12

Global 3D Terrain Maps for Agricultural Applications

Francisco Rovira-Más
Polytechnic University of Valencia
Spain

1. Introduction

At some point in life, everyone needs to use a map. Maps tell us where we are, what is around us, and what route needs to be taken to reach a desired location. Until very recently, maps were printed in paper and provided a two-dimensional representation of reality. However, most of the maps consulted at present are in electronic format with useful features for customizing trips or recalculating routes. Yet, they are still two-dimensional representations, although sometimes enriched with real photographs. A further stage in mapping techniques will be, therefore, the addition of the third dimension that provides a sense of depth and volume. While this excess of information may seem somewhat capricious for people, it may be critical for autonomous vehicles and mobile robots. Intelligent agents demand high levels of perception and thus greatly profit from three-dimensional vision. The widespread availability of global positioning information in the last decade has induced the development of multiple applications within the framework of precision agriculture. The main idea beyond this concept is to supply the right amount of input at the appropriate time for precise field locations, which obviously require the knowledge of field coordinates for site-specific applications. The practical implementation of precision farming is, consequently, tied to geographical references. However, prescription and information maps are typically displayed in two dimensions and generated with the level of resolution normally achieved with satellite-based imagery. The generation of *global three-dimensional (3D) terrain maps* offers all the advantages of *global localization* with the extra benefits of *high-resolution local perception* enriched with three dimensions plus color information acquired in real time.

Different kinds of three-dimensional maps have been reported according to the specific needs of each application developed, as the singular nature of every situation determines the basic characteristics of its corresponding 3D map. Planetary exploration, for example, benefits from virtual representations of unstructured and unknown environments that help scouting rovers to navigate (Olson et al., 2003; Wang et al., 2009); and the military forces, the other large group of users of 3D maps for recreating off-road terrains (Schultz et al., 1999), rely on stereo-based three-dimensional reconstructions of the world for a multiplicity of purposes. From the agricultural point of view, several attempts have been made to apply the mapping qualities of compact binocular cameras to production fields. Preceding the advent of compact cameras with real-time capabilities, something that took place at the turn of this century, airborne laser rangefinders allowed the monitoring of soil loss from gully erosion

by sensing surface topography (Ritchie & Jackson, 1989). The same idea of a laser map generator, but this time from a ground vehicle, was explored to generate elevation maps of a field scene (Yokota et al., 2004), after the fusion of several local maps with an RTK-GPS. Due to the fact that large extensions of agricultural fields require an efficient and fast way for mapping, unmanned aircrafts have offered a trade-off between low-resolution non-controllable remote sensing maps from satellite imagery and high-resolution ground-based robotic scouting. MacArthur et al. (2005), mounted a binocular stereo camera on a miniature helicopter with the purpose of monitoring health and yield in a citrus grove, and Rovira-Más et al. (2005) integrated a binocular camera in a remote controlled medium-size helicopter for general 3D global mapping of agricultural scenes. A more interesting and convenient solution for the average producer, however, consists of placing the stereo mapping engine on conventional farming equipment, allowing farmers to map while performing other agronomical tasks. This initiative was conceived by Rovira-Más (2003) — later implemented in Rovira-Más et al. (2008)—, and is the foundation for the following sections. This chapter explains how to create 3D terrain maps for agricultural applications, describes the main issues involved with this technique while providing solutions to cope with them, and presents several examples of 3D globally referenced maps.

2. Stereo principles and compact cameras

The geometrical principles of stereoscopy were set more than a century ago, but their effective implementation on compact off-the-shelf cameras with the potential to correlate stereo-based image pairs in real time, and therefore obtain 3D images, barely covers a decade. Present day compact cameras offer the best solution for assembling the mapping engine of an intelligent vehicle: favorable cost-performance ratio, portability, availability, optimized and accessible software, standard hardware, and continuously updated technology. The perception needs of today's 3D maps are mostly covered by commercial cameras, and very rarely will be necessary to construct a customized sensor. However, the fact that off-the-shelf solutions exist and are the preferred option does not mean that they can be simply approached as "plug and play." On the contrary, the hardest problems appear after the images have been taken. Furthermore, the configuration of the camera is a crucial step for developing quality 3D maps, either with retail products or customized prototypes. One of the early decisions to be made with regards to the camera configuration is whether using fixed baseline and permanent optics or, on the contrary, variable baselines and interchangeable lenses. The final choice is a trade-off between the high flexibility of the latter and the compactness of the former. A compact solution where imagers and lenses are totally fixed, not only offers the comfort of not needing to operate the camera after its installation but adds the reliability of precalibrated cameras. Camera calibration is a delicate stage for cameras that are set to work outdoors and onboard off-road vehicles. Every time the baseline is modified or a lens changed, the camera has to be calibrated with a calibration panel similar to a chessboard. This situation is aggravated by the fact that cameras on board farm equipment are subjected to tough environmental and physical conditions, and the slightest bang on the sensor is sufficient to invalid the calibration file comprising the key transformation parameters. The mere vibration induced by the diesel engines that power off-road agricultural vehicles is enough to unscrew lenses during field duties, overthrowing the entire calibration routine. A key matter is, therefore, finding out what is the best camera configuration complying with the expected needs in the field, such that a precalibrated rig

can be ordered with no risk of losing optimum capabilities. Of course, there is always a risk of dropping the precalibrated camera and altering the relative position between imagers, but this situation is remote. Nevertheless, if this unfortunate accident ever happened, the camera would have to be sent back to the original manufacturer for the alignment of both imaging sensors and, subsequently, a new calibration test. When the calibration procedure is carried out by sensor manufacturers, it is typically conducted under optimum conditions and the controlled environment of a laboratory; when it is performed *in-situ*, however, a number of difficulties may complicate the generation of a reliable calibration file. The accessibility of the camera, for example, can cause difficulties for setting the right diaphragm or getting a sharp focus. A strong sun or unexpected rains may also ruin the calculation of accurate parameters. At least two people are required to conduct a calibration procedure, not always available when the necessity arises. Another important decision to be made related to the calibration of the stereo camera is the size of the chessboard panel. Ideally, the panel should have a size such that when located at the targeted ranges, the majority of the panel corners are found by the calibration software. However, very often this results in boards that are too large to be practical in the field, and a compromise has to be found. Figure 1 shows the process of calibrating a stereo camera installed on top of the cabin of a tractor. Since the camera features variable baseline and removable lenses (Fig. 5b), it had to be calibrated after the lenses were screwed and the separation between imagers secured. Notice that the A-4 size of the calibration panel forces the board holder to be quite close to the camera; a larger panel would allow the holder to separate more from the vehicle, and consequently get a calibration file better adjusted to those ranges that are more interesting for field mapping. Section 4 provides some recommendations to find a favorable camera configuration as it represents the preliminary step to design a compact mapping system independent of weak components such as screwed parts and in-field calibrations.

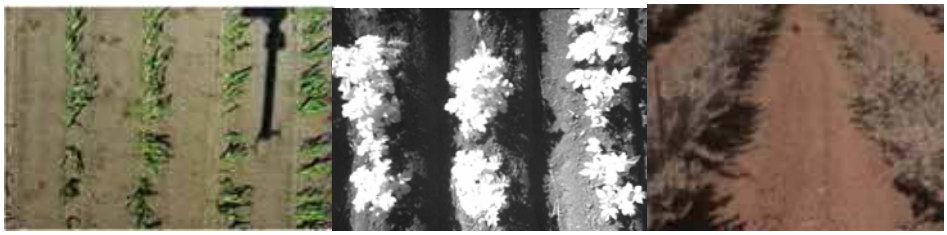
Fig. 1. Calibration procedure for non-precalibrated stereo cameras

The position of the camera is, as illustrated in Fig. 1, a fundamental decision when designing the system as a whole. The next section classifies images according to the relative position between the camera and the ground, and the system architectures discussed in Section 5 rely on the exact position of the sensor, as individual images need to be fused at the correct

position and orientation. It constitutes a good practice to integrate a mapping system in a generic vehicle that can perform other tasks without any interference caused by the camera or related hardware. Apart from the two basic configuration parameters —i. e. baseline and optics—, the last choice to make is the image resolution. It is obvious that the higher resolution of the image the richer the map; however, each pair of stereo images leads to 3D clouds of several thousand points. While a single stereo pair will cause no trouble for its virtual representation, merging the 3D information of many images as individual building blocks will result in massive and unmanageable point clouds. In addition, the vehicle needs to save the information in real time and, when possible, generate the map "on the fly." For this reason, high resolution images are discouraged for the practical implementation of 3D global mapping unless a high-end computer is available onboard. In summary, the robust solutions that best adapt to off-road environments incorporate precalibrated cameras with an optimized baseline-lenses combination and moderate resolutions as, for instance, 320 x 240 or 400 x 300.

3. Mapping platforms, image types, and coordinate transformations

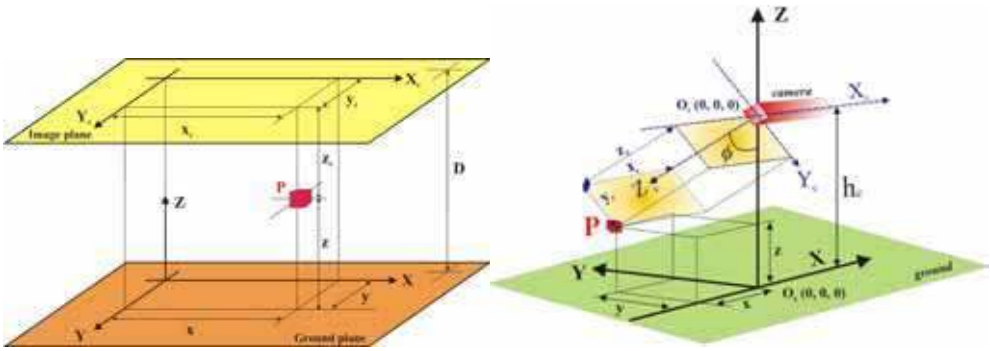
The final 3D maps should be independent of the type of stereo images used for their construction. Moreover, images taken under different conditions should all contribute to a unique globally-referenced final map. Yet, the position of the camera in the vehicle strengthens the acquisition of some features and reduces the perception of others. Airborne images, for instance, will give little detail on the position of tree trunks but, on the other hand, will cover the top of canopies quite richly. Different camera positions will lead to different kind of raw images; however, two general types can be highlighted: ground images and aerial images. The essential difference between them is the absence of perspective —and consequently, a vanishing point— in the latter. *Aerial images* are taken when the image plane is approximately parallel to the ground; and *ground images* are those acquired under any other relative position between imager and ground. There is a binding relationship between the vehicle chosen for mapping, the selected position of the camera, and the resulting image type. Nevertheless, this relationship is not exclusive, and aerial images may be grabbed from an aerial vehicle or from a ground platform, according to the specific position and orientation of the camera. Figure 2 shows an aerial image of corn taken from a remote-controlled helicopter (a), an aerial image of potatoes acquired from a conventional small tractor (b), and a ground image of grapevines obtained from a stereo camera mounted on top of the cabin of a medium-size tractor (c). Notice the sense of perspective and lack of parallelism in the rows portrayed in the ground-type image.



a b c

Fig. 2. Image types for 3D mapping: aerial (a and b), and ground (c)

The acquisition of the raw images (left-right stereo pairs) is an intermediate step in the process of generating a 3D field map, and therefore the final map must have the same quality and properties regardless of the type of raw images used, although as we mentioned above, the important features being tracked might recommend one type of images over the other. What is significantly different, though, is the coordinate transformation applied to each image type. This transformation converts initial camera coordinates into practical ground coordinates. The camera coordinates (x_c, y_c, z_c) are exclusively related to the stereo camera and initially defined by its manufacturer. The origin is typically set at the optical center of one of the lenses, and the plane X_cY_c coincides with the image plane, following the traditional definition of axes in the image domain. The third coordinate, Z_c, gives the depth of the image, i. e. the ranges directly calculated from the disparity images. The camera coordinates represent a generic frame for multiple applications, but in order to compose a useful terrain map, coordinates have to meet two conditions: first, they need to be tied to the ground rather than to the mobile camera; and second, they have to be globally referenced such that field features will be independent from the situation of the vehicle. In other words, our map coordinates need to be *global* and *grounded*. This need is actually accomplished through two consecutive steps: first from local camera coordinates (x_c, y_c, z_c) to local ground coordinates (x, y, z); and second, from local ground coordinates to global ground coordinates (e, n, z_g). The first step depends on the image type. Figure 3a depicts the transformation from camera to ground coordinates for aerial images. Notice that ground coordinates keep their origin at ground level and the z coordinate always represents the height of objects (point P in the figure). This conversion is quite straightforward and can be mathematically expressed through Equation 1, where D represents the distance from the camera to the ground. Given that ground images are acquired when the imagers plane of the stereo camera is inclined with respect to the ground, the coordinate transformation from camera coordinates to ground coordinates is more involving, as graphically represented in Figure 3b for a generic point P. Equation 2 provides the mathematical expression that allows this coordinate conversion, where h_c is the height of the camera with respect to the ground and ϕ is the inclination angle of the camera as defined in Figure 3b.



a b

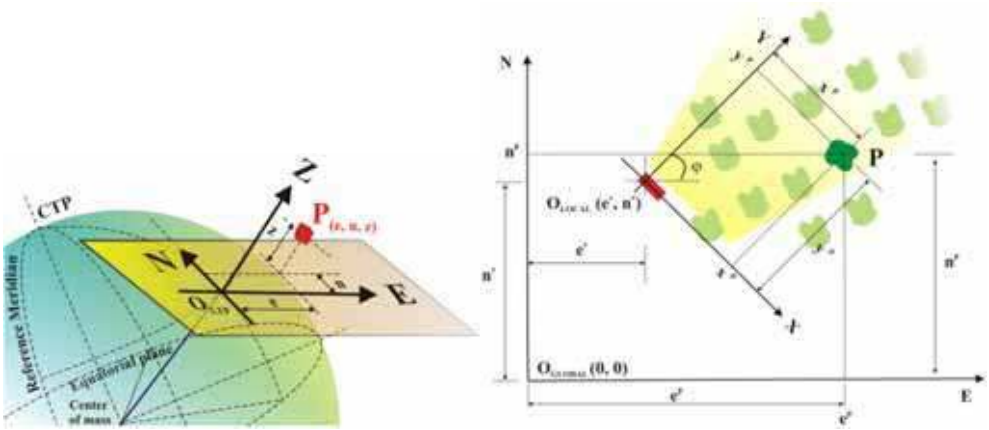
Fig. 3. Coordinate transformations from camera to ground coordinates for aerial images (a) and ground images (b)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + D \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\cos\varnothing & \sin\varnothing \\ 0 & -\sin\varnothing & -\cos\varnothing \end{bmatrix} \times \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} + h_c \times \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (2)$$

The transformation of equation 2 neglects roll and pitch angles of the camera, but in a general formulation of the complete coordinate conversion to a global frame, any potential orientation of the stereo camera needs to be taken into account. This need results in the augmentation of the mapping system with two additional sensors: an inertial measurement unit (IMU) for estimating the pose of the vehicle in real time, and a global positioning satellite system to know the global coordinates of the camera at any given time. The first transformation from camera coordinates to ground coordinates occurs at a local level, that is, the origin of ground coordinates after the application of Equations 1 and 2 is fixed to the vehicle, and therefore travels with it. The second stage in the coordinate transformation establishes a static common origin whose position depends on the global coordinate system employed. GPS receivers are the universal global localization sensors until the upcoming completion of Galileo or the full restoration of GLONASS. Standard GPS messages follow the NMEA code and provide the global reference of the receiver antenna in geodetic coordinates latitude, longitude, and altitude. However, having remote origins results in large and inconvenient coordinates that complicate the use of terrain maps. Given that agricultural fields do not cover huge pieces of land, the sphericity of the earth can be obviated, and a flat reference (ground) plane with a user-set origin results more convenient. These advantages are met by the *Local Tangent Plane* (ENZ) model which considers a flat surface containing the plane coordinates *east* and *north*, with the third coordinate (*height*) z_g perpendicular to the reference plane, as schematized in Figure 4a. Equation 3 gives the general expression that finalizes the transformation to global coordinates represented in the Local Tangent Plane. This conversion is applied to every single point of the local map —3D point cloud— already expressed in ground coordinates (x, y, z). The final coordinates for each transformed point in the ENZ frame will be (e, n, z_g). Notice that Equation 3 relies on the global coordinates of the camera's instantaneous position—center of the camera coordinate system—given by (e^c, n^c, z^c_g), as well as the height h_{GPS} at which the GPS antenna is mounted, and the distance along the Y axis between the GPS antenna and the camera reference lens d_{GPS}. The attitude of the vehicle given by the pitch (α), roll (β), and yaw (φ), has also been included in the general transformation equation (3) for those applications where elevation differences within the mapped field cannot be disregarded. Figure 4b provides a simplified version of the *coordinate globalization* for a given point P, where the vehicle's yaw angle is φ and the global position of the camera when the image was taken is determined by the point O_{LOCAL}. A detailed step-by-step explanation of the procedure to transform geodetic coordinates to Local Tangent Plane coordinates can be followed in Rovira-Más et al. (2010).

$$\begin{bmatrix} e \\ n \\ z_g \end{bmatrix} = \begin{bmatrix} e^c \\ n^c \\ z_g^c - h_{GPS} \cdot \cos\beta \cdot \cos\alpha \end{bmatrix} + \begin{bmatrix} \cos\varphi \cdot \cos\beta & \cos\varphi \cdot \sin\beta \cdot \sin\alpha - \sin\varphi \cdot \cos\alpha & \cos\varphi \cdot \sin\beta \cdot \cos\alpha + \sin\varphi \cdot \sin\alpha \\ \sin\varphi \cdot \cos\beta & \sin\varphi \cdot \sin\beta \cdot \sin\alpha + \cos\varphi \cdot \cos\alpha & \sin\varphi \cdot \sin\beta \cdot \cos\alpha - \cos\varphi \cdot \sin\alpha \\ -\sin\beta & \cos\beta \cdot \sin\alpha & \cos\beta \cdot \cos\alpha \end{bmatrix} \times \begin{bmatrix} x \\ y + d_{GPS} \\ z \end{bmatrix} \quad (3)$$



a b

Fig. 4. Local Tangent Plane coordinate system (a), and transformation from local vehicle-fixed ground frame XYZ to global reference frame ENZ (b)

4. Configuration of 3D stereo cameras: choosing baselines and lenses

It was stated in Section 2 that precalibrated cameras with fixed baselines and lenses provide the most reliable approach when selecting an onboard stereo camera, as there is no need to perform further calibration tests. The quality of a 3D image mostly depends on the quality of its corresponding depth map (disparity image) as well as its further conversion to three-dimensional information. This operation is highly sensitive to the accuracy of the calibration parameters, hence the better calibration files the higher precision achieved with the maps. However, the choice of a precalibrated stereo rig forces us to permanently decide two capital configuration parameters which directly impact the results: *baseline* and *focal length* of the lenses. In purity, stereoscopic vision can be achieved with binocular, trinocular, and even higher order of multi-ocular sensors, but binocular cameras have demonstrated to perform excellently for terrain mapping of agricultural fields. Consequently, for the rest of the chapter we will always consider binocular cameras unless noted otherwise.

Binocular cameras are actually composed of two equal monocular cameras especially positioned to comply with the stereoscopic effect and epipolar constriction. This particular disposition entails a common plane for both imagers (arrays of photosensitive cells) and the (theoretically) perfect alignment of the horizontal axes of the images (usually x). In practice, it is physically achieved by placing both lenses at the same height and one besides the other

at a certain distance, very much as human eyes are located in our heads. This inter-lenses separation is technically denominated the *baseline* (B) of the stereo camera. Human baselines, understood as inter-pupil separation distances, are typically around 60 - 70 mm. Figure 5 shows two stereo cameras: a precalibrated camera (a), and a camera with interchangeable lenses and variable baseline (b). Any camera representing an intermediate situation, for instance, when the lenses are removable but the baseline fixed, cannot be precalibrated by the manufacturer as every time a lens is changed, a new calibration file has to be immediately generated. The longer the baseline the further ranges will be acceptably perceived, and vice versa, short baselines offer good perceptual quality for near distances. Recall that 3D information comes directly from the disparity images, and no correlation can be established if a certain point only appears in one of the two images forming the stereo pair; in other words, enlarging the baseline increases the minimum distance at which the camera can perceive, as objects will not be captured by both images, and therefore pixel matching will be physically impossible. The effect of the *focal length* (f) of the lenses on the perceived scene is mainly related to the field of view covered by the camera. Reduced focal lengths (below 6 mm) acquire a wide field of view but possess lower resolution to perceive the background. Large focal lengths, say over 12 mm, are acute sensing the background but completely miss the foreground. The nature and purpose of each application must dictate the baseline and focal length of the definite camera, but these two fundamental parameters are coupled and should not be considered independently but as a whole. In fact, the same level of perception can be attained with different B-f combinations; so, for instance, 12 m ranges have been optimally perceived with a baseline of 15 cm combined with 16 mm lenses, or alternatively, with a baseline of 20 cm and either lenses of 8 mm or 12 mm (Rovira-Más et al., 2009). Needless to say that both lenses in the camera have to be identical, and the resolution of both imagers has to be equally set.

a b

Fig. 5. Binocular stereoscopic cameras: precalibrated (a), and featuring variable baselines and interchangeable lenses (b)

5. System architecture for data fusion

The coordinate transformation of Equation 3 demands the real time acquisition of the vehicle pose (roll, pitch, and yaw) together with the instantaneous global position of the camera for each image taken. If this information is not available for a certain stereo pair, the resulting 3D point cloud will not be added to the final global map because such cloud will lack a global reference to the common user-defined origin. The process of building a global

3D map from a set of stereo images can be schematized in the pictorial of Figure 6. As represented below, the vehicle follows a —not always straight— course while grabbing stereo images that are immediately converted to 3D point clouds. These clouds of points are referenced to the mapping vehicle by means of the ground coordinates of each point as defined in Figure 3. As shown in the left side of Figure 6, every stereo image constitutes a local map with a vehicle-fixed ground coordinate system whose pose with relation to the global frame is estimated by an inertial sensor, and whose origin's global position is given by a GPS receiver. After all the points in the 3D cloud have been expressed in vehicle-fixed ground coordinates (Equations 1 and 2), the objective is to merge the local maps in a unique global map by reorienting and patching the local maps together according to their global coordinates (Equation 3). The final result should be coherent, and if for example the features perceived in the scene are straight rows spaced 5 m, the virtual global map should reproduce the rows with the same spacing and orientation, as schematically represented in the right side of Figure 6.

Fig. 6. Assembly of a global map from individual stereo-based local maps

The synchronization of the local sensor —stereo camera— with the attitude and positioning sensors has to be such that for every stereo image taken, both inertial measurements (α, β, φ) and geodetic coordinates are available. Attitude sensors often run at high frequencies and represent no limitations for the camera, which typically captures less than 30 frames per second. The GPS receiver, on the contrary, usually works at 5 Hz, which can easily lead to the storage of several stereo images (3D point clouds) with exactly the same global coordinates. This fact requests certain control in the incorporation of data to the global map, not only adjusting the processing rate of stereo images to the input of GPS messages, but considering as well the forward speed of the mapping vehicle and the field of view covered by the camera. Long and narrow fields of view (large B and large f) can afford longer sampling rates by the camera as a way to reduce computational costs at the same time overlapping is avoided. In addition to the misuse of computing resources incurred when overlapping occurs, any inaccuracy in either GPS or IMU will result in the appearance of artifacts generated when the same object is perceived in various consecutive images poorly transformed to global coordinates. That phenomenon can cause, for example, the representation of a tree with a double trunk. This issue can only be overcome if the mapping engine assures that all the essential information inserted in the global map has been acquired with acceptable quality levels. As soon as one of the three key sensors produces unreliable data, the assembly of the general map must remain suspended until proper data reception is resumed. Sensor noise has been a common problem in the practical generation of field maps, although the temporal suspension of incoming data results in incomplete, but correct, maps, which can be concluded in future missions of the vehicle. There are many ways to be aware of, and ultimately palliate, sensor inaccuracies. IMU drift can be assessed with the yaw estimation calculated from GPS coordinates. GPS errors can be reduced with

the subscription to differential signals, and by monitoring quality indices such as dilution of precision or the number of satellites in solution. Image noise is extremely important for this application as perception data constitute the primary source of information for the map; therefore, it will be separately covered in the next section. The 3D representation of the scene, composed of discrete points forming a cloud determined by stereo perception, can be rendered in false colors, indicating for example the height of crops or isolating the objects located at a certain placement. However, given that many stereo cameras feature color (normally RGB) sensors, each point P can be associated with its three global coordinates plus its three color components, resulting in the six-dimensional vector $(e, n, z_g, r, g, b)_P$. This 3D representation maintains the original color of the scene, and besides providing the most realistic representation of that scene, also allows the identification of objects according to their true color. Figure 7 depicts, in a conceptual diagram, the basic components of the architecture needed for building 3D terrain maps of agricultural scenes.

Fig. 7. System architecture for a stereo-based 3D terrain mapping system

6. Image noise and filters

Errors can be introduced in 3D maps at different stages according to the particular sensor yielding spurious data, but while incorrect position or orientation of the camera may be detected and thus prevented from being added to the global map, image noise is more difficult to handle. To begin with, the perception of the scene totally relies on the stereo camera and its capability to reproduce the reality enclosed in the field of view. When

correlating the left and right images of each stereo pair, mismatches are always present. Fortunately, the majority of miscorrelated pixels are eliminated by the own filters embedded in the camera software. These unreliable pixels do not carry any information in the disparity image, and typically represent void patches as the pixels mapped in black in the central image of Fig. 8. However, some mismatches remain undetected by the primary filters and result in disparity values that, when transformed to 3D locations, point at unrealistic positions. Figure 8 shows a disparity image (center) that includes the depth information of some clouds in the sky over an orchard scene (left). When the clouds were transformed to 3D points (right), the height of the clouds was obviously wrong, as they were place below 5 m. The occurrence of outliers in the disparity image is strongly dependent on the quality of the calibration file, therefore precalibrated cameras present an advantage in terms of noise. Notice that a wrong GPS message or yaw estimation automatically discards the entire image, but erroneously correlated pixels usually represent an insignificant percentage of the point cloud and it is neither logical nor feasible to reject the whole image (local map). A practical way to avoid the presence of obvious outliers in the 3D map is by defining a *validity box* of logical placement of 3D information. So, when mapping an orchard, for instance, negative heights make no sense (underground objects) and heights over the size of the trees do not need to be integrated in the global map, as they very likely will be wrong. In reality, field images are rich in texture and disparity mismatches represent a low percentage over the entire image. Yet, the information they add is so wrong that it is worth removing them before composing the global map, and the definition of a validity box has been effective to do so.

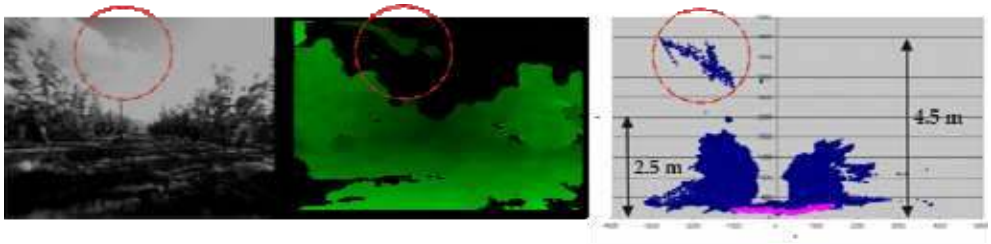


Fig. 8. Correlation errors in stereo images

7. Real-time 3D data processing

The architecture outlined in Figure 7 is designed to construct 3D terrain maps "on the fly", that is, while the vehicle is traversing the field, the stereo camera takes images that are converted to 3D locally-referenced point clouds, and in turns added to a global map after transforming local ground coordinates to global coordinates by applying Equation 3. The result is a large text file with all the points retrieved from the scene. This file, accessible after the mapping mission is over, is ready for its virtual representation. This *online* procedure of building a 3D map strongly relies on the appropriate performance of localization and attitude sensors. An alternative method to generate a 3D field map is when its construction is carried out *off-line*. This option is adequate when the computational power onboard is not sufficient, if memory resources are scarce, or if some of the data need preprocessing before the integration in the general map. The latter has been useful when the attitude sensor has

been inaccurate or not available. To work offline, the onboard computer needs to register a series of stereo images and the global coordinates at which each stereo image was acquired. A software application executed in the office transforms all the points in the individual images to global coordinates and appends the converted points to the general global map. The advantage of working off-line is the possibility of removing corrupted data that passed the initial filters. The benefit of working on-line is the availability of the map right after the end of the mapping mission.

8. Handling and rendering massive amounts of 3D Data

The reason behind the recommendation of using moderate resolutions for the stereo images is based on the tremendous amount of data that gets stored in 3D field maps. A typical 320 x 240 image can easily yield 50000 points per image. If a mapping vehicle travels at 2 m/s (7 km/h), it will take 50 s to map a 100 m row of trees. Let us suppose that images are acquired every 5 s, or the equivalent distance of 10 m; then the complete row will require 10 stereo images which will add up to half million points. If the entire field comprises 20 rows, the whole 3D terrain map will have *10 million points*. Such a large amount of data poses serious problems when handling critical visual information and for efficiently rendering the map. *Three-dimensional virtual reality chambers* are ideal to render 3D terrain maps. Inside them, viewers wear special goggles which adapt the 3D represented environment to the movement of the head, so that viewers feel like they were actually immersed in the scene and walking along the terrain. Some of the examples described in the following section were run in the John Deere Immersive Visualization Laboratory (Moline, IL, USA). However, this technology is not easily accessible and a more affordable alternative is necessary to make use of 3D maps with conventional home computers. Different approaches can be followed to facilitate the management and visualization of 3D maps. Many times the camera captures information that is not essential for the application pursued. For example, if the objective is to monitor the growth of canopies or provide an estimate of navigation obstacles, the points in the cloud that belong to the ground are not necessary and may occupy an important part of the resources. A simple redefinition of the validity box will only transfer those points that carry useful information, reducing considerably the total amount of points while maintaining the basic information. Another way of decreasing the size of file maps is by enlarging the spacing between images. This solution requires an optimal configuration of the camera to prevent the presence of gaps lacking 3D information. When all the information in the scene is necessary, memory can be saved by condensing the point cloud in regular grids. In any case, a mapping project needs to be well thought in advance because not only difficulties can arise in the process of map construction but also in the management and use that comes afterwards. There is no point in building a high-accuracy map if no computer can ever handle it at the right pace. More than being precise, 3D maps need to fulfill the purpose for which they were originally created.

9. Examples

The following examples provide general-purpose agricultural 3D maps generated by following the methodology developed along the chapter. In order to understand the essence of the process, it is important to pay attention to the architecture of the system on one hand, and to the data fusion on the other. No quality 3D global map can be attained unless both

local and global sensors perform acceptably. When the assembly of the map is carried out off-line, it is helpful to plot first the positions of the stereo camera where images were taken to make sure that global coordinates are correct. Only when this happens, we can proceed with the integration of the independent images into the global map by performing all the coordinate transformations given before. The scene portrayed in Figure 9 represents a series of wooden posts used as supporting structures for growing hops. Posts were either straight up or inclined, but they were equally spaced, which gave us a good opportunity to test if the transformation from local to global provided coherent results. The images were taken with two commercial stereo cameras; one with a (9 cm) fixed baseline but removable 4.8 mm lenses, and the other with fixed lenses and (22 cm) baseline, both mounted on the front of a utility off-road vehicle. A GPS receiver provided global references and no attitude sensor was integrated in the vehicle, which forced the map construction to be off-line. The field was flat and both roll and pitch were negligible. The yaw angle was directly estimated from the GPS-determined trajectory followed by the vehicle (Fig. 9, top-center). This map was created without including the true color code of each correlated point, and as a result its 3D representation is only available in false colors. It proved that 3D terrain mapping with the degree of detail given by an in-field stereo camera plus the benefits of global references is feasible as long as the architecture of the system is properly designed.

The perceptual data of the second map, given in Figure 10, was acquired with a precalibrated camera, with a fixed baseline of 22 cm and integrated miniature lenses. The presence of a fiber optic gyroscope (FOG) and an RTK-GPS in the vehicle allowed for on-line map construction, following the architecture schematized in Figure 7. As color images were saved and decoded, true color was incorporated to the final global map as rendered in the multiple views included in Figure 10. This scene represents a turf lane bounded by two rows of trees, although the map mostly includes the right row. On the whole, 12 images were acquired to cover the approximately 30 m of the row, summing up a total of 379,517 points. The GPS-based camera positions show that the vehicle basically moved straight in the direction W-E. The top view demonstrates an adequate performance of the IMU when measuring yaw angles. The side view shows that the terrain was fairly even, as corroborated by the pitch angles estimated with the FOG, always below 3^{o}. The availability of true color helps to discriminate the yellowish turf from the darker hues of the trees. However, note that the color of the pixels corresponding to the top of the trees got confused with that of the pixels representing the sky, and consequently they display a false white color.

10. Conclusion

The adoption of new technologies by agricultural producers is a matter of time. Thirty years ago the GPS was a military device; today, it is offered by all major manufacturers of farm equipment. Social and economic problems in the production of competitive agricultural goods are pushing these technologies even further, especially precision farming and agricultural robotics. GPS-based automatic guidance solutions are being implemented in the field, and precision agriculture applications are successfully combined with conventional production systems. However, most of the maps utilized so far are two-dimensional and lack the level of detail that can be reached with a stereoscopic camera. These cameras have been successfully used in other robotic applications, but the ever growing need of handling richer and more updated information for better and faster decision making in the competitive agriculture of the future makes them a favorite resource.

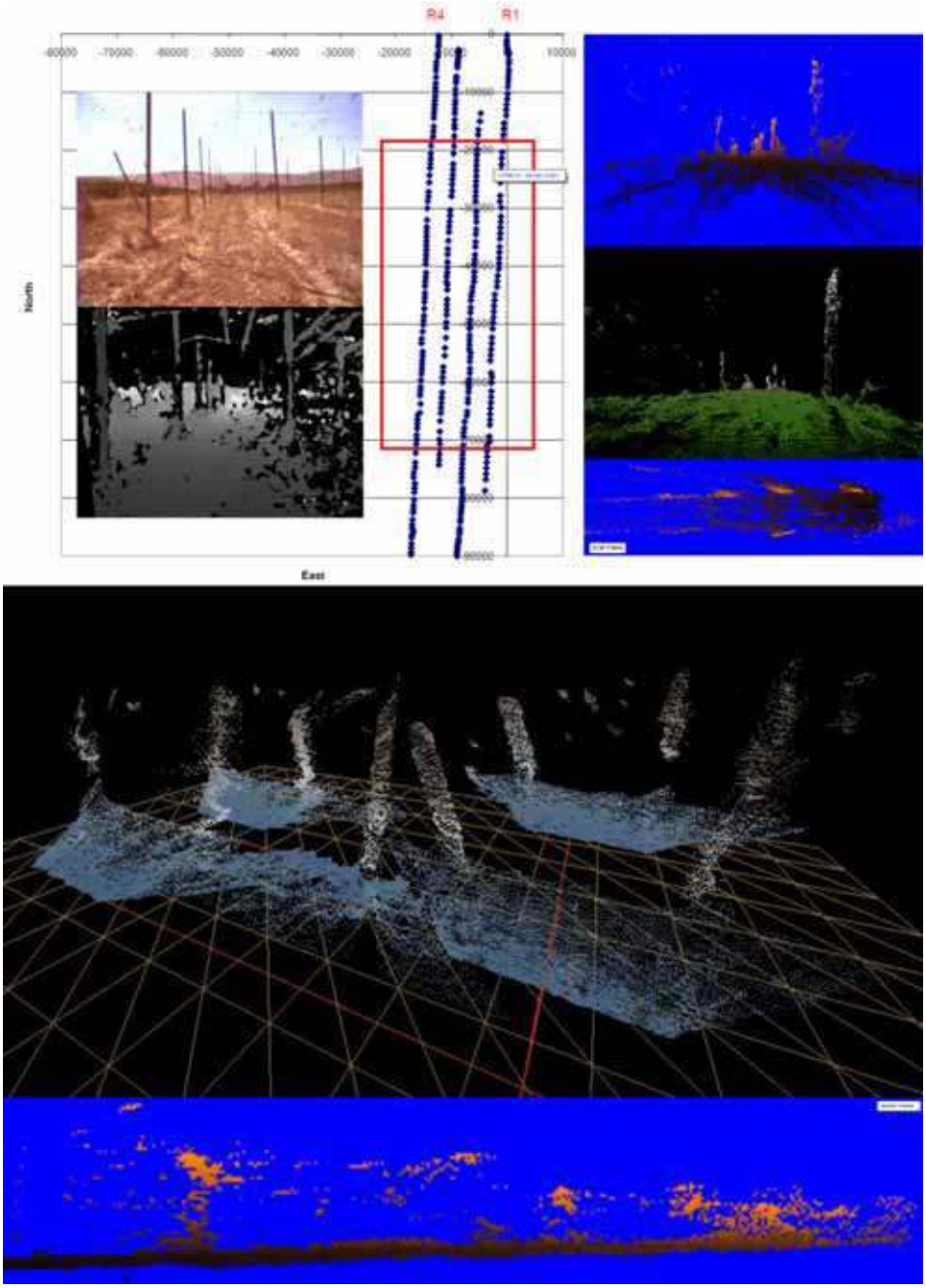


Fig. 9. Three-dimensional terrain map of a barren field with crop-supporting structures

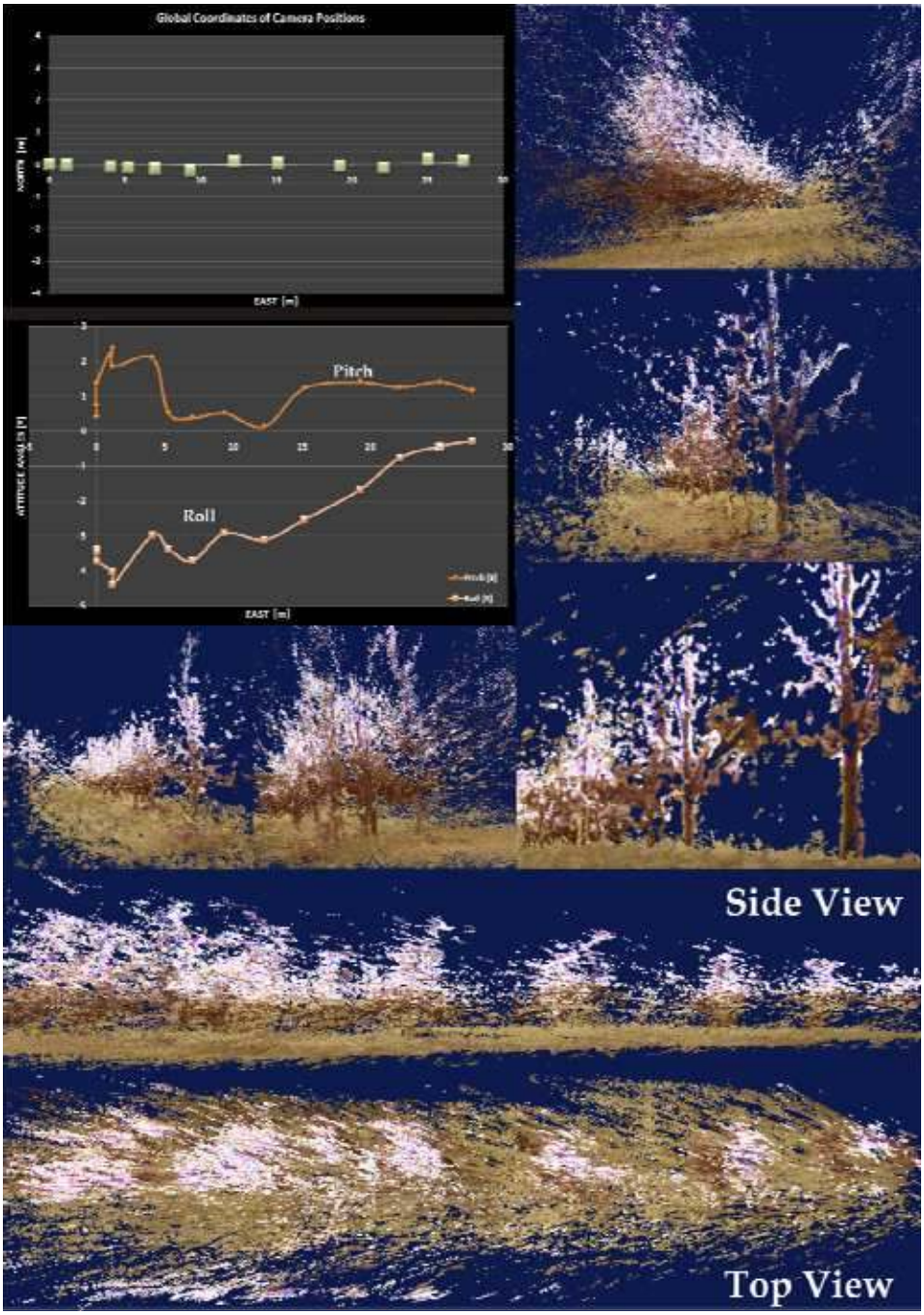


Fig. 10. True color three-dimensional terrain map

Yet, three-dimensional field information without a general frame capable of providing global references is not very practical. For that reason, the methodology elaborated along the chapter provides a way to build globally-referenced maps with the highest degree of visual perception, that in which human vision is based on. This theoretical framework, dressed with numerous practical recommendations, facilitates the physical deployment of real 3D mapping systems. Although not in production yet, the information attained with these systems will certainly help to the development and progress of future generations of intelligent agricultural vehicles.

11. References

MacArthur, D. K.; Schueller, J. K. & Crane, C. D. (2005). Remotely-piloted mini-helicopter imaging of citrus. ASAE Publication 051055, ASABE, St. Joseph, MI

Olson, C. F.; Abi-Rached, H.; Ye, M. & Hendrich, J. P. (2003). Wide-baseline stereo vision for Mars rovers, *Proceedings of the International Conference on Intelligent Robots and Systems*, pp. 1302-1307, IEEE

Ritchie, J. C. & Jackson T. J. (1989). Airborne laser measurements of the surface topography, *Transactions of the ASAE*, Vol. 32(2), pp. 645-658

Rovira-Más, F. (2003). *Applications of stereoscopic vision to agriculture*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign

Rovira-Más, F.; Zhang, Q. & Reid, J. F. (2005). Creation of three-dimensional crop maps based on aerial stereoimages, *Biosystems Engineering*, Vol. 90(3), pp. 251-259

Rovira-Más, F.; Zhang, Q. & Reid, J. F. (2008). Stereo vision three-dimensional terrain maps for precision agriculture, *Computers and Electronics in Agriculture*, Vol. 60, pp. 133-143

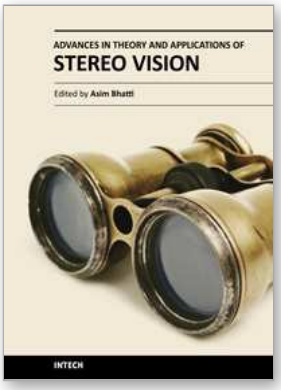
Rovira-Más, F.; Wang, Q. & Zhang, Q. (2009). Design parameters for adjusting the visual field of binocular stereo cameras, *Biosystems Engineering*, Vol. 105, pp. 59-70

Rovira-Más, F.; Zhang, Q. & Hansen A. C. (2010). *Mechatronics and intelligent systems for off-road vehicles*, Springer, UK, Chapter 3

Schultz, H.; Riseman, E. M.; Stolle, F. R. & Woo, D. (1999). Error detection and DEM fusion using self-consistency, *Proceedings of the Seventh IEEE International Conference on Computer Vision*, pp. 1174-1181, Vol. 2, IEEE

Wang, W.; Shen, M.; Xu, J.; Zhou, W. & Liu, J. (2009). Visual traversability analysis for micro planetary rover, *Proceedings of the International Conference on Robotics and Biomimetics*, pp. 907-912, Guilin, China, December 2009, IEEE

Yokota, M.; Mizushima, A.; Ishii, K. & Noguchi, N. (2004). 3-D map generation by a robot tractor equipped with a laser range finder. *Proceedings of the Automatic Technology for Off-road Equipment Conference*, pp. 374-379, Kyoto, Japan, October 2004, ASAE Publication 701P1004, ASABE, St. Joseph, MI



Advances in Theory and Applications of Stereo Vision
Edited by Dr Asim Bhatti

ISBN 978-953-307-516-7
Hard cover, 352 pages
Publisher InTech
Published online 08, January, 2011
Published in print edition January, 2011

The book presents a wide range of innovative research ideas and current trends in stereo vision. The topics covered in this book encapsulate research trends from fundamental theoretical aspects of robust stereo correspondence estimation to the establishment of novel and robust algorithms as well as applications in a wide range of disciplines. Particularly interesting theoretical trends presented in this book involve the exploitation of the evolutionary approach, wavelets and multiwavelet theories, Markov random fields and fuzzy sets in addressing the correspondence estimation problem. Novel algorithms utilizing inspiration from biological systems (such as the silicon retina imager and fish eye) and nature (through the exploitation of the refractive index of liquids) make this book an interesting compilation of current research ideas.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821