Photo: Monopoly919 (AdobeStock)

Computer vision applied to agriculture

6

Thiago Teixeira Santos | Jayme Garcia Arnal Barbedo | Sônia Ternes | João Camargo Neto | Luciano Vieira Koenigkan | Kleber Xavier Sampaio de Souza

Introduction

Computer vision, in a simple and comprehensive definition, is a field of artificial intelligence dedicated to extracting information from digital images. In the context of digital agriculture, computer vision can be used in the detection of diseases and pests, in yield estimation and in the non-invasive evaluation of attributes such as quality, appearance and volume, and it is also an essential component in agricultural robotic systems. According to Duckett et al. (2018), field robotics could enable a new range of agricultural equipment: small and intelligent machines capable of reducing waste and environmental impact[1] and providing economic viability, thus increasing food sustainability. Also according to Duckett et al. (2018), there is considerable potential to increase the window of opportunity for interventions, for example, in wet soil operation, night operation and constant crop monitoring.

A class of problems addressed by computer vision are the alleged perceptual problems: the detection and classification of patterns in images that are associated with an object of interest, as for instance fruits (Sa et al., 2016; Santos et al., 2020), animals (Barbedo et al., 2019) or symptoms of diseases and pests (Ferentinos, 2018; Barbedo, 2019).

Constant and efficient monitoring can be carried out based on images captured by field teams or obtained by cameras attached to tractors, implements, robots or drones: the search for crop or livestock anomalies; the evaluation of crop spatial variability for intervention, according to the precepts of precision agriculture; and autonomous action by machines and implements. Figure 1 shows an example of detection of grape bunches in images obtained from vineyards.

[1] Due to the sparing and intelligent use of pesticides or simply mechanical intervention: the physical removal of pests

Photos: Thiago Teixeira Santos

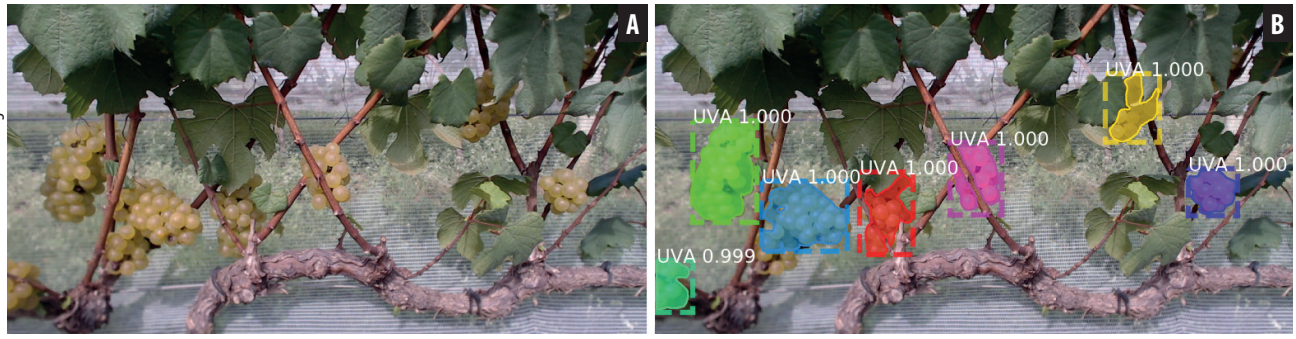


Figure 1. Examples of a perceptual task, the detection of grapes in images: image taken in a winery of a Chardonnay vine (A); detection result using a neural network (B).

Illustration: Thiago Teixeira Santos

Another class of problems are geometric ones. In forming an image, the light captured by the lens is projected onto a surface so that the three-dimensional scene produces a 2-D representation. Much of the scene structure is in the image, but depth information (the distance between the camera and the objects in the scene) is lost. One of the greatest contributions of geometric computer vision was the development of algorithms for recovering lost three-dimensional information from a set of images of the same scene. This is one of the most widely used computer vision applications in the market today: three-dimensional mapping and the production of maps from imagery obtained by Unmanned Autonomous Vehicles (UAVs – popularly known as drones, see Figure 2). Methodologies based on geometric computer vision have been employed in geological studies (Westoby et al., 2012), in pasture height assessment (Forsmoo et al., 2018) and in crop mapping (Comba et al., 2018), among other uses. Commercially, it is the core technology behind 3-D mapping and reconstruction services by UAVs extensively used in agriculture, such as Pix4D mapper and Agisoft PhotoScan/Metashape.

There is a growing number of computer vision applications in agricultural research. Consider, for example, the journal *Computers and Electronics in Agriculture*, which specializes in new software, hardware, and electronics applications in agriculture. A search for articles related to computer vision reveals that 23.7% of all works published in 2018 are associated with computer vision, rising to 29.1% in 2019. From January to June 2020, 115 of the 319 works (36.0%) published are related to computer vision. This volume of articles also translates into impact: of the 25 most cited works by June 2020, 14 are computer vision applications. Some simple factors explain this growth. Digital cameras are affordable and widely available devices in various configurations, easily integrated into larger systems (such as smartphones and UAVs). The advances in algorithms and hardware over the last ten years are reflected in the current dynamism of the area.

The next sections will present the recent innovations in the application of computer vision to agriculture, focusing on the contributions by Embrapa Digital Agriculture over the last 3 years. These advances are the result from both perceptual computer vision, the recognition of elements in the scene (Section 2), and geometric computer vision, the retrieval of three-dimensional information from images (Section 3). The combination of both fronts (Section 4) opens the way for systems that can perform highly complex operations, such as field robotics. Section 5 closes the chapter with some final remarks.

Perception: pattern recognition in images

Pattern recognition can be seen as the role of finding a representation for the pattern sought that is sufficiently versatile to cover observable variations, yet simple enough to be processed in a timely

Figure 2. UAV mapping: images are used to identify the three-dimensional structure of the area, and the position and orientation of the aircraft, displayed in red (A); the geolocated three-dimensional model is then projected onto a plane, forming a map. (B).

Illustration: Thiago Teixeira Santos

manner by the machine. In other words, it is an adequate pattern description to allow the machine to find it in the input data, yet succinct so that its interpretation is carried out within operating time constraints.

Visual patterns in natural images can be incredibly intricate, with regularities and variations that are difficult to describe. In agriculture, patterns assumed by fruits, leaves, grains, plants and symptoms of pathologies exhibit enormous variability, amplified by differences in lighting, position, occlusion and different sources of noise (dirty lenses, dust, interference, etc.). Figure 3 illustrates some of the difficulties a fruit detection system faces in real field growing conditions: severe occlusion between fruits, leaves and branches; color similarity between green fruits and the canopy; lighting variations between images; specular reflection (direct reflection of sunlight that saturates the camera sensor); and focus problems. Notwithstanding some success from the use of machine learning techniques (Gongal et al., 2015), pattern recognition in natural images began to reach high levels of accuracy with the arrival of convolutional neural networks (Lecun et al., 2015), quickly adopted for image recognition in agriculture (Kamilaris; Prenafeta-Boldú, 2018).

In neural networks, an architecture or model is a sequence of modules that perform simple operations on the data so that a module receives data from previous modules and propagates the result of its operations to the following modules. In computer vision, the most used neural networks are the convolutional neural networks (CNNs), in which the main operation employed is convolution, a linear combination of values in the vicinity of the input pixels. Neural networks are said to be deep if there is a large sequence of linked modules. The deeper the network, the greater its ability to learn representations for complex patterns, since each module is able to compose the representations of previous modules in a hierarchy. In the case of images, there is an intuitive interpretation for this behavior: the initial modules are able to find lines and edges of objects, the following modules compose these patterns into simple textures and structures like triangles and spots, which are then combined into other structures like parts of leaves, branches and berries. Finally, the final modules combine these elements into objects of interest: a plant, a bunch of grapes, an ox.

Photos: João Camargo Neto

Figure 3. Examples of the difficulties faced in fruit detection. In the images, we can observe problems of focus, specular reflection, severe occlusion by leaves, branches and other fruits, light variations and similarities in the color pattern between fruits and leaves.

The modules have parameters that need to be adjusted so that the joint operation of the entire network produces the expected results. A frequently used metaphor is to imagine that each parameter is adjusted by a dimmer. Adjusting a neural network would be performing the adjustment of millions of dimmers, each of which could affect the pattern recognition performance. Manually, however, this adjustment would be impractical and virtually impossible. The training of neural networks is an automated process for adjusting these parameters, so that the network "learns" the appropriate representations for the recognition problem in question.

In supervised learning of image patterns, this training is carried out using observations, images whose desired answer is known ("there is an orange in this image", "there are signs of coffee rust on this leaf"). This training requires thousands of observations, which is directly linked to the size of the network: more parameters require more observations, although it is difficult to determine an exact relationship between the number of parameters and the number of observations required. When the network processes the input image, the produced result is compared to the expected result, and their error is computed. The parameters are then adjusted to reduce the previous error, in a process known as backpropagation (Goodfellow et al., 2016). In practice, observations are grouped into batches, the network processes the batch and the observed error is computed. The backpropagation algorithm is used to adjust the parameters, starting with the final modules of the network and proceeding towards the parameters of the initial modules (hence the name of the procedure). Training proceeds with the next batch, and the procedure is repeated until the error reaches an observable minimum[2]. In short, deep neural networks automate the process of searching for adequate representations in pattern recognition problems, provided there is a sufficiently large set of observations for training in order to adequately represent the variability of the intended pattern. It is precisely this ability that makes the methodology so attractive to the intricate problems of recognition in agriculture.

Identification of plant diseases

The detection and classification to diagnose disease, pests and plant nutritional deficiencies in images are of great interest in agriculture. Automatic detection enables constant monitoring and searching for crop anomalies, based on images captured by field teams or obtained by cameras attached to tractors, implements, robots or UAVs. On the other hand, classification associates the detected anomalies to the disease, deficiency or pest, assisting the producer in the correct intervention. Neural networks can be used in both tasks, even simultaneously.

As seen above, thousands of observations are required before a neural network is able to produce accurate results. This need is amplified for plant disease recognition due to the large number of combinations resulting from the crossing between target cultures, pathologies, stage of disease development and imaging condition (manual collection, aerial monitoring by UAVs, capture at the ground level by machine, camera position, among others). This situation points to the need for large shared databases (Barbedo, 2018; Ferentinos, 2018), as considerable effort is required for their production.

The process of collecting and annotating the images, in other words, associating each image with the desired result for the supervised learning stage, is usually lengthy and costly. However, some strategies can be used to increase the number of observations. Barbedo (2019) showed that multiple lesions

[2] The ideal error would be zero, but there is no guarantee that an architecture will be able to achieve this. It is also an open problem to determine *a priori* what is the smallest error a network will be able to achieve for a given training set.

of the same pathologies which occur on the same leaf can be exploited to increase the number of observations from the same collection. Several examples of symptoms can be obtained from a single leaf or plant tissue sample, as seen in Figure 4. This strategy allowed an original database, containing 1575 observations (Barbedo, 2018), to be expanded to 46409 observations (Barbedo, 2019), producing gains in disease classification accuracy of, on average, 12%.

Photos: Jayme Garcia Arnal Barbedo

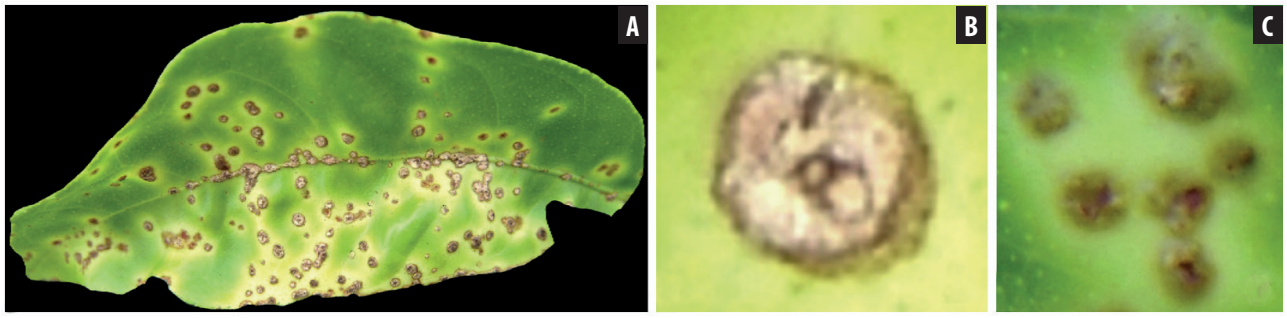


Figure 4. Examples of observations used in training systems for plant disease recognition: a sample of a diseased leaf, collected in the field (A); an observation of symptoms associated with the pathology (B); clusters of symptoms that also form a discernible pattern associated with the pathology (C).

Barbedo (2019) showed that a convolutional neural network, the GoogLeNet architecture (Szegedy et al., 2015), can be applied in the classification of many pathologies in different cultures, reaching accuracy values of 80% (passion fruit) up to 100% (cassava, cabbage, cotton, wheat, and sugarcane), as shown in Table 1. The database used, termed as Digipathos, was made publicly available[3]. Although the classification results are promising, there are still major challenges, especially with regard to detection ("are there symptoms present in the observation?"), which is crucial in autonomous monitoring for pest and disease management, but which still does not present the same classification accuracy ("what is the pathology for the observed symptom?"). In his experiments, Barbedo (2019) shows that accurate detections can be produced when symptoms are already severe, but not when the symptoms are still mild or do not occupy large portions of plant tissue, which is the ideal time for intervention by the farmer. False positive detection errors (healthy tissue detected as diseased) are often caused by factors such as the presence of dust, debris or even water droplets. It is also not clear yet what number of samples is needed so that

Table 1. Accuracy of the classification of pathologies in different cultures. For the cassava and kale images, the accuracy reached 100% in all tests.

Crop	Number of images	Accuracy (%)
Bean	3,079	94 ± 0.8
Cassava	895	100 ± 0.0
Citrus	1,868	96 ± 0.6
Coconut	1,504	98 ± 0.6
Corn	10,480	75 ± 4.4
Coffee	1,899	89 ± 1.9
Cotton	2,023	99 ± 0.3
Cashew	4,509	98 ± 0.5
Grape	2,330	96 ± 0.8
Kale	196	100 ± 0.0
Passion fruit	280	80 ± 4.2
Soy	13,733	87 ± 3.6
Sugar cane	2773	99 ± 0.4
Wheat	840	99 ± 0.5
Total	46,135	94 ± 2.0

Source: Adapted from Barbedo (2019).

[3] Available in: https://www.digipathos-rep.cnptia.embrapa.br

the characteristics of symptoms can be properly learned by neural networks (still an open question in computer vision in general).

Detection of animals in pastures

Barbedo et al. (2019) present an example of how UAV technologies and computer vision can be combined for monitoring large areas, for example detecting cattle in extensive livestock production. Given the dynamics of the animals and the enormous size of the pasture areas, the ranchers face great difficulties monitoring the herds in the pastures.

A database composed of 1853 images containing 8629 Canchim animals was produced based on images obtained by a commercially available quadrirotor[4]. Barbedo et al. (2019) tested 15 different neural network architectures at 3 distinct spatial resolutions (1, 2 cm/pixel and 4 cm/pixel), in order to analyze the performance resulting from different flight heights. The results showed that most of the tested architectures were able to reach high levels of accuracy, above 95%. The NasNet architecture (Zoph et al., 2018), a very deep network with great capacity to learn complex patterns, achieved accuracy close to 100%. These results are expressive, especially considering the complexity of the problem, as shown in Figure 5: several situations, from severe occlusion by trees and drinking fountains to differences in

Photos: Jayme Garcia Arnal Barbedo

Figure 5. Examples of situations observed in the detection of animals in pastures: animal in high pasture (A); dry pasture (B); exposed soil (C); tree occlusions (D); covering of drinking fountains (E) and electrical cables (F).

[4] In this case, a DJI Phantom 4 Pro vehicle.

lighting and pasture conditions, in addition to the position and disposition of the animals, all of which present highly variable situations. Even so, the accuracy of most of the architectures tested is expressive. Another particularly interesting effect from an operational point of view was that most models present better results at the 2 cm/pixel resolution and not at the maximum 1 cm/pixel resolution, which may be due to the resolution of the convolutional modules with which these architectures were originally designed. In practice, this enables flights at higher heights, which allows covering areas in less time.

Detection and counting of fruits

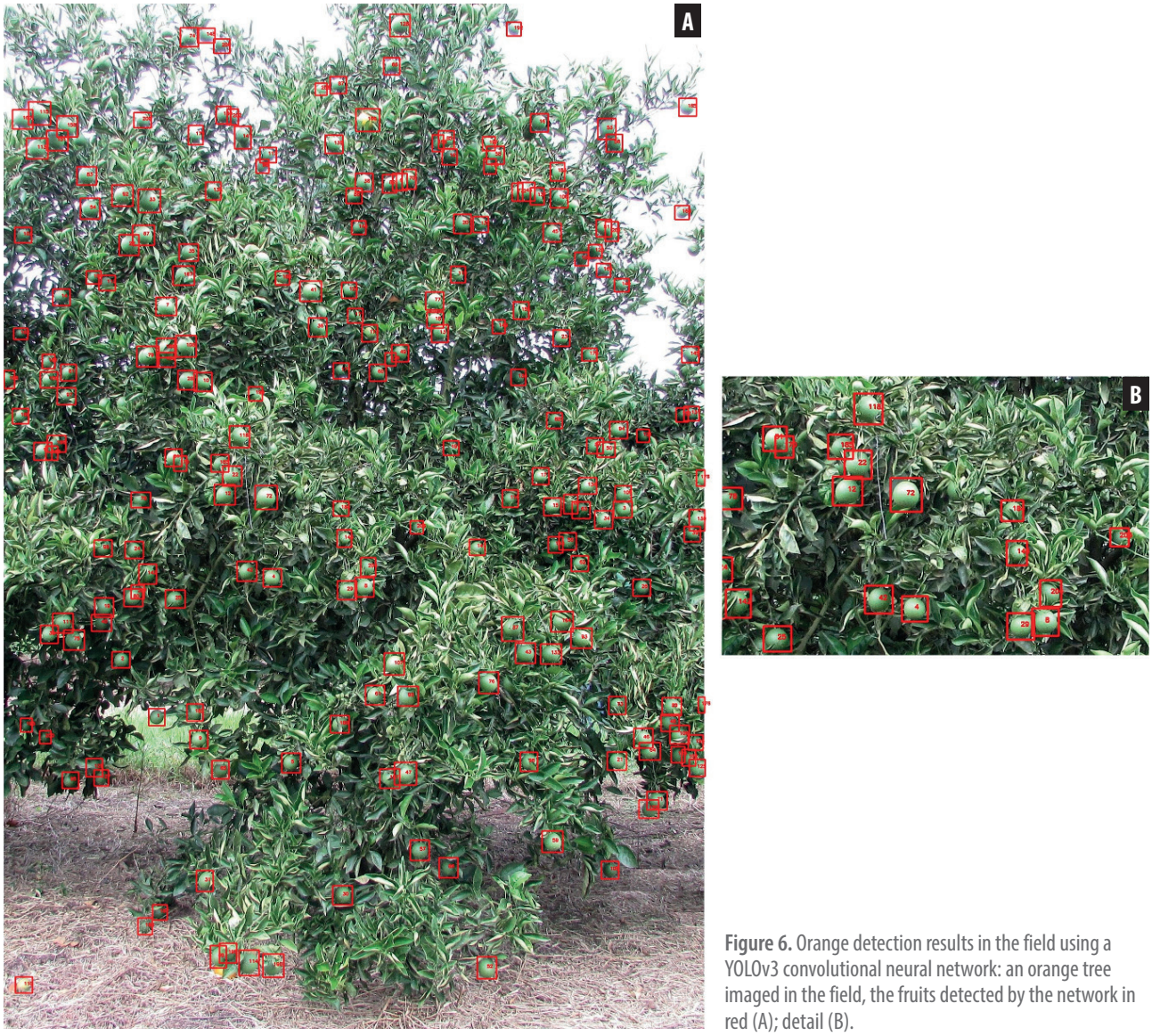
Automatic fruit detection is an enabling component for many agricultural applications. It can help estimate production, which is useful in logistical planning and in negotiations between rural producers and buyers. If detection is combined with precise spatial location, new applications can be developed in precision agriculture, assisting in the proper management of spatial crop variability. Fruit detection can also be a preliminary step in monitoring disease and nutritional deficiencies (see item "Identification of plant diseases"), restricting the areas in the images that should be inspected for symptoms. Given the decline in the agricultural workforce, fruit detection is also a technology that enables automated spraying and harvesting systems (Duckett et al., 2018; Xiong et al., 2020).

As discussed earlier, there are several factors that hinder the detection process, from occlusion by leaves and branches to camera focus and lighting issues (Figure 3). In some crops, the fruits also have various shapes, compactness and orientation, such as viticulture (Santos et al., 2020). Despite some success with other machine learning techniques (Gongal et al., 2015), fruit detection has recently gained traction with the improvements in convolutional neural networks (Sa et al., 2016; Bargoti; Underwood, 2017; Kamilaris; Prenafeta-Boldú, 2018).

Camargo Neto et al. (2019) produced a dataset with 3,066 images of oranges collected in the field, from different devices, such as cameras and smartphones. Most of the images were provided by the Crop Estimation Program (PES) of the Citriculture Defense Fund (Fundecitrus). The fruits, from different varieties of orange, had different levels of maturation, with a predominance of green fruits (Figure 3). From these images, a subset of 2036 observations was used in the training of a YOLOv3 neural network (Redmon et al., 2016; Redmon; Farhadi, 2018). The authors evaluated the network trained in the 1030 remaining images and verified the correct detection of more than 90% of the fruits, with an accuracy also above 90%, that is, less than 10% of the detections produced were false positives. Figure 6 shows an example of fruit detection in an orange tree image taken in the field.

Santos et al. (2020) showed that for the grapes in viticulture that present high variation in shape, color, size and compactness, bunches can be detected and segmented using architectures such as Mask-RNN and YOLO. The authors produced a new annotation tool that can speed up the process of associating pixels to fruits, discriminating exactly which pixels belong to which bunches. The generated dataset, named WGISD (Embrapa Wine Grape Segmentation Dataset) and publicly available[5], contains 4,432 bunches in 300 images, covering five wine varieties. The authors evaluated three different neural network architectures, YOLOv2 (Redmon; Farhadi, 2017), YOLOv3 (Redmon; Farhadi, 2018) and Mask-RCNN (He et al., 2017), the latter responsible for the most promising results. In a test base composed of 837 bunches, the network identified 87% of the bunches, with precision of 90.7%. Examples of the produced detections are shown in Figure 1 (B).

[5] Available at: https://doi.org/10.5281/zenodo.3361736.



Photos: João Camargo Neto

Figure 6. Orange detection results in the field using a YOLOv3 convolutional neural network: an orange tree imaged in the field, the fruits detected by the network in red (A); detail (B).

However, a complete fruit counting application needs a methodology that can integrate the detections reported in several images, so that fruits seen in more than one image are not counted multiple times. In other words, fruits (and objects of interest in general) observed in several images must be associated with each other. This data association task can be performed by integrating pattern recognition with geometric computer vision, as shown as follows.

Three-dimensional mapping and reconstruction

One of the greatest contributions of geometric computer vision was developing algorithms capable of recovering three-dimensional information from a set of images of the same scene. As results from decades of research in areas such as projective geometry and continuous optimization, these algorithms can transform even a simple webcam into a powerful 3-D scanner. Perhaps even more importantly, they allow a mobile agent, such as a UAV, not only to map the three-dimensional structure of the environment,

but also to determine its precise location (Figure 2), paving the way for autonomous agents that can navigate and interact with its surroundings (Stachniss et al., 2016).

Images must be obtained from different positions, by multiple cameras or by a single camera moving through the scene. This is the meaning of the term structure from motion (SfM), used in computer vision to define the problem of recovering the three-dimensional structure of a scene and the position of the camera from a set of images. Figure 7 illustrates the process of projecting a point in the scene as the camera is moved to three different positions. If we can determine correspondences between points in different images, it is possible to determine, with the help of projective geometry techniques, the position of the camera at the time each image was captured, more precisely the location of its projection centers, represented in the Figure 7 for points C_1, C_2 and C_3[6]. Once the location of the projection centers has been determined, it is then possible to estimate the position of the point in three-dimensional space based on its projections on the images (the points x1, x2, and x3 in Figure 7), a process known as *triangularization*. A detailed description of the entire process can be seen in Hartley and Zisserman (2003). The determination of image correspondence is also obtained automatically, using algorithms specialized in finding visually salient points (the points x1, x2, and x3) and, by comparing the pixels in their neighborhoods, associating different image points (Lowe, 2004; Detone et al., 2018).

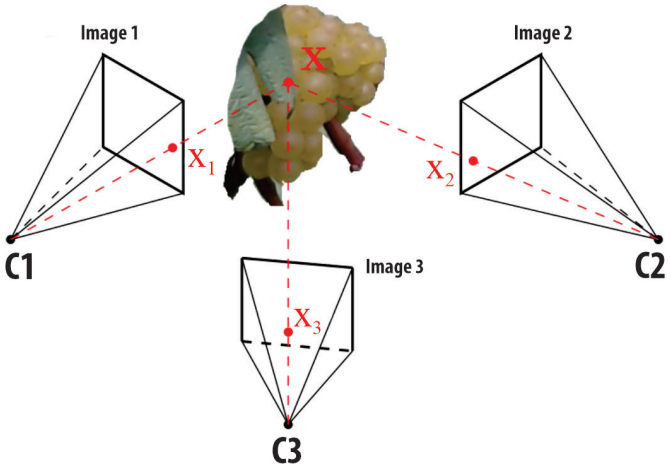


Figure 7. Structure from motion. An X point on a scene surface is projected onto the image plane at different positions as the camera is moved to positions C1,C2 and C3.

Illustration: Thiago Teixeira Santos

Santos et al. (2017) showed that an SfM system using a simple webcam can build accurate three-dimensional plant models in the field. Figure 8 shows an example, for a Chardonnay vine. As we will see in the next section, these three-dimensional models can be used to estimate 3-D attributes, such as

Figure 8. Three-dimensional reconstruction with SfM for a Chardonnay vine in the field: the red prisms indicate the camera (a commercial webcam) position and orientation, when each image was captured (A); same 3-D model observed from another angle (B).

Illustration: Thiago Teixeira Santos.

[6] Additional information, obtained by calibration methods, is needed to determine the correct scale, that is, the distance in a known unit such as meters or millimeters.

fruit volume and position. The 3-D system used and developed at Embrapa Digital Agriculture, named 3dmcap, is freely available[7] for non-commercial use.

The use of three-dimensional information in agriculture is expected to intensify in the coming years, not only through the use of the SfM technique (already commercially used by 3-D mapping services with UAVs), but also by the falling costs of stereo cameras, which provide depth information in the image, and by LIDAR sensors. Recent examples are the use of stereo cameras in vineyard phenotyping (Milella et al., 2019) and the detection of apples using LIDAR (Gené-Mola et al., 2020).

Figure 8 shows the three-dimensional reconstruction with SfM for a Chardonnay vine in the field: red prisms indicate the position and camera orientation (a commercial webcam, at the time of capture of each image (A); same 3-D model from another angle (B).

Combination of structure and recognition

If the SfM retrieves the three-dimensional structure from the scene and the imaging itself (the camera position(s) during the capture time), and the recognition identifies objects of interest in the scene, such as symptoms, fruits, plants or animals, the combination of the two pieces of information allows a broad assessment of the observed environment.

One of the uses of this combination is fruit mapping: the 3-D information combined with the detection of fruit in each image allows the spatial position of each fruit to be determined and that the same fruit is not counted more than once when it appears in multiple images.

Santos et al. (2020) used SfM to obtain a three-dimensional reconstruction of a row of vines in the field, based on the frames of a video sequence produced by a camera embedded in a service vehicle. A neural network was used to detect bunches of grapes in each image. By projecting the 3-D points of the scene onto the images, it was possible to associate detections with positions in the three-dimensional space and, therefore, determine the consistency between the bunches observed in one video frame and the bunches seen in the following frames[8] (Figure 9).

The joint use of 3-D models obtained by SfM and convolutional networks for fruit

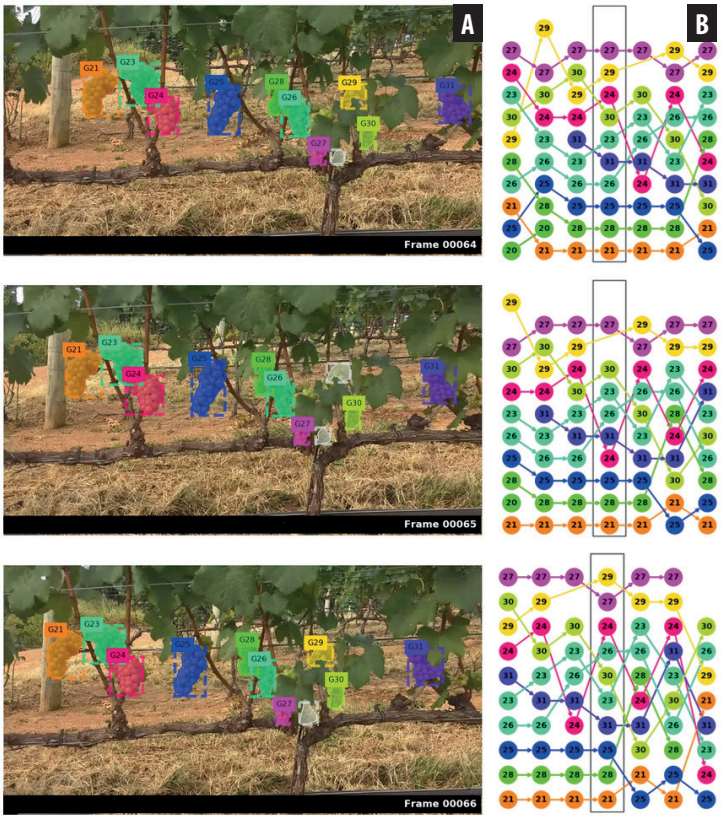


Figure 9. Tracking grape bunches in a video sequence obtained in the field: video frames were extracted and submitted to fruit detection by neural networks (A); the nodes represent bunches of grapes, in the order in which they were found by the neural network (each column of nodes represents a frame of the video sequence). The arrows inform the association between nodes from one frame to another, performed using 3-D information obtained by SfM (B).

Illustration: Thiago Teixeira Santos.

[7] Available at: https://github.com/thsant/3dmcap

[8] 8 A video demonstrating the tracking of grape bunches is available at: https://www.youtube.com/watch?v=1Hji3GS4mm4

detection and counting was also explored by Liu et al. (2019) in mango orchards and by Häni et al. (2020) in apple orchards.

Attributes of great interest in agronomic applications can be extracted from three-dimensional information. Santos et al. (2017) used a machine learning algorithm to identify which regions of the three-dimensional vine models corresponded to the bunches of grapes, as shown in Figure 10 (A). The volume of bunches was then estimated based on these regions. Fruit volume has a strong correlation with its weight, as can be seen in Figure 10 (B). These computer vision-based systems can provide a non-invasive and non-destructive methodology for estimating fruit weight, without having to remove them from the plant. Such technology can be used to assess growth throughout the crop cycle, without the need to remove (collect) samples.

Figure 10 shows the estimation of fruit weight based on volume in three-dimensional models.

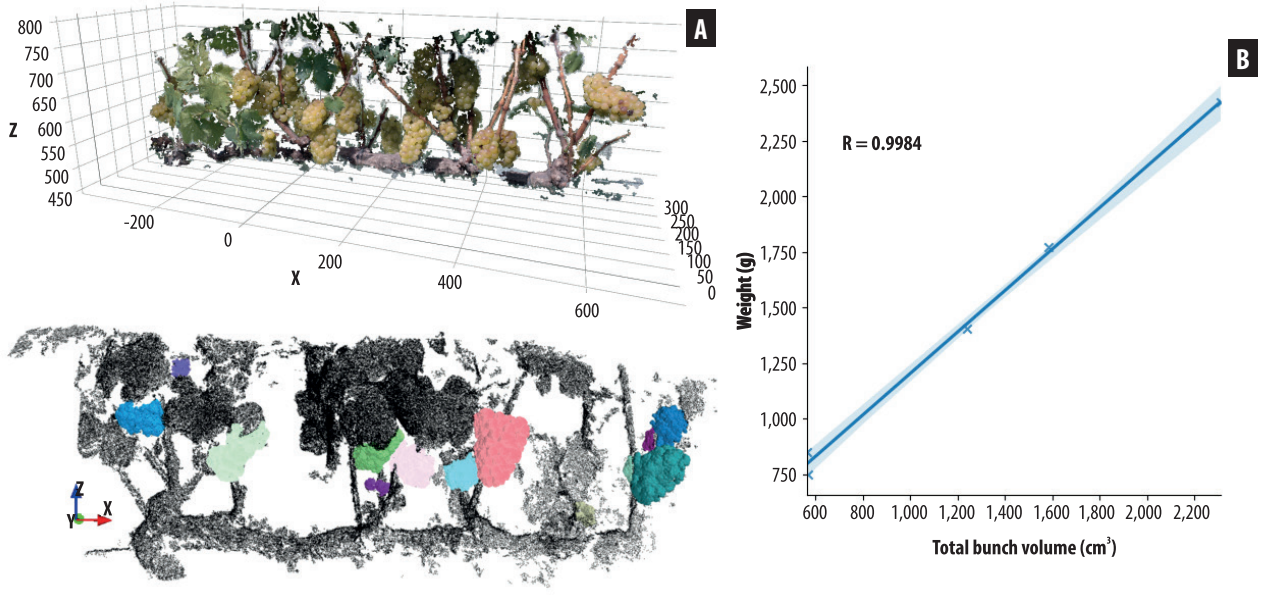


Figure 10. Estimation of fruit weight from the volume in three-dimensional models: grape bunches are identified (in colors) and separated from the rest of the plant (in black) (A); coefficient of determination between the estimated volume and the total weight of fruits in five different vines (B).

Source: Adapted from Santos et al. (2017).

Performance and intervention: field robotics

The combination of SfM and recognition is precisely one of the enabling technologies for one of the most challenging and impactful applications in agricultural automation: field robotics. Take, for example, a major challenge in agricultural robotics: automated fruit harvesting. While crops such as grains, sugarcane and coffee have their own machinery for automated harvesting, the same does not apply to horticulture and fruit cultivation – especially for the latter – due to the existing complexity in the structure of the orchards. Fruit harvesting depends on manual harvesting, which is unsettling considering the decreasing availability of labor in the field (Roser, 2013).

Automatic harvesting systems require two components of computer vision: the perceptual, for identifying fruits and obstacles, and the geometric, for the automatic positioning of the robot and its

handlers. Several research groups have applied these two components in the development of automated harvesting systems. Taking apple farming as an example, Silwal et al. (2017) developed a robotic apple harvesting system, evaluated in a commercial orchard. Their computer vision system was accurate, taking an average of 1.5 s to locate each fruit. The system was successful in harvesting 85% of the fruits, with an average time of 6 s per fruit. In addition to pomiculture, other crops have been investigated for implementing robotic harvesting, such as peppers (Bac et al., 2017), lettuce (Birrell et al., 2020), strawberry (Xiong et al., 2020), kiwi fruit (Williams et al., 2020), among others.

Final considerations

Computer vision has enormous potential for application in the area of digital agriculture. Several products and services based on computer vision components are expected to reach producers in the coming years. However, many challenges still depend on research and development endeavors.

A major bottleneck is the need for large databases to train neural networks for perceptual tasks. Research in the area of semi-supervised and unsupervised learning is currently being conducted by the computer vision community. The idea is to be able to learn patterns of interest with few examples and obtain systems with good accuracy in order to detect patterns such as fruits, symptoms and animals.

In robotics, the challenge continues to be developing robust systems that are capable of autonomously operating in the field for long periods, but which are safe for people and animals circulating in the field. These systems need to map the environment quickly, respond promptly, accurately find the objects to be monitored, and carry out the interventions for which they are designed. Despite the immense challenges, the computer vision and robotics communities have made great advances in recent years, which will soon be reflected in various agricultural applications, from monitoring to performance.

Finally, the authors emphasize that the results in fruit detection were financed by the Embrapa SEG 11.14.09.001.05.04 and FAPESP 2017/19282-7 projects. The results related to disease detection were financed by projects FAPESP 2013/06884-8 and Embrapa SEG 02.14.09.001.00.00. The results related to animal detection experiments were funded by FAPESP 2018/12845-9 project. The images for citriculture research were provided by PES/Fundecitrus. In addition, the GPUs used to train the neural networks were donated by NVIDIA Corporation.

References

BAC, C. W.; HEMMING, J.; TUIJL, B. A. J. van; BARTH, R.; WAIS, E.; HENTEN, E. J. van. Performance evaluation of a harvesting robot for sweet pepper. **Journal of Field Robotics**, v. 34, n. 6, p. 1123-1139, Sept. 2017. DOI: 10.1002/rob.21709.

BARBEDO, J. G. A.; KOENIGKAN, L. V.; SANTOS, T. T.; SANTOS, P. M. A study on the detection of cattle in UAV images using deep learning. **Sensors**, v. 19, n. 24, article number 5436, Dec. 2019. DOI: 10.3390/s19245436.

BARBEDO, J. G. A. Plant disease identification from individual lesions and spots using deep learning. **Biosystems Engineering**, v. 180, p. 96-107, Apr. 2019. DOI: 10.1016/j.biosystemseng.2019.02.002.

BARBEDO, J. G. Factors influencing the use of deep learning for plant disease recognition. **Biosystems Engineering**, v. 172, p. 84-91, Aug. 2018. DOI: 10.1016/j.biosystemseng.2018.05.013.

BARGOTI, S.; UNDERWOOD, J. Deep fruit detection in orchards. In: IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION, 2017, Singapore. **Proceedings**... Singaropre: IEEE, 2017. p. 3626-3633. DOI: 10.1109/ICRA.2017.7989417.

BIRRELL, S.; HUGHES, J.; CAI, J. Y.; IIDA, F. A field-tested robotic harvesting system for iceberg lettuce. **Journal of Field Robotics**, v. 37, n. 2, p. 225-245, Mar. 2020. DOI: 10.1002/rob.21888.

CAMARGO NETO, J.; TERNES, S.; SOUZA, K. X. S. de; YANO, I. H.; QUEIROS, L. R. Uso de redes neurais convolucionais para detecção de laranjas no campo. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 12., 2019, Indaiatuba. **Anais [**...] Ponta Grossa: SBIAGRO, 2019. p. 312-321. Organizadores: Maria Fernanda Moura, Jayme Garcia Arnal Barbedo, Alaine Margarete Guimarães, Valter Castelhano de Oliveira. SBIAgro 2019. Available at: https://www.alice.cnptia.embrapa.br/ alice/bitstream/doc/1125722/1/PC-Redes-neurais-SBIAGRO-2019.pdf. Accessed on : 10 May 2020.

COMBA, L.; BIGLIA, A.; AIMONINO, D. R.; GAY, P. Unsupervised detection of vineyards by 3D point-cloud UAV photogrammetry for precision agriculture. **Computers and Electronics in Agriculture**, v. 155, p. 84-95, Dec 2018. DOI: 10.1016/j.compag.2018.10.005.

DETONE, D.; MALISIEWICZ, T.; RABINOVICH, A. SuperPoint: self-supervised interest point detection and description. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS, 2018, Salt Lake City. **Proceedings**... IEEE, 2018. p. 337-349. DOI: 10.1109/CVPRW.2018.00060.

DUCKETT, T.; PEARSON, S.; BLACKMORE, S.; GRIEVE, B. Agricultural robotics: the future of robotic agriculture. **UK-RAS Network**, June 2018. DOI: 10.31256/WP2018.2.

FERENTINOS, K. P. Deep learning models for plant disease detection and diagnosis. **Computers and Electronics in Agriculture**, v. 145, p. 311-318, Feb. 2018. DOI: 10.1016/j.compag.2018.01.009.

FORSMOO, J.; ANDERSON, K.; MACLEOD, C. J. A.; WILKINSON, M. E.; BRAZIER, R. Drone-based structure-from-motion photogrammetry captures grassland sward height variability. **Journal of Applied Ecology**, v. 55, n. 6, p. 2587-2599, Nov. 2018. DOI: 10.1111/1365-2664.13148.

GENÉ-MOLA, J.; GREGORIO, E.; CHEEIN, F. A.; GUEVARA, J.; SANZ-CORTIELLA, R.; ESCOLÀ, A.; ROSELL-POLO, J. R. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. **Computers and Electronics in Agriculture**, v. 168, 105121, Jan. 2020. DOI: 10.1016/j.compag.2019.105121.

GONGAL, A.; AMATYA, S.; KARKEE, M.; ZHANG, Q.; LEWIS, K. Sensors and systems for fruit detection and localization: a review. **Computers and Electronics in Agriculture**, v. 116, p. 8-19, Aug. 2015. DOI: 10.1016/j.compag.2015.05.021.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. Cambridge; London: MIT Press, 2016.

HÄNI, N.; ROY, P.; ISLER, V. A comparative study of fruit detection and counting methods for yield mapping in apple orchards. **Journal of Field Robotics**, v. 37, n. 2, p. 263-282, Mar. 2020.

HARTLEY, R.; ZISSERMAN, A. **Multiple view geometry in computer vision**. 2nd ed. New York: Cambridge University Press, 2003. DOI: 10.1017/CBO9780511811685.

HE, K.; GKIOXARI, G.; DOLLÁR, P.; GIRSHICK, R. Mask R-CNN. In: IEEE INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2017, Venice. **Proceedings**... IEEE, 2017. p. 2980-2988. DOI: 10.1109/ICCV.2017.322.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: a survey. **Computers and Electronics in Agriculture**, v. 147, p. 70-90, Apr 2018. DOI: 10.1016/j.compag.2018.02.016.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, May 2015. DOI: 10.1038/nature14539.

LIU, X.; CHEN, S. W.; LIU, C.; SHIVAKUMAR, S. S.; DAS, J.; TAYLOR, C. J.; UNDERWOOD, J.; KUMAR, V. Monocular camera based fruit counting and mapping with semantic data association. **IEEE Robotics and Automation Letters**, v. 4, n. 3, p. 2296-2303, 2019. DOI: 10.1109/LRA.2019.2901987.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91-110, Nov. 2004. DOI: 10.1023/B:VISI.0000029664.99615.94.

MILELLA, A.; MARANI, R.; PETITTI, A.; REINA, G. In-field high throughput grapevine phenotyping with a consumer-grade depth camera. **Computers and Electronics in Agriculture**, v. 156, p. 293-306, Jan 2019. DOI: 10.1016/j.compag.2018.11.026.

REDMON, J.; DIVVALA, S.; GIRSHICK, R.; FARHADI, A. You only look once: unified, real-time object detection. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas. **Proceedings**... IEEE, 2016. p. 779-788. DOI: 10.1109/CVPR.2016.91.

REDMON, J.; FARHADI, A. YOLO v3: an incremental improvement [DB]. **arXiv preprint arXiv:1612.08242**, 2018.

REDMON, J.; FARHADI, A. YOLO9000: better, faster, stronger. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2017, Honolulu. **Procedings**... IEEE, 2017. p. 6517-6525. DOI: 10.1109/CVPR.2017.690.

ROSER, M. **Employment in agriculture**. 2013. Available at: https://ourworldindata. org/employment-in-agriculture. Accessed on: 19 May 2020.

SA, I.; GE, Z.; DAYOUB, F.; UPCROFT, B.; PEREZ, T.; MCCOOL, C. DeepFruits: a fruit detection system using deep neural networks. **Sensors**, v. 16, n. 8, p. 1222, Aug. 2016. DOI: 10.3390/s16081222.

SANTOS, T. T.; BASSOI, L. H.; OLDONI, H.; MARTINS, R. L. Automatic grape bunch detection in vineyards based on affordable 3D phenotyping using a consumer webcam. In: CONGRESSO BRASILEIRO DE AGROINFORMÁTICA, 11., 2017, Campinas. **Ciência de**

dados na era da agricultura digital: anais. Campinas: Editora da Unicamp: Embrapa Informática Agropecuária, 2017. p. 89-98. SBIAgro 2017. Available at: https://www.alice.cnptia.embrapa.br/alice/bitstream/doc/1083291/1/AutomatcgrapeSBIAgro.pdf. Accessed on: 16 Oct . 2020.

SANTOS, T. T.; SOUZA, L. L. de; SANTOS, A. A. dos; AVILA, S. Grape detection, segmentation, and tracking using deep neural networks and three-dimensional association. **Computers and Electronics in Agriculture**, v. 170, article 105247, p. 1-17, Mar. 2020. DOI: 10.1016/j.compag.2020.105247.

SILWAL, A.; DAVIDSON, J. R.; KARKEE, M.; MO, C.; ZHANG, Q.; LEWIS, K. Design, integration,and field evaluation of a robotic apple harvester. **Journal of Field Robotics**, v. 34, n. 6, p. 1140–1159, sept. 2017. DOI: 10.1002/rob.21715.

STACHNISS, C.; LEONARD, J. J.; THRUN, S. Simultaneous localization and mapping. In: SICILIANO, B.; KHATIB, O. (ed.). **Springer handbook of robotics**. Cham: Springer International Publishing, 2016. p. 1153-1176. DOI: 10.1007/978-3-319-32552-1_46.

SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2015, Boston. **Proceedings**... IEEE, 2015. p. 1-9. DOI: 10.1109/CVPR.2015.7298594.

WESTOBY, M.; BRASINGTON, J.; GLASSERA, N. F.; HAMBREY, M. J.; REYNOLDS, J. M. 'Structure-from-Motion' photogrammetry: a low-cost, effective tool for geoscience applications. **Geomorphology**, v. 179, p. 300-314, Dec 2012. DOI: 10.1016/j.geomorph.2012.08.021.

WILLIAMS, H.; TING, C.; NEJATI, M.; JONES, M. H.; PENHALL, N.; LIM, J.; SEABRIGHT, M.; BELL, J.; AHN, H. S.; SCARFE, A.; DUKE, M.; MACDONALD, B. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. **Journal of Field Robotics**, v. 37, n. 2, p. 187-201, Mar. 2020. DOI: 10.1002/rob.21890.

XIONG, Y.; GE, Y.; GRIMSTAD, L.; FROM, P. J. An autonomous strawberry-harvesting robot: design, development, integration, and field evaluation. **Journal of Field Robotics**, v. 37, n. 2, p. 202-224, Mar 2020. DOI: 10.1002/rob.21889.

ZOPH, B.; VASUDEVAN, V.; SHLENS, J.; LE, Q. V. Learning transferable architectures for scalable image recognition. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2018, Salt Lake City. **Proceedings**... IEEE, 2018. p. 8697-8710. DOI: 10.1109/CVPR.2018.00907.