Real-time Fruit Detection Using Deep Neural Networks

Barna Keresztes[1,2], Florent Abdelghafour[1,2], Dimby Randriamanga[1,2], Jean-Pierre Da Costa[1,2], Christian Germain[1,2]

[1]Univ. Bordeaux, IMS UMR 5218, F-33405 Talence, France

[2]CNRS, IMS UMR 5218, F-33405 Talence, France

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. Proximal imaging using tractor-mounted cameras is a simple and cost-effective method to acquire large quantities of data in orchards and vineyards. It can be used for the monitoring of vegetation and for the management of field operations such as the guidance of smart spraying systems for instance. One of the most prolific research subjects in arboriculture is fruit detection during the growing season. Estimations of fruit-load can be used for early yield assessments and for the monitoring of harvest and thinning. In addition, the visual aspects of fruits enable to appraise their growth and ripening status. This paper proposes a new approach for real-time fruit detection, combining a fast geometrical pre-processing whose output feeds a deep neural network (DNN) classifier. The first step is a radial Hough-like operator, which aims at identifying quickly the regions of interest, restricting the use of the DNNs to the most probably genuine candidates. The proposed method is generic enough to be applied on most near-spherical fruits. It was tested in two contexts: grapes and apples, with different varieties and phenological stages. In both cases the proposed method provided promising results. Correlation coefficients with manual counting and real harvest loads are up to 0.96 for grapes and up to 0.85 for apples.

Keywords. Proximal sensing, image analysis, fruit detection, deep learning, radial Hough transform

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 14th International Conference on Precision Agriculture. Keresztes, B., Abdelghafour, F., Randriamanga, D., Da Costa, J.P. and Germain, C. *Real-time Fruit Detection Using Deep Neural Networks*. In: Proceedings of the 14th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Introduction

Since the beginning of the 1980's, automated detection and counting of fruits in vineyards and orchards has been a major concern and a very prolific research subject in computer vision. The implementation of computer vision algorithms enables applications such as the quality grading of fruits, yield mappings, robotized picking and disease detection (Gemtos *et al.*, 2013). Jimenez et al. (2000) summarizes the most noticeable approaches for fruit detection assisted by computer vision. Some fruits, like oranges or tomatoes are easy to detect at maturity, since their color differs a lot from the foliage. In other cases, template matching algorithms, texture analysis or shape identification (*e.g.* circular Hough transform) have been used. The use of special cameras (stereo cameras, multispectral imaging, laser ranging) is also a well-established solution to improve detection efficiency (Gao *et al.*, 2010). However, fruits present a variety of shapes, sizes and colors. In natural environments, fruits can be partially occulted and are exposed to varying illuminations. In these conditions automated fruit detection within in-field images is still an ongoing challenge. In recent years deep learning accomplished major improvements such as the development of convolutional neural networks (CNN) which is a promising non parametric alternative to standard image analysis methods that proves very efficient for image recognition. Dedicated networks are able to identify specific objects in an image (regional CNNs) (Girschlik et al, 2014) or to segment an image (SegNets) (Badrinarayanan et al, 2017). Deep learning is a very robust approach that can be suitable for agricultural applications such as deep fruit detection in orchards (Bargoti and Underwood, 2017).

This paper presents a new method for fruit detection and counting intended for images acquired by proximal sensing. It combines a parametric pre-processing step, based on the circular Hough transform meant to identify the locations of fruit candidates and a deep neural network acting as a decision process. This combination enables to reduce the number of samples to be processed compared to an R-CNN, while keeping the robustness of the deep neural networks, so that the presented algorithm is able to identify any kind of circular fruit. In addition, a convenient method for the creation of large training database is proposed.

The following sections present the acquisition device and the different algorithms designed for the two-step fruit detection. The third section presents the results of the proposed method both for grapes and apples.

Materials and methods

Image acquisition

The aim of this project is to create a system able to count the visible fruits in orchards using a computer vision system that can be easily integrated in common farm infrastructures and equipment. For this purpose, the device needs to be compact and easy to operate during the main farm works on conventional machines (quads, tractors or high clearance tractors). The device should work autonomously at the regular work rates (around 8km/h), *i.e.* acquisitions covering up to three pictures per second. The camera is oriented perpendicularly to the trellising plane. Image processing has to be implemented on an on-board computer, providing direct georeferencing of the fruit headcount.

a. b. c.

Figure 1. a. The imaging system; b. the camera mounted on a tractor and c. on a quad.

The imaging system is built around a Basler Ace industrial camera with a resolution of 5 megapixels. The lens has a 55° horizontal field of view, capturing a $0.8m^2$ area in the vineyards and $2m^2$ area in the orchards, which is sufficient to observe the fruit-zone in each case.

The camera is controlled by an on-board computer which is also used to store and process the acquired images. The computer is built around a low consumption 4-core ARM chip, and does not contain any mechanical component, to reduce the risk of failure caused by vibrations. In order to run the deep neural networks more effectively, the computer is equipped with a dedicated Movidius AI processor.

To overcome the uncontrolled variability in illumination levels encountered in outdoor conditions, a high-power xenon flash-bulb is used. This flash removes any unwanted background on images, since the rows behind target receive much less light than the plants in the foreground. In addition the use of a short exposure time $(250 - 300\mu sec)$ ensures that images are not blurred by the movement of the camera. The system contains an ultrasonic telemeter to measure the distance between the camera and the plants. Distance information is used to detect missing plants and the end of rows.

Finally, the system is controlled by a GNSS unit connected to the computer. The position of the tractor is monitored, and the acquisition starts automatically as it enters the defined areas, there is then no requirement for any user interaction when operating measurements. The cruising speed is used to determine automatically the acquisition rate. Each picture is georeferenced, so that results can be processed further in a GIS software. Some examples of images acquired by the system are presented in figure 4.

Image analysis

The proposed method is a two-step process. The first step is a detection of potential fruits in the canopy; the locations of the detected candidates are passed through the second step which is a binary classification of fruit candidates that determines if candidates are indeed a fruit or an error. The following subsections describe the two steps.

Detection

This section presents a fast detection method for the selection of fruit candidates. It is an entirely shape-based method that retrieves circular objects regardless of their color or illumination. This algorithm is then able to detect any near-spherical fruit even when the color of fruits matches the color of the leaves, while being robust to variations in luminosity.

It can be assumed that the fruits are the only spherical objects in the image. Considering the very short distance between the target and the acquisition device, the illumination power of the embedded flash occults natural sunlight. The directions of the main light source and of the focal plane are then the same.

In these conditions, images can be described using the *Lambertian reflection model* (Oren and Nayar, 94). On a point P of the surface of a sphere:

$$I_D = I_L C cos\alpha \tag{1}$$

where I_D is the reflected light, I_L is the incident light, C is the color reflexion coefficient of the surface and α is the angle between the normal vector **N** to the surface and the incident light vector. Assuming that the light source is close to the camera *i.e.* the image plane is orthogonal to the light direction:

$$\alpha = arcsin(\mathrm{d/r}) \tag{2}$$

where r is the radius of the fruit and d is the distance of the pixel from point P to the line passing though the sphere center O in the light direction. On the image plane, *d* is the distance between the projections P' and O' of the point P and the sphere center O.

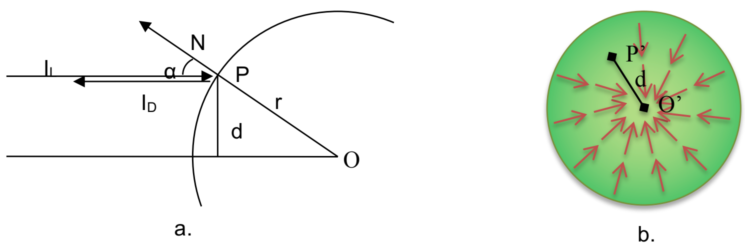

Figure 2. The reflection model of the spherical fruits: schematic side view (a), Schematic view of a fruit in the image plane (b). Gradients are represented as red arrows, converging towards the centre of the disk.

Putting together equations (1) and (2), we get:

$$I_D = I_L C \sqrt{1 - \frac{d^2}{r^2}} \tag{3}$$

The derivative of I_D according to d is then:

$$\frac{\partial I_D}{\partial d} = -\frac{I_L C}{r} \frac{d}{\sqrt{r^2 - d^2}} \tag{4}$$

which is negative for $d > 0$, which means that the light intensity diminishes with the distance from the fruit center projection O' and form circular isocontours around it. This also means that the image gradients are positive towards the center of the fruit, then all the gradient vectors point toward this point, as presented by the red arrows in figure 2.b.

In order to estimate the positions of the fruits, the radial Hough transform is defined where the accumulator space is formed by the gradient vectors of the image. If these vectors converge, they add up and create a peak in the accumulator. These peaks indicate the centers of the fruits.

a. b.

Figure 3. Radial Hough transform on a grape bunch (a); Radial Hough transform on an image of apples (b)

Figure 3 shows the result of radial Hough operator on the image of an apple tree (b) and a bunch of grapes (a).

The results show that the radial Hough transform gives reliable results for spherical fruits, regardless of their color, and development stage. Fruits are also correctly identified by their center even when they are partially occulted by another fruit, a leaf or a branch. It is also possible to discriminate individual fruits within a group of contiguous fruits. The proposed method can also work in other direct lighting conditions, even when the light source does not have the same direction as the camera (e.g. natural light, flash not close to the camera). In this case, the gradients converge toward the projection of the fruit center on the fruit surface, in the direction of light. The image gradients are computed using a two-dimensional Gaussian derivative function. This is a separable fast computing operator along the x and y axes. It also smooths the image, so that gradient directions are robust to noise. The smoothing degree of this function can be modified using the standard deviation σ of the Gaussian kernel. The radial Hough transform is a fast algorithm; as it required to be operated only with one pass through images. Both the computation of gradients and the radial Hough transform can be parallelized to achieve real-time processing.

Classification

The method presented above identifies regions in the image with convergent gradients. However, in some cases some circular patches on leaves or shiny junctions of branches present similar geometrical properties and are then falsely selected. The Hough radial transform is not robust to these marginal cases. In order to achieve a more reliable detection, an additional step based on a DNN classifier identifies the fruits and discards other objects.

Classifying images of fruits is not a trivial task, because there is a substantial variability regarding the illumination, geometric and colorimetric properties of fruits. At a single stage, fruits differ in shapes, sizes and colors. The differences are even more pronounced, when different varieties and different development stages are taken into account. The classifier should be robust to these intrinsic variations. A *deep neural network* can meet the requirements above, so it is used as classifier in this application.

The description of fairly simple objects like spherical fruits does not require the extraction of very complex features. With a small number of simple convolutional filters, it is possible to obtain features that describe and discriminate fruits within images. The LeNet architecture (LeCun et al, 1999) presented in figure 4 is sufficient to meet the requirements of the intended applications. This architecture has a low number of trainable parameters that makes its training simple and fast, compared to the recently developed, complex deep neural networks. Even with a low number of training samples overfitting can be avoided. The processing of a ROI is also very fast.

At the considered resolution, the image of a grape berry fits in the $[32 \times 32]$ pixel sized input layer of the LeNet architecture. For apples and other fruits, the maximal size of the fruit is considered as initial size for the ROI, which is resized to $[64 \times 64]$ pixel patches. The deep neural network is scaled accordingly.

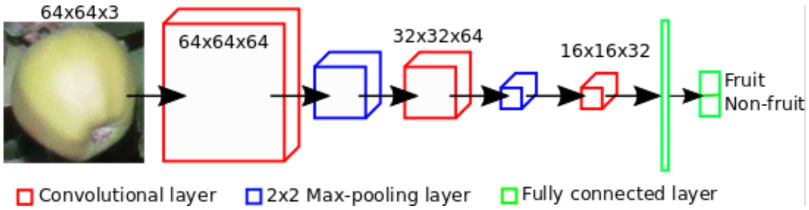

Figure 4. The deep neural network architecture

For the training a binary a database of 45.000 samples for grapes and 150.000 samples for apples, labeled with *fruit*/*not fruit* is used. The training samples contained fruit images from different varieties and development stages. The training was conducted over 50 epochs. The batch size was set to 20.000 samples for the grapes and 5.000 samples for the apples, to be able to fit into the GPU memory. 10% of the samples were randomly selected for the testing database. The training of the DNN was conducted offline, on a dedicated workstation.

The classification algorithm consists in locating the maxima of the accumulator space of the radial Hough transform, and the window around this point is passed through the deep neural network. This operation is repeated until the classifier outputs are all negatives. The sum of the positive responses gives the number of fruits in the image.

CPU's are usually not adapted to run DNN, it results in slow computations of convolutional layers. However, the use of a dedicated processor considerably improves the computational times. The Movidius AI processor allows real-time processing, with power consumption inferior to 1W.

Post-processing

Concerning grape berry detection, an additional post-processing step is necessary to eliminate some of the remaining false detections. Since grape berries form bunches, isolated detections of berries should then be discarded.

The use of an adjacency graph of the detection results allows assessing that two berries are connected, if the distance between their centers is less than twice the fruit diameter. A detection is considered isolated if it has no neighbors.

It has also to be considered that connected clusters of the graph may form a grape bunch. This is a quite simplistic approach, as it is very difficult to discriminate the real number and morphology of bunches when some of them are contiguous. However, it provides an approximation of grape distribution in images and the compactness of bunches.

Training

One of the main difficulties in the field of deep learning is the construction of a large training database, which is time-consuming and labor-intensive. Based on the number of trainable parameters of the network, a database of at least 20.000 samples is required.

To overcome this problem, an interactive method for the creation of the training database is proposed. The training images were processed using the radial Hough transform presented in section 2.2.1, and then a simple decision method was applied on the local maxima, based mostly on the color and luminance (Keresztes et al., 2012). The process continued until the numbers of positive and negative detections were equal, i.e. we get the same number of fruit and non-fruit detections.

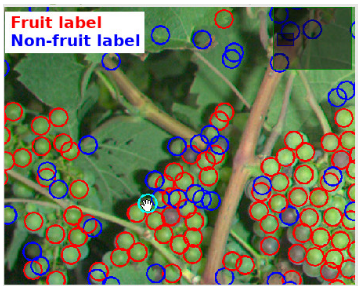

Figure 5. Interactive fruit labeling.

Then, the detected objects are overlaid on images with different colors according on the labels. The operator can manually change the labels by clicking on an object which is incorrectly labeled. When all the labels are correct, the positions and labels are added to the database.

Using this method, a sufficient training database can be built in a few hours by a single operator.

Results

Dataset

This paper presents the results of the proposed algorithm on grape and apple detection. The data acquisition for grapes was conducted in a 12 ha vineyard in the Bordeaux area. The plots are planted with red wine grape varieties *Merlot Noir*, *Cabernet Franc* and a white grape variety *Cabernet Sauvignon*. Four ha are dedicated for each variety. Images were acquired at four different phenological stages defined on the BBCH (Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie) scale (Lorentz *et al.*, 1995):

- *Cluster closening* in mid-July (*BBCH 79*)
- *Beginning veraison* in mid-august (*BBCH 83*)
- *Full veraison* (*BBCH 85*) *i.e.* three weeks before harvest
- *Right before harvest* in October (*BBCH 89*).

Concerning the cultivation systems, rows were defoliated on North or East sides before cluster closing, images were all acquired only on the defoliated sides, where the fruits are the most visible.

a. b.

Figure 6. Two example images from the dataset: a. grapevine a. and b. apple tree.

In apple orchards, the experiments were conducted in the Bergerac region, on two 1ha experimental plots. The plots contained three apple cultivars: *Gala*, *Golden Delicious* and *Pink Lady*. Images were acquired between May and September (fruit diameter over 30mm), with 3 acquisitions per month. The images were taken on both sides of the apple trees.

Fruit detection

The presented algorithm was tested on every image of the grapevine and apple tree dataset.

The image by image results show a very good detection rate for the grape berries (figure 7). The presented images were taken at *cluster closening* development stage, for Merlot, Cabernet Franc and Cabernet Sauvignon varieties respectively.

Figure 7. Results of the grape detection at the *cluster closening* stage (BBCH 79)

All the detected berries are encircled, the different detected bunches were coded with different colors.

The results of grape berry detection were also compared with the results of the automatic and manual berry counting on a small dataset of 10 images. As expected, the results show an excellent correlation, with $R^2 = 0.96$ (p-value < 10^{-6}).

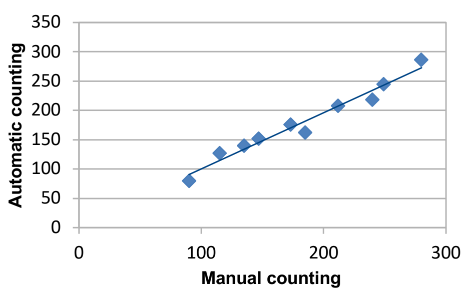

Figure 8. Correlation between automatic and manual grape counting at cluster closening

The apples were also well detected in images. Figure 9. presents the fruit detection results on two images presenting early stages of fruit development. This stage is the most difficult to process, as the fruits are smaller than the leaves and have the same color.

Figure 8. The results of apple detection

The apple detection algorithm was compared with manual counting on a large number of images from two stages corresponding to mean diameters *D* equal to 40 mm and 50 mm (see fig 9).

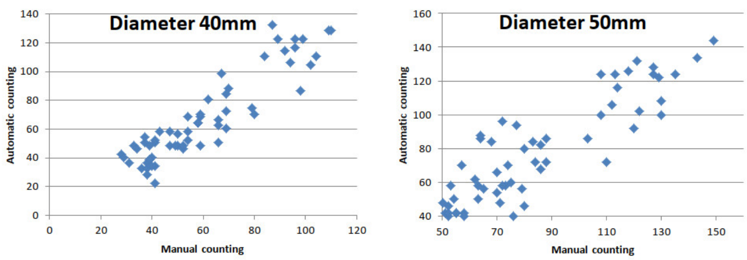

Figure 9. Comparison of automatic and manual apple counting on two development stages: mean diameter equal to 40 mm (left) and 50 mm (right)

The correlation between the manual and automatic counting is highly significant (p-value < 10^{-23}) with R^2 values of 0.85 and 0.83, respectively for $D = 40mm$ and $D = 50mm$. The correlation is slightly lower than for grape detection, as the DNN was trained to recognize only the fruits that are more than 50% visible.

Commented [JPDC1]: Barna, si tu pouvais intégrer des résultats de classification (cf. rapport de Dimby) ce serait parfait.

Conclusion and perspectives

In this paper we presented a new computer vision method for detecting spherical fruits. By combining a shape-based algorithm with a deep neural network, we created a detection method that is both fast and robust. The proposed method provides a reliable detection and counting of the visible fruits in the trellising plan.

By providing training data for different varieties and development stages, the algorithm is able to identify correctly the fruits in various conditions, even in early development stages when their color blends into the foliage. The R^2 correlation between the manual and automatic counting is over 0.83 even for early development stages.

An intelligent camera is used for image acquisition and embedded processing. This is a low-cost system which is able to acquire large quantities of photos and integrates easily in the existing infrastructure: tractors and other machines. The photos can be acquired during regular field work.

The controlled lighting, provided by a strong on-board flash eliminated the variability of the exterior lighting conditions, allowing in the meantime the development of a fast and reliable method to estimate the positions of the fruits.

The presented method does not provide a direct measurement of yields but rather a statistical estimation of crop load and it spatial repartition at a very local scale. The robustness of this method in terms of fruit detection has been proven for different development stages and varieties. However the eventual agronomic parameters resulting from the classification are based on inferences. Indeed it is still required to approximate unitary fruit weights and occultation coefficients *i.e.* an estimation of the number of fruits hidden by leaves, branches and other fruits. Such a coefficients mostly depend on cultivation systems, varieties and pedo-climatic contexts.

The information that can be extracted with this method provides an insight for many agronomic parameters that are still not included in management strategies. For instance the assessment of fruit distribution on plants using adjacency could be used as a measure of cluster density. It provides then a feedback assessment of the efficiency of mechanical thinning for apple trees or assessment of the ventilation and sun exposure of fruits in the canopy. It is now possible to obtain reliable estimations of agronomic parameters with extensive measurements and automated processing of enormous date set. Yet the development of precision farming still requires more combined field data to estimate the inferences coefficients and build models for Decision support tools.

Acknowledgement

This work was supported by the European Union's Horizon 2020 research and innovation program under grant agreement *no. 731884*

References

Badrinarayanan, V., Kendall, A. and Cipolla, R.: SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence Vol. 39, Issue 12, Dec. 1 2017

Bargoti, S. and Underwood, J. 2017. Deep fruit detection in orchards. 2017 IEEE International Conference on Robotics and Automation (ICRA) (2017): 3626-3633.

Gao H., Zhu F., Cai J. (2010) A Review of Non-destructive Detection for Fruit Quality. In: Li D., Zhao C. (eds) Computer and Computing Technologies in Agriculture III. CCTA 2009. IFIP Advances in Information and Communication Technology, vol 317. Springer, Berlin, Heidelberg

Gemtos, T., Fountas, S., Tagarakis, A., & Liakos, V. (2013). Precision agriculture application in fruit crops: Experience in handpicked fruits. Procedia Technology, 8, 324–332.

Girshlick, R., Donahue, J., Darrell, T. and Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation, CVPR 2014

Jimenez, A.R., Ceres, R. and Pons, J.L.: A Survey of Computer Vision Methods for Locating Fruit on Trees. Transaction of the ASAE, Vol. 43(6): 1911-1920, 2000

Keresztes, B., Lavialle, O., Grenier, G., Vaysse, P.: Apple yield estimation during the growth season using image analysis, International Conference on Agricultural Engineering, Valencia, July 2012

LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P.: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998

Lorenz, D.H., Eichhorn, K.W., Bleiholder, H., Klose, R., Meier, U., Weber, E. (1995). Growth Stages of the Grapevine: Phenological growth stages of the grapevine (Vitis vinifera L. ssp. vinifera)-Codes and descriptions according to the extended BBCH scale.

Oren, M. and Nayar, S.K.: Generalization of Lambert's reflectance model, Siggraph proceedings, 1994, pp.239-246

Sa , I., Ge, Z., Dayoub F., Upcroft, B., Perez, T. and McCool, C.: DeepFruits: A Fruit Detection System Using Deep Neural Networks. Sensors 16(8):1222, Basel 2016