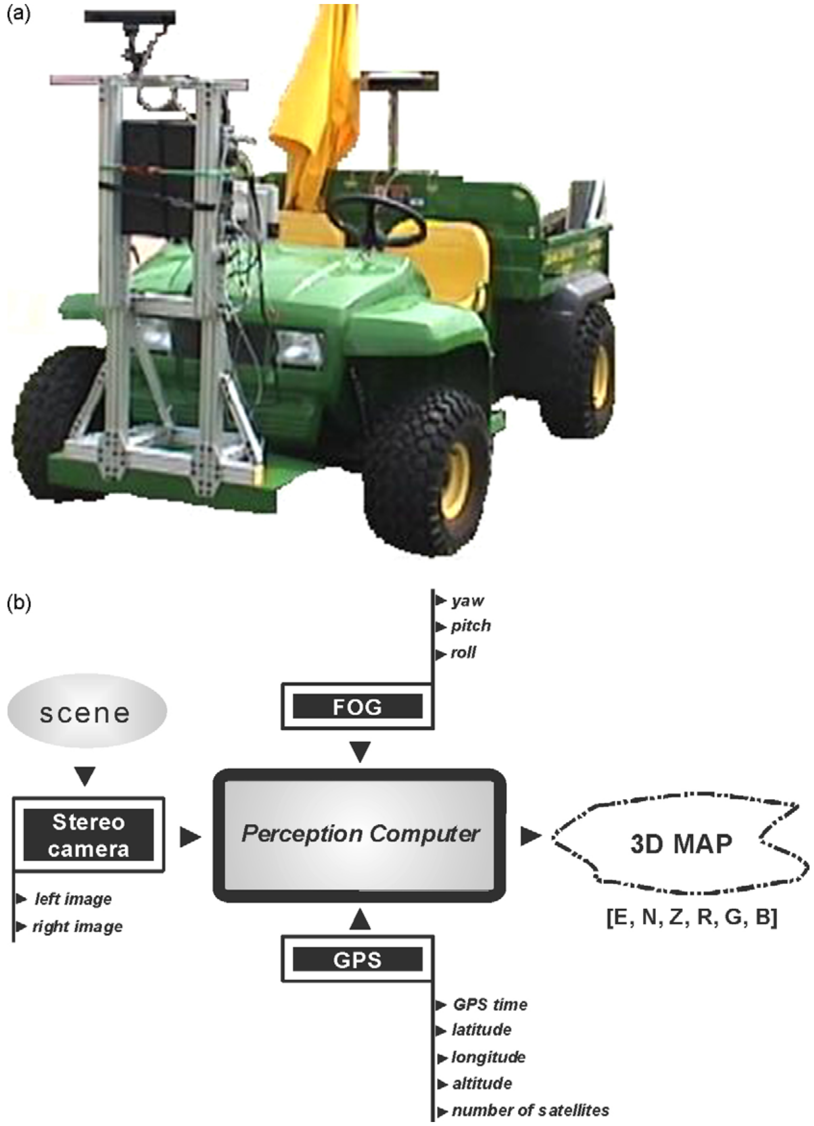

Fig. 1 – (a) System implementation: system mounted on a utility vehicle. (b) System architecture for 3D map generator.

1. The MEGA-D is a commercial all-digital stereo camera made by Videre Design (Menlo Park, California). It supports C mount lenses (4.8 mm), and a fixed 9 cm baseline. Its imagers are 2/3 in., and it uses the IEEE-1394 bus for direct digital input.
2. The Tyzx, Inc. 3Daware DeepSea Development System is a complete, high-performance stereo vision system comprised of a real-time stereo processor and a Tyzx stereo camera. The camera possesses a 22 cm baseline and pre-focused lenses that do not require calibration by the user. Communication between camera and board is attained by LVDS electrical interface for left and right images as well as an I2C interface.

2.2. *Mobile platforms*

Several platforms have been used to generate 3D field maps in the course of the experiments reported in this research; however, the majority of images studied in this article were captured with the Gator^{TM} utility vehicle (John Deere, Moline, IL) shown in Fig. 1(a). The stereo camera was located on the front of the utility vehicle at heights that ranged from 1 to 1.8 m. The vehicle carried a desktop computer for perception analysis, either a Real-Time-Kinematic (RTK) GPS or a Starfire^{TM} DGPS, and a Fiber Optic Gyroscope (FOG) to acquire inertial measurements. Other platforms that were prepared to perform 3D mapping include agricultural tractors and an unmanned helicopter for aerial images.

2.3. *Methodology: system architecture*

The perception computer was responsible for assimilating all the information delivered by the sensors to build the 3D map. For a given scene, the stereo camera captured the stereo pair of images (left and right) and sent them out to the computer. The left image was typically recorded in color, except for the cases when an infrared filter was mounted on the lenses. The processing card and software on the computer calculated the

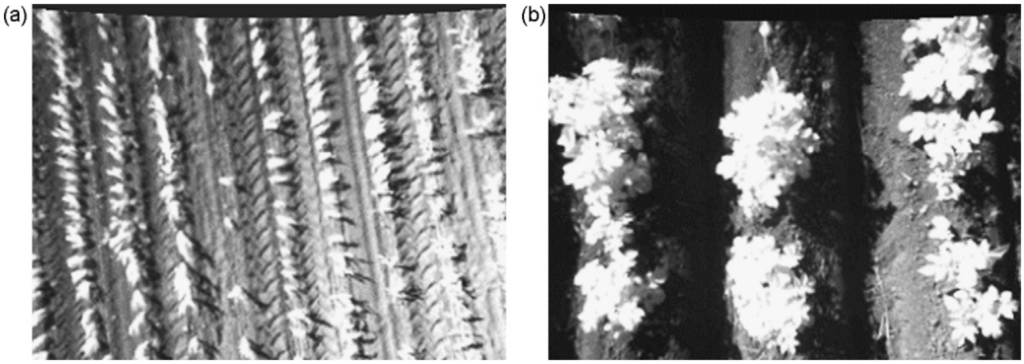

Fig. 2 – Aerial infrared images: corn from a remote-controlled helicopter (a) and potatoes captured from a tractor (b).

disparity image that was used to determine the camera coordinates of the points in the scene. For every stereo pair of images, the GPS estimated the global coordinates of the position of the camera. In the early stages of the project, inertial measurements from the FOG were not available, which led to the yaw calculation from the GPS coordinates. The number of satellites, together with other quality indices such as the dilution of precision (DOP), was tracked to avoid fusing images with the wrong coordinates. Once the FOG was incorporated to the system, the availability of instantaneous inertial measurements (roll, pitch, and yaw) allowed for a real-time generation of the map as well as pitch-roll compensation in uneven terrains. With the transformation of every point captured with the camera to global coordinates, all the images were merged in a unique global map with one common origin of coordinates. The final 3D terrain map comprised point clouds with six different parameters for every point given by the stereo camera: East coordinate, North coordinate, altitude Z, Red color component, Green color component, and Blue color component. With these parameters, a 3D representation of the virtual scene was achieved. Fig. 1(b) illustrates the system architecture of the 3D field mapping system.

3. Image type and transformation of coordinates

3.1. *Image configuration for 3D mapping*

There are unlimited ways of positioning the camera with respect to the targeted scene. Two particular configurations were found especially favorable for the purpose sought in this project: *ground images* and *aerial images*. Aerial images are those images taken when the image planes (imagers of a digital stereo camera) are approximately parallel to the ground, producing a top view visualization of the scene. Ground images are images generally taken from ground vehicles, and in this configuration the image plane is perpendicular to the ground or forming certain inclination angle with respect to the ground. The most significant difference between both types of images is the effect of perspective, and, therefore, the presence of a *vanishing point* in ground images. The following figures show several examples of images captured with different stereo rigs. In Fig. 2(a) an aerial image taken through an infrared filter from a remotely operated helicopter covered eight rows of corn. Fig. 2(b) is another aerial image, but this time from a tractor traveling in a field of potatoes. The basic difference between these two aerial images is the proximity of the camera to the crop; the helicopter had to remain at a certain distance to the ground, whereas the tractor kept a more reduced and constant separation. Fig. 3(a and b) are ground images obtained from a camera mounted at the front of a tractor and on a utility vehicle respectively. The former portrays rows of soybeans, and the latter shows an orchard of cherry trees.

3.2. *Transformation of coordinates*

The coordinates registered by the stereo camera are given in the camera system of coordinates $X_cY_cZ_c$, whose origin is set at the optical center of the left lens. The Z_c coordinate rep-

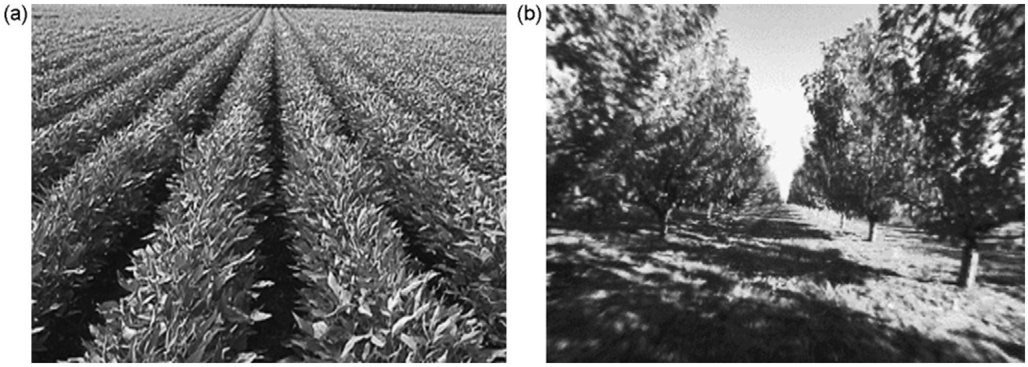

Fig. 3 – Ground images: soybeans (a) and cherry trees (b).

resents the *range* or camera-object distance, and the X_c–Y_c plane is coincident with the image plane. The camera coordinates are not convenient for global terrain mapping where several images must be fused together. Furthermore, aerial images and ground images will require different transformations to ease the 3D representation of point clouds. This article mainly discusses the generation of 3D field maps composed from ground images, but a deeper explanation of maps created from aerial images can be found in Rovira-Más et al. (2005).

Very often, in a ground image configuration, the camera needs to be inclined some degrees to ensure that the key visual information of the scene is included in the camera field of view. In these cases, the camera system of coordinates does not adequately represent the objects of the scene since the X_c–Y_c plane is tilted according to the camera inclination angle and, as a result, the Z_c axis presents an angle with respect to the ground. A more intuitive system of coordinates (*ground coordinates*) was devised to facilitate the representation and integration of individual stereo images. Both the camera system of coordinates and the ground system of coordinates are represented in Fig. 4. The *camera coordinates*, as defined above, are represented by the frame $X_cY_cZ_c$. The ground coordinates keep the origin at the left lens plane, but it is displaced to the ground level (marked as (0, 0, 0) in Fig. 4). The X axis of the ground coordinate system is coincident with X_c, the Y coordinate represents the distance object-camera, and the Z coordinate always gives the height of the selected point regardless of the camera inclination angle ϕ. The diagram of Fig. 4 clearly shows the dependence of $X_cY_cZ_c$ on ϕ, and the difficulties of representing 3D point clouds in camera coordinates. As a matter of fact, images captured with different inclination angles are easier to combine in a unique map after the transformation to ground coordinates has been performed.

The transformation between both systems of coordinates is given in Eq. (1) where (x_c, y_c, z_c) are the camera coordinates (x, y, z) are the newly defined ground coordinates, h_c is the camera height measured at the optical center of the lens from the ground, and ϕ is the camera inclination angle. The reader should not be distracted from the importance of the transformation of Eq. (1) by its apparent straightforwardness: this is a crucial stage in the map creation that many papers in this area tend to overlook. A similar transform is defined by Lobo et al. (2003), but the methodology is more sophisticated and it requires the estimation of gravity with an IMU. The application of Eq. (1) reported satisfactory results without the need of more complex procedures.

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

4. Local maps

4.1. Image noise and validity box

A *global map* consists of an orderly arrangement of individual map-portions of the sensed field. Any of such portions come from a particular stereo image where camera coordinates have been transformed to locally defined ground coordinates according to Eq. (1). Since these individual maps are locally referenced to the camera position, they are designated *local maps*. Every stereo pair will create a local map, and the ground coordinates of any sensed object will be referenced to a particular local map. Due to the fact that a global map is composed of a set of local maps, it is essential to generate good quality local maps in order to obtain an acceptable global map. The quality of a local map can be assessed by the value of the disparity image associated with the stereo pair of images that define the local map. Generally speaking, two fundamental problems related to local maps were found: *scarce matching* and *unfiltered mismatching*. The basic operation in stereo vision is the correlation between the left and the right image. The Tyzx$^{\text{TM}}$ stereo camera, for instance, features the "Census Algorithm" for pixel correlation, a non-parametric summary of local spatial structure (Zabih and Woodfill, 1994).

In spite of the high performance shown by commercial matching algorithms, the presence of mismatches giving wrong coordinates is frequent. Pixels in the disparity image without disparity information will not yield stereo information. A disparity image with few matched pixels is regarded as a scarce matching situation. Reasons for obtaining such images include lack of texture in the sensed scene, deficient ambient illumination, and the occurrence of pixels that have been filtered out by the correlation algorithm, and therefore do not carry any 3D information.

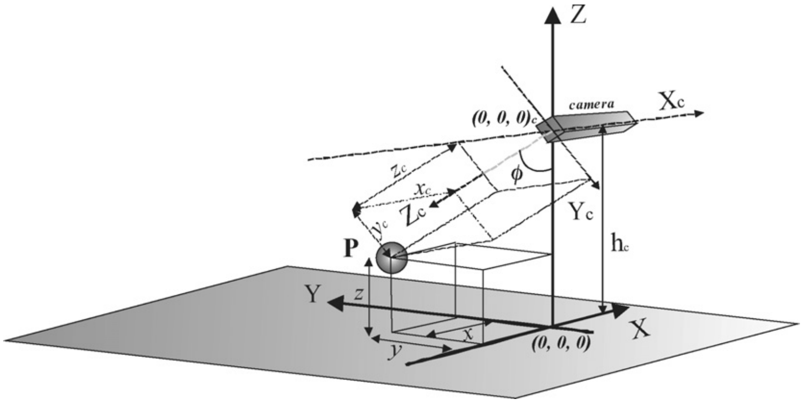

Fig. 4 – Transformation from image coordinates to ground coordinates.

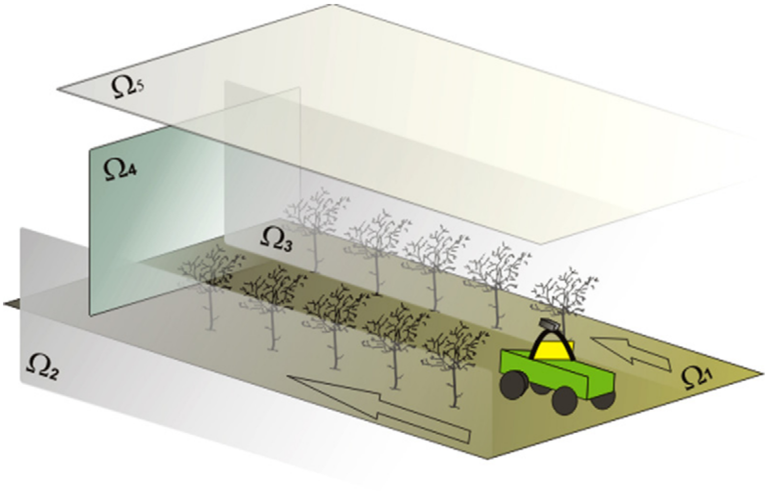

Fig. 5 – Concept of validity box.

Agricultural fields and orchard scenes are generally well illuminated and possess rich texture patterns, which typically yields disparity images with an extensive coverage. Unfiltered mismatches are pixels with erroneous stereo information that will lead to meaningless locations. A straightforward approach to this problem, at least to a certain extent, is a space of logical coordinate-placement and neglect positions outside that space. This space is termed the *validity box* for local maps, and it proved to be very effective in filtering out underground misplacements as well as incorrect matching at the top half of images provoked by cloudy skies. Fig. 5 illustrates the concept of the validity box for a ground vehicle in an orchard scene.

Since the vehicle represented in Fig. 5 is traveling between tree rows, and the trees block the view of the neighboring rows, no points can be found in adjacent inter-row spaces. Planes Ω_2 and Ω_3 avoid noisy points falling in nearby rows, plane Ω_1 prevents negative locations, Ω_4 limits the ranges to a reasonable distance, and plane Ω_5 eliminates distractions caused by clouds or other background objects. The set of conditions given in Eq. (2) represent the concept of the validity box in a mathematical format. The plane Ω_1 does not necessarily have to be located at $Z=0$, but the Y coordinate cannot be negative since the camera is looking ahead. P is a point of the 3D cloud with ground coordinates (x, y, z).

$$\text{Validity box definition}: L_1(\text{x-coordinate}) \times L_2(\text{y-coordinate}) \times L_3(\text{height}), \begin{cases} \Omega_1 \equiv z = 0 \\ \Omega_2 \equiv x = -\dfrac{1}{2}L_1 \\ \Omega_3 \equiv x = \dfrac{1}{2}L_1 \\ \Omega_4 \equiv y = L_2 \\ \Omega_5 \equiv z = L_3 \end{cases},$$

$$P(x, y, z) \subset \text{Validity box if} \begin{cases} -\dfrac{1}{2}L_1 < x < \dfrac{1}{2}L_1 \\ 0 < y < L_2 \\ 0 < z < L_3 \end{cases} \tag{2}$$

4.2. Concept of 3D density and density grids

The *cloud of points* representing the targeted scene is usually yielding the maximum resolution that a particular stereo system can produce. There are some applications, though, where less resolution is preferred, for example, a real-time navigation system that is processing obstacle maps might be more useful if the processing speed is high, even though that implies a less detailed representation of reality. Such a navigational map will often require the record of position and dimension of obstacles to be avoided and will not improve by including very small objects. A satisfactory methodology to simplify the resolution of 3D field maps while maintaining the key information is through the concept of 3D *density* and *density grids*. The idea of the 3D density is rooted in the properties of the conventional density, which establishes a relationship between the mass of a substance and the volume that it occupies. Similarly, the 3D density is defined as the number of stereo matched points detected per unit volume. The mathematical definition of the term is provided in Eq. (3) below:

$$d_{3D} = \frac{N}{V} \tag{3}$$

where V is the volume of space considered, N the total number of points inside V, and d_{3D} is the 3D density. Fig. 6 gives a graphical description of the 3D density, where the d_{3D} of the cell represented is $14/d^3$.

A practicable approach for applying the concept of 3D density is by dividing the space sensed by the stereo camera into a regular grid, and then computing the 3D density for every cell of the grid. A grid where the 3D density is computed and represented is denoted as a density grid. A 3D density grid is shown in Fig. 7. The concept of d_{3D} and its embodiment through density grids leads to issues such as the loss of density with the increment of ranges: pixels in the foreground carry more information than those close to the vanishing point. The decay was found to be quadratic and a compensation formula was deduced to equalize densities within the map (Rovira-Más et al., 2006).

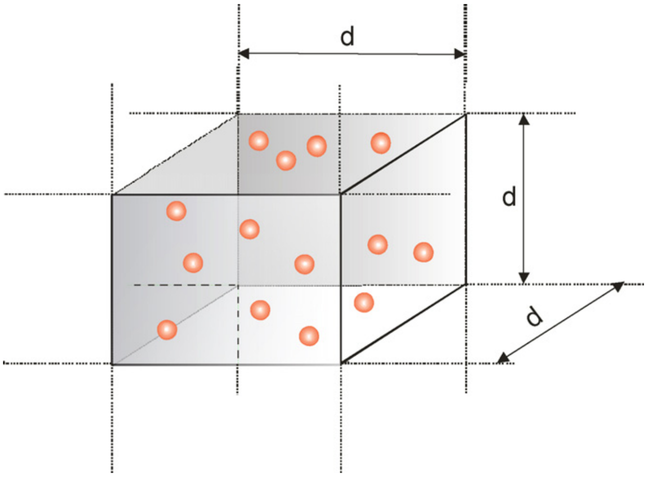

Fig. 6 – Concept of 3D density.

5. Global map construction

5.1. *Transformation from local to global coordinates*

In essence, a global map is the arrangement of local maps in the appropriate position and orientation. Before arriving to the merging phase, local maps have to be adequately treated to palliate noise and mismatching errors. The 3D locations of the objects captured in the scenes are given in ground coordinates referenced to the local frame (camera) center of coordinates. Since the global coordinates of the camera position at any time are known with a GPS receiver, any point in the scene can theoretically be converted to global coordinates to constitute a global map. However, in addition to the global (geodetic) coordinates of the camera center of coordinates, the angle of the camera system of coordinates with respect to the global system of coordinates must be known as well. The estimation of the orientation angle between systems of coordinates led to the two alternatives described in the following paragraphs. After the non-linear transformation from local to global coordinates is completed, all the local maps will be related to the same origin and system of coordinates, constituting the global terrain map.

The objective of this operation (global map construction) is the transformation of every point that belongs to a local map and is therefore expressed in ground coordinates, into a point given in *global coordinates* (East, North) where every location has its reference in the Tangent Plane System of Coordinates. The origin of the global frame is arbitrarily set for every map. The first approach pursued to build the map was a simple solution for field scenes structured in straight rows. The searched parameter was the *orientation angle* ψ, defined as the angle between the orientation of the crop rows and the East direction. The hypothesis for this transformation was that all the rows are parallel so that only one of them is necessary to determine the orientation angle ψ. If one particular row, say row R, consists of k images, the orientation angle is given by Eq. (4), where the global coordinates of the origin for the first image are given by (E_1, N_1), and the global coordinates of the center of coordinates for the last image of row R are (E_k, N_k).

$$\psi = \arctan\frac{N_k - N_1}{E_k - E_1} \tag{4}$$

If n is a specific point belonging to image j, the ground coordinates for point n will be known and given by the pair (X_n, Y_n). Since point n belongs to image j, the global location of the origin of local (ground) coordinates will be (E_j, N_j). In view of the fact that image j is also a component of the global map, point n can be transformed to global coordinates (E_n, N_n) according

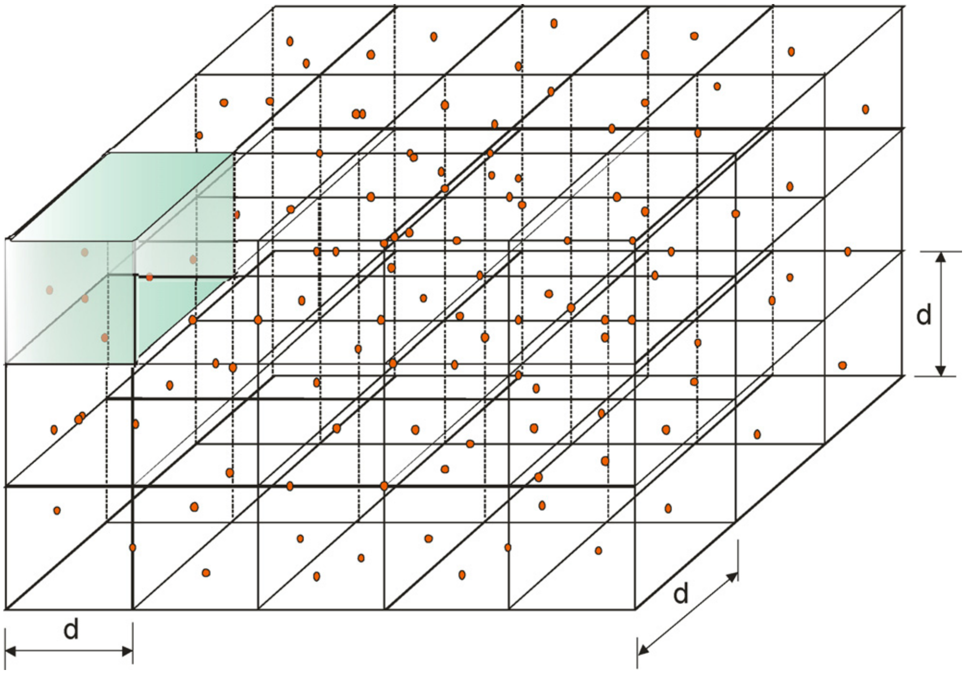

Fig. 7 – Three-dimensional density grid.

to Eq. (5), taking into account that ψ is the orientation angle found in Eq. (4) and $\Pi = 1$ when the vehicle travels in the forward direction, and $\Pi = -1$ when the row is traversed in the returning direction.

$$\left.\begin{aligned} N_n &= N_j + [y_n \sin\psi + x_n \cos\psi]\,\Pi \\ E_n &= E_j + [y_n \cos\psi - x_n \sin\psi]\,\Pi \end{aligned}\right\} \tag{5}$$

In the preceding approach, the on-board computer was programmed to register the geodetic coordinates of the camera every time an image was acquired, while the orientation angle between local and global frames was deduced from the GPS locations recorded. Eq. (5) represents a transformation of coordinates for flat fields and straight rows. In this particular case, which is the most common, the map can be assembled without the need of an inertial measurement unit. However, in the most generic case, the three Euler angles (roll, pitch, and yaw) have to be known, and consequently inertial sensors need to be implemented in the vehicle. The alternative approach, based on a real-time generation of 3D maps, retrieves not only the coordinates of the camera for every image taken but also its yaw, pitch and roll angles. Consequently, every single point is transformed to global coordinates in real time and added to the global map whose construction is in progress. As a result, the output yielded by the computer is the complete global map ready to be displayed with virtual reality tools. The orientation angle (yaw), together with other attitude measurements, was estimated by a Fiber Optic Gyro (FOG) manufactured by JAE (Japan Aviation Electronics Industry, Tokyo, Japan). The real-time generated 3D terrain maps, besides the three global coordinates *East*, *North* and *Z*, also registered the color code (*Red*, *Green*, and *Blue*) of each single point. Since the three attitude angles (*yaw*, *pitch* and *roll*) were available for this approach, the expression given in Eq. (5) had to be augmented to the more complete Eq. (6) to include the three measurements from the FOG. Having the complete orientation of the camera allowed for 3D maps of uneven terrain, including hilly fields. However, the majority of the applications studied in this project could consider flat topography and the inertial measurement unit was basically employed to sense the yaw angle, i.e., the orientation angle of the camera with respect to the global axes East and North (heading).

$$\begin{bmatrix} E_n \\ N_n \\ Z_n \end{bmatrix} = \begin{bmatrix} E_j \\ N_j \\ Z_j - h\cos\beta\cos\alpha \end{bmatrix} + \begin{bmatrix} \cos\psi\cos\beta & \cos\psi\sin\beta\sin\alpha - \sin\psi\cos\alpha & \cos\psi\sin\beta\cos\alpha + \sin\psi\sin\alpha \\ \sin\psi\cos\beta & \sin\psi\sin\beta\sin\alpha + \cos\psi\cos\alpha & \sin\psi\sin\beta\cos\alpha - \cos\psi\sin\alpha \\ -\sin\beta & \cos\beta\sin\alpha & \cos\beta\cos\alpha \end{bmatrix} \begin{bmatrix} x_n \\ y_n + d \\ z_n \end{bmatrix} \tag{6}$$

The transformation equation in Eq. (6) can be understood as an amplification of Eq. (5) where three dimensions, and therefore six degrees of freedom, are taken into account: the Cartesian ground coordinates of any point (X_n, Y_n, Z_n), and the Euler angles (ψ, α, β) of the stereo camera. Additionally, the height at which the GPS is mounted with respect to the ground (*h*), and the horizontal distance (*Y* direction) between the GPS and the optical center of the reference lens (*d*) also need to be considered in the transformation to global coordinates. The *geodetic coordinates* of the camera at the instant when image *j* was acquired are given by (E_j, N_j, Z_j), and the global coordinates of the transformed point are represented by (E_n, N_n, Z_n).

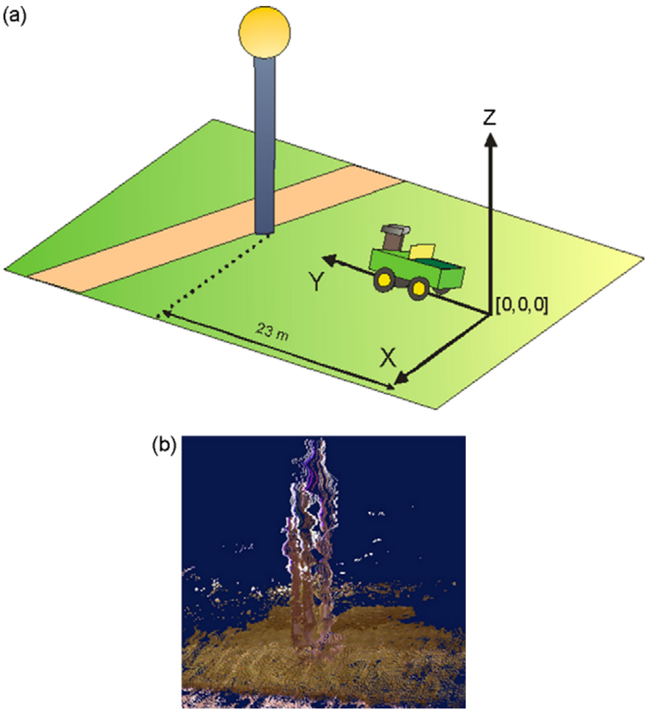

Fig. 8 – Localization sensor failure: (a) experiment setup and (b) field map.

5.2. *Challenges for terrain global mapping*

It remains to complete the survey of those general limitations that can be placed upon the success of terrain mapping using this methodology. One of the earliest issues that came upon was the creation of sparse maps resulting from a poor synchronization between the forward speed of the vehicle and the processing speed of the on-board computer. These maps presented extensive void areas where no 3D information was recorded. By increasing the processing speed for the on-board computer, the problem of insufficient overlap was easily overcome. An error of different nature takes place when inaccuracies in the inertial measurement unit or the GPS receiver occur unexpectedly. The experiment shown in Fig. 8 illustrates the problem of mapping under GPS failure. A utility vehicle was directed toward a lamppost as schematized in Fig. 8(a). The resulting map of Fig. 8(b) depicts a triplicate of the sensed post. Localization errors were coped with quality estimators: number of satellites in solution and GPS quality index. Agricultural equipment typically navigates through the fields several times each season, and since a field map does not necessarily have to be fully determined during the first pass, only when sensor data was acceptable was the local map integrated in the global map, otherwise it was ignored and no further data was added until the quality indices exceeded certain threshold.

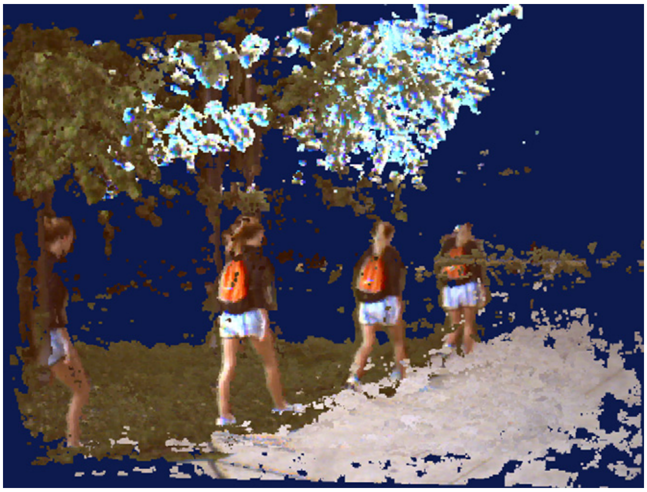

Fig. 9 – Three-dimensional map generated when a moving object crossed the field of view of the camera.

6. Results and discussion

In this study, the creation of field maps focused especially on the detection of field structures for navigation and the estimation of crop properties for monitoring purposes. The advantage of using a stereo camera for perception also brings the added benefit of the potential use of the range for safeguarding. When autonomous operations are taken into account, safety is probably the primary concern, mainly in the case of agricultural fields where people and heavy machines coexist. The map represented in Fig. 9 was built by merging four images (i.e., local maps). While the vehicle was being prepared for the mapping session, a person crossed the field of view of the camera and was registered four times. Every image provided the distance from the vehicle to the object (range or Y coordinate), and this information output was the basis for safeguarding decision-making: continue moving, reduce speed, or halt the vehicle immediately. The environment reconstruction was consistent in the four images, and therefore the sidewalk, trees, and grass portion were correctly situated. However, the color code of some trees was blended with the hues of the sky: stereo matching is principally based on texture correlation, and color sensor quality is typically a secondary requisite for compact stereo cameras.

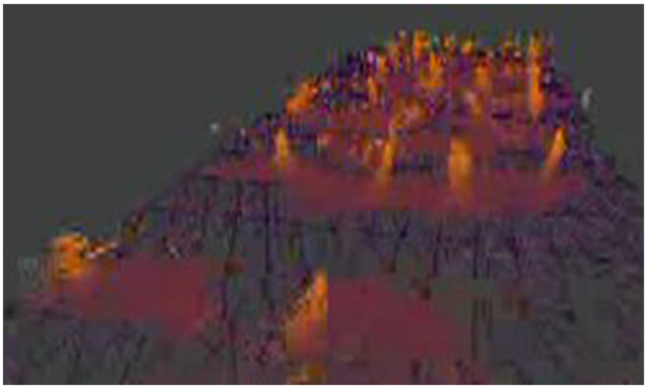

Fig. 11 – Barren field prepared for hops production: top view.

Since the system architecture (Fig. 1(b)) required the coordination of several sensors, the initial proof to verify was the correct and logical placement of local maps (*inter-map coherence*), that is, a general corroboration that the results are meaningful before focusing on the intra-map accuracy. The inter-map correctness included the right position of the ground truth as well as the proper relative orientation between sub-maps. Fig. 10(a) shows the actual scene used to test the mapping system. A barren field prepared to plant hops, a twining vine used to impart a bitter flavor to malt liquors, consisted of wooden posts arranged in equally spaced rows. The posts were placed following two configurations: perfectly vertical (left row in Fig. 10(a)) and tilted certain angle (right row in Fig. 10(a)). The desired and expected response of the algorithm was a global map with the following properties: ground consistency, right spacing for posts, posts correct orientation, and posts distinct vertical configuration. Fig. 10(b) renders a ground-level view of the global map associated to the scene of Fig. 10(a).

In the global map depicted in Fig. 10(b), the position of the ground in the field was correct. Several posts are distinguishable in the scene, and as expected, some of them are vertical and some of them tilted. A top view of the map (Fig. 11) confirms the equality in the rows spacing as well as in the distance between posts belonging to the same row. The approximate separation between posts was fairly close to the actual spacing. The map is depicted in false colors to facilitate the distinction between ground and posts.

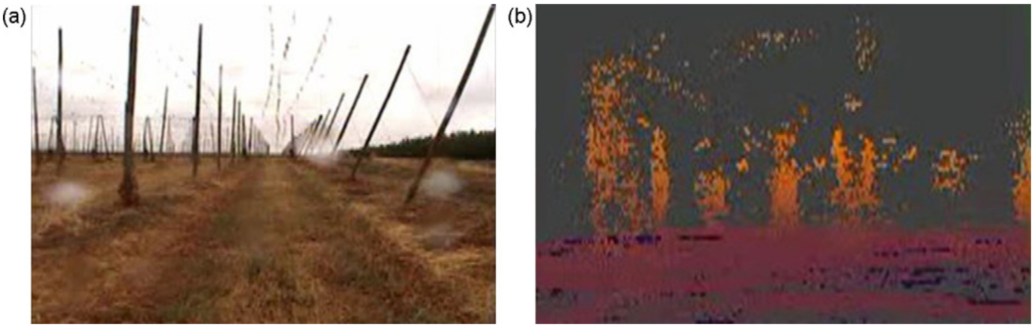

Fig. 10 – Barren field prepared for hops production: (a) real scene and (b) global map, ground-level point of view.

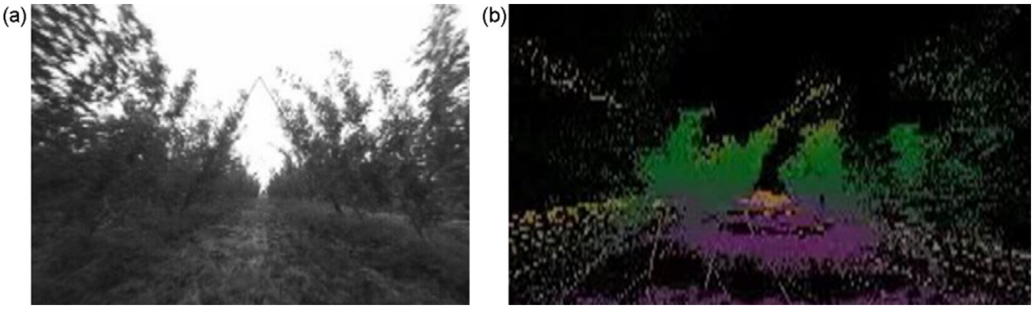

Fig. 12 – Apple tree field structured in V-shape posts: (a) real scene and (b) global map, ground-level point of view.

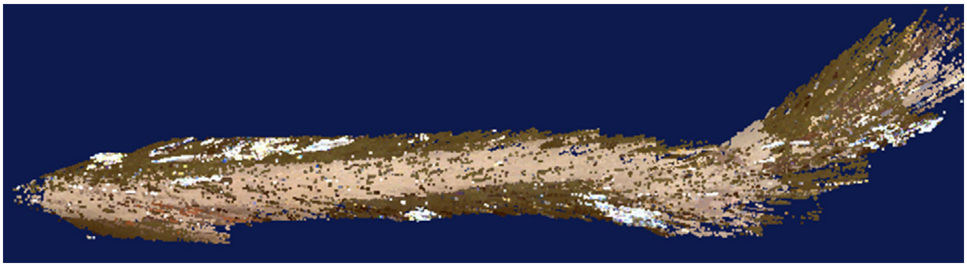

Fig. 13 – Three-dimensional terrain map generated with a utility vehicle: top view.

A further step, and consequently broaden challenge, was set in the field scene represented in Fig. 12(a). The apple trees were configured in a V-shape structure resulting in two new difficulties: (1) vegetation blended with the wooden structure, limiting its visibility and reducing the ambient illumination that reached the camera and (2) the structures guided the trees in such a way that they partially closed the space above the vehicle, increasing the probability of a GPS signal blockage. The distinct triangular configuration of the trees shown in Fig. 12(a) was replicated in the virtual map of Fig. 12(b). From the three-dimensional field map (Fig. 12(b)) such information as crop size and shape, row spacing, structure position, and vehicle traversability could be extracted to feed the control unit of an intelligent vehicle.

As previously stated, one of the applications of 3D mapping to agriculture with highest potential is the automatic guidance of agricultural vehicles. Virtual terrain maps can aid autonomous navigation, and, if not as the unique perception system, they can definitely contribute to the main control engine of the vehicle. Fig. 13 represents the top view of a 3D map generated by a utility vehicle driven along a sidewalk in a university campus. The trajectory followed by the vehicle included straight sections as well as certain curvature and a turn at the end of the path. The overlap between local maps was acceptable, and the GPS and IMU provided satisfactory data to register the scene with enough precision. The utility vehicle traveled over the paved path, which had several trees along the sides on a wide turf area. At the end of the run, the vehicle followed a slight curve and left the paved path to enter into the green area. The width and alignment of the path was consistent in all the images (local maps), delimiting a potential trajectory to be followed by an autonomous mobile platform traveling from the beginning to the end of the sidewalk, avoiding the trees, or other potential obstacles, located on both sides. The pixels representing the path (high intensity) indicated the traversable terrain for the vehicle. Unlike the maps rendered in Figs. 10–12, true color was essential in this map to discriminate the position of the path from the non-passable terrain where campus trees and art statues stand.

A front view of the terrain map of Fig. 13 can show the height and position of the side trees, complementing the navigational information by adding details with respect to possible lateral hazards to the vehicle. A 3D map provides the most complete representation of reality, offering the advantage of choosing the most adequate view for each application pursued. The field map of Fig. 13 was composed of 20 images, which produced a point cloud of 492428 points. The number of GPS satellites in solution fluctuated between six and seven throughout the entire image acquisition phase. The difficulties encountered to handle and manipulate maps of such, or larger, magnitude were alleviated by displaying the field maps in a specialized 3D virtual reality chamber (John Deere Technology Center Immersive Visualization Laboratory, Moline, IL), where the viewer could walk around the confined space of the chamber wearing 3D vision glasses and perceiving the virtual environment as if it was real.

7. Summary and conclusions

The role that information technology is playing in present day engineering applications has led to a growing interest in acquiring as much information as possible from available sensors, mainly with regard to localization and perception. In that stream of research, a methodology to generate 3D terrain maps was developed by combining information obtained with a compact stereo camera, a localization sensor and an inertial measurement unit. Agricultural applications can benefit from

3D field maps in such implementations as yield monitoring or autonomous navigation. The technique devised through this research was validated in multiple field tests with different stereo cameras and positioning sensors. Several coordinate transformations and concepts such as 3D density were developed in the course of this investigation. The accuracy of 3D representations through stereo is known to be inferior to that obtained with a laser range finder; however, the complexity of mounting a laser unit on a vehicle with two planes of scanning is substituted by the simple task of affixing a camera to the vehicle. Generally speaking, in many current applications, standing-alone stereo cameras as 3D perception units are not capable of providing visual information reliably enough to succeed without the aid of other redundant sensors, but stereoscopic vision is becoming an essential component in most of the perception units of robotic applications, and its significance seems to be increasing fast.

Acknowledgements

The material presented in this article was based on work supported partially by Deere & Company and USDA Hatch Funds (ILLU-10-352 AE). Any opinions, findings, and conclusions expressed in this publication are those of the authors and do not necessarily reflect the views of Deere & Company, USDA and the University of Illinois.

REFERENCES

Dissanayake, M.W.M.G., Newman, P., Clark, S., Durrant-Whyte, H.F., Csorba, M., 2001. A solution to the simultaneous localisation and map building (SLAM) problem. IEEE Trans. Rob. Autom. 17 (3), 229–241.

Han, S., Zhang, Q., 2001. Map-based control functions for autonomous tractors. ASAE paper No 011191. St. Joseph, MI.

Hrabar, S., Sukhatme, G.S., Corke, P., Usher, K., Roberts, J., 2005. Combined optic-flow and stereo-based navigation of urban canyons for a UAV. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 3309–3316.

Huber, D., Carmichael, O., Hebert, M., 2000. 3-D map reconstruction from range data. In: Proc. 2000 IEEE International Conference on Robotics & Automation, San Francisco, CA, April, pp. 891–897.

Lee, S., Jang, D., Kim, E., Hong, S., Han, J., 2005. A real-time 3D workspace modeling with stereo camera. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 2140–2147.

Lobo, J., Almeida, L., Alves, J., Dias, J., 2003. Registration and segmentation for 3D map building. In: Proc. IEEE International Conference on Robotics and Automation (ICRA '03), vol. 1, pp. 139–144.

McBride, J., Snorrason, M., Goodsell, T., Eaton, R., Stevens, M.R., 2005. Single camera stereo for mobile robot surveillance. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 3.

Miller, R., Amidi, O., 1998. 3-D site mapping with the CMU autonomous helicopter. In: Proc. 5th International Conference on Intelligent Autonomous Systems (IAS-5), June.

Olson, C.F., Abi-Rached, H., Ye, M., Hendrich, J.P., 2003. Wide-baseline stereo vision for Mars rovers. In: Proc. 2003 International Conference on Intelligent Robots and Systems (IROS 2003), vol. 2, pp. 1302–1307.

Rovira-Más, F., Zhang, Q., Reid, J.F., 2005. Creation of three-dimensional crop maps based on aerial stereoimages. Biosyst. Eng. 90 (3), 251–259, doi:10.1016/j.biosystemseng.2004.11.013.

Rovira-Más, F., Reid, J.F., Zhang, Q., 2006. Stereovision data processing with 3D density maps for agricultural vehicles. Transac. ASABE 49 (4), 1213–1222.

Se, S., Lowe, D., Little, J., 2002. Vision-based mapping with backward correction. In: Proc. 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems, EPFL, Lausanne, Switzerland, October, pp. 153–158.

Wang, L.K., Hsieh, S., Hsueh, E., Hsaio, F.B., Huang, K.Y., 2005. Complete pose determination for low altitude unmanned aerial vehicles using stereo vision. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp. 108–113.

Weast, A.B., Ota, J.M., Kitts, C.A., Bulich, C.A., Laurence, A.M., Lwin, C.M., Wigle, T.D., Perkins, W.B., Cook, J.F., 1999. Integrating digital stereo cameras with Mars Pathfinder technology for 3D regional mapping underwater. In: Proc. 1999 IEEE Aerospace Conference, Snowmass, CO, March, pp. 253–259.

Zabih, R., Woodfill, J., 1994. Non-parametric local transform for computing visual correspondence. In: Proc. of the third European Conference on Computer Vision, Stockholm, May, pp. 151–158.

Zhao, H., Aggarwal, J.K., 2000. 3D reconstruction of an urban scene from synthetic fish-eye images. In: IEEE Southwest Symposium on Image analysis and Interpretation, Austin, Texas, April.