# 3D ACQUISITION SYSTEM APPLIED TO AGRONOMIC SCENES

**B. Billiot, P. Gouton**

LE2I
University of Burgundy
Dijon, France

**F. Cointault**

Agrosup Dijon
Dijon, France

## ABSTRACT

To improve results in automatic wheat ear counting by proxy-detection for early yield prediction, we need depth information of the scene. In this paper, we describe our 3D acquisition system dedicated to reconstruction of agronomic scenes. This system is composed of a camera mounted on a linear displacement driven by a microcontroller. The linear displacement allows acquiring a set of images in different distances to the scene. This image stack is used to apply shape from focus technique which is a passive and monocular 3D reconstruction method. This technique consists in the application of a focus measure for every pixel in the stack. An approximation method is used to refine the results of focus measure. Depth information of each point on the scene is determined by the maximum of focus measure. With depth information of each pixel, we create the depth map of the scene. The parameters previously inaccessible with 2D images like a leaf volume can be retrieve thanks to depth information.

**Keywords:** 3D reconstruction, acquisition system, agronomic scenes, crop analysis

## INTRODUCTION

To enable a better decision making by the farmer in order to optimize the crop management, it is essential to provide a set of information on basic parameters of the crops. These information are numerous and the image processing is increasingly used for diseases detection (Al Hiary et al., 2011), weeds detection (Burks et al., 2000) or yield estimation (Germain et al., 1995). We will focus initially on assessing the yield of a wheat crop in automatic way. This yield is directly related to the number of ears per square meter for which the counting is currently done manually. To gather more accurate data than these obtained by remote sensing or not accessible by this technique, we developed a proxy-detection system. The advantages of this kind of system in comparison to remote sensing are: a best resolution, temporality and precision but also a simplified access and lower costs. With images obtained by the 2D acquisition system, we detect and count wheat ears by hybrid spaces (Cointault et al., 2008) or texture analysis (Cointault et al., 2008). To go further in the analysis of agronomic scene parameters, the contribution of three-dimensional information using 3D acquisition of scene is a relevant solution. Thus, the design of a 3D acquisition system is required in order to obtain new parameters related to crops. In literature, there are numerous 3D acquisition techniques. In first time, we use a Time-of-Flight scanner to reconstruct the scene. This kind of system provides good results but the prohibitive cost of such device as well as its incompatibility with easy use in the field does not make this one a possible solution. Then, we focused toward a common approach in computer vision: stereovision also called shape from stereo. The principle of this technique is detailed in (Barnard and Fischler, 1982) and it is based on multiview geometry using at least two cameras. The major drawback of this technique is its sensibility to occlusion phenomenon (a point on the scene viewed by a camera is not necessarily viewed by the other). These occlusions problems do not allow us to obtain good results due to the kind of scene where this phenomenon is very active (crops). Thus, we need a 3D reconstruction technique that frees oneself from occlusion problems. We choose to use shape from focus technique (Nayar and Nakagawa, 1994) which is currently applied in microscopic domain. This is a monocular technique that provides a depth map of the scene based on 2D image stack. Firstly, we detail the acquisition system created to obtain the image sequence. Then, we explain the different steps of image processing used to perform shape from focus method. Finally, we present and analyze the results obtained.

## ACQUISITION SYSTEM

The data set requiring the use of shape from focus technique is an image sequence of the same scene acquired by moving the focal plane. A schematic representation of acquisition process can be found in figure 1.

The rugged notebook is connected on camera and motor control board by USB cable to retrieve information and control the system. The motor control board is composed of a microcontroller and a power card. A stepper motor is coupled to a linear displacement with trapezoid screw. This motor is controlled by the motor control card and allows moving the camera incrementally and precisely.

The real acquisition system is presented in figure 2. The vision system is composed of a camera and a lens as well as two power LEDs to illuminate the scene. The vision system is centered on the desired field of view and it is mounted on the linear displacement. All of this is integrated into a global metal frame which is composed of two frames: upper and lower ones. When the system is not used, the upper frame could be integrated in lower frame to reduce the size of system.

The choice of step corresponding to the displacement between each acquisition will depend on the depth of field induced by vision system. Indeed, accuracy of shape from focus method and the number of images to acquire will depend on this depth of field. The depth of field phenomenon appears when we use a real optical system. It corresponds to the area of the scene where the points projected on the sensor will be sharp. All scene points located before and after this area will appear blurred in the image. The depth of field depends on four parameters of the vision system: camera/scene distance, aperture of the lens, focal length and diameter of circle of confusion. The focal length and aperture value will therefore depend on the kind of lens used. The diameter of the circle of confusion is equivalent to the pixel size of camera sensor.

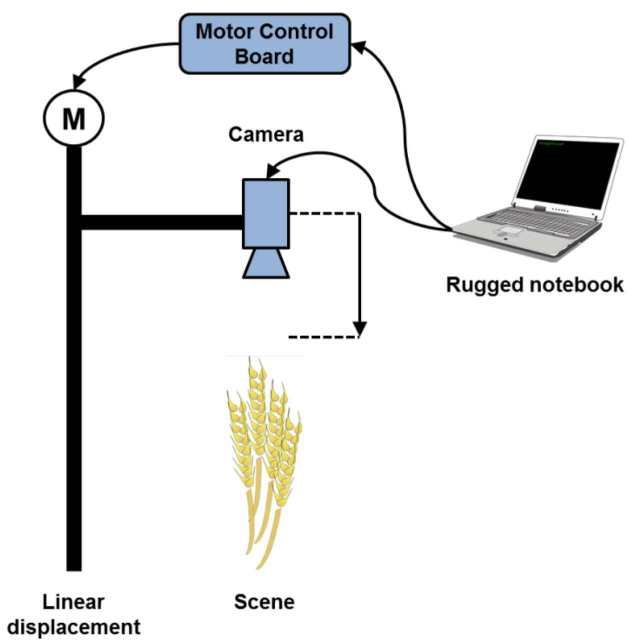


**Fig. 1. Acquisition process**

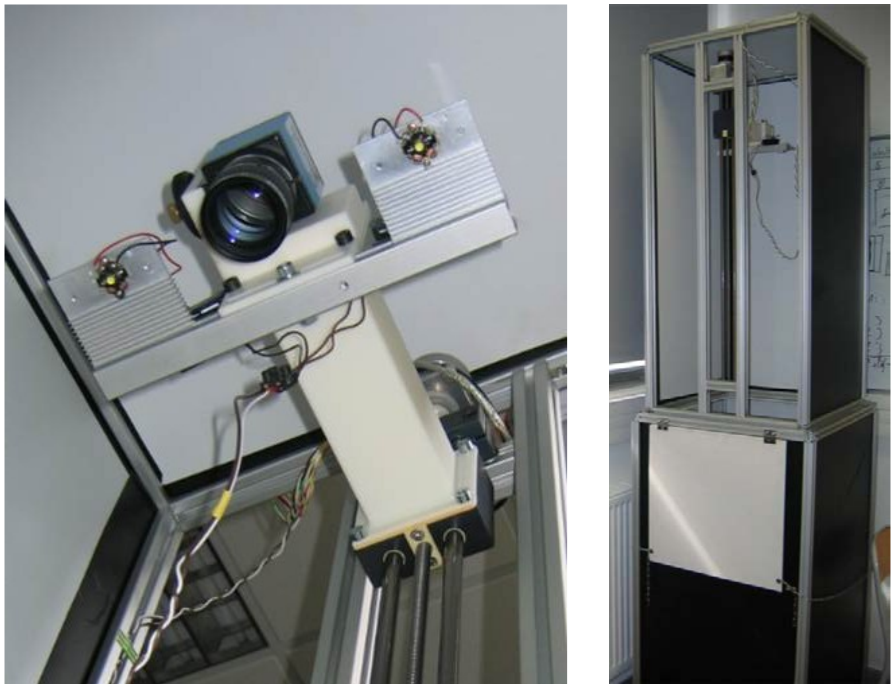

**Fig. 2. Acquisition system**

To resume the impact of these different optical parameters, the depth of field decreases when the focal length or aperture value increases. In the same way, it increases when the diameter of circle of confusion or the camera/scene distance increases. The depth of field is calculated as follows:

$$
\begin{aligned}
D_n &= \frac{D}{1 + C * A * \frac{D - F}{F^2}} \\
D_f &= \frac{D}{1 - C * A * \frac{D - F}{F^2}} \qquad (1) \\
DoF &= D_f - D_n
\end{aligned}
$$

Where *D* is camera/scene distance, *C* is diameter of circle of confusion, *F* is focal length and *A* is aperture. The diameter of circle of confusion can be defined like the pixel size of sensor.

Our vision system includes a CCD camera with a 1280 by 960 resolution and pixel size of 4.65 µm. We use a 50 mm lens with an aperture of 1.4 which allows the depth of field presented in table 1. In this table, we find different available fields of view and depths of field according to the kind of lens.

**Table 1. Fields of view and depths of field according to the kind of lens (values in millimeter and our values highlighted)**

| D | 25mm f1.4 | | | 35mm f1.4 | | | 50mm f1.4 | | |
|---|---|---|---|---|---|---|---|---|---|
| | width | height | DoF | width | height | DoF | width | height | DoF |
| 800 | 204.8 | 153.6 | 12.91 | 146.2 | 109.7 | 6.5 | 102.4 | 76.8 | 3.12 |
| 900 | 230.4 | 172.8 | 16.4 | 164.5 | 123.4 | 8.27 | 115.2 | 86.4 | 3.98 |
| 1000 | 256 | 192 | 20.31 | 182.8 | 137.1 | 10.25 | 128 | 96 | 4.94 |

## IMAGE PROCESSING

When the acquisition of image stack is done, different image processing techniques can be applied on the images to perform shape from focus method. There are four different steps: image registration, focus measure, approximation and depth map construction.

### Image registration

The acquisition process involves a magnification effect tied to the displacement of the vision system between each image acquisition. In microscopic domain, this magnification is not considered due to the very small displacement of the system. However, with our application we need to register this magnification. The image registration technique used is phase correlation method detailed in (Raj and Staunton, 2007). This method consist in calculate the cross power spectrum resulting from two images to estimate the translation between them. When the registration of all images of the sequence is done, we crop and scale all images according to the last image. Thus, we have a sequence of images where all images represent the same area of the scene.

### Focus measure and approximation

When the image registration is done, we must measure the sharpness of all pixels of image stack to find where pixels are most sharp. We can consider this measure like the following function:

$$f_i(x,y) = \max_i(FM_i(x,y)) \qquad (2)$$

Where $i = 1 \ldots N$ and $N$ the number of images of sequence. $FM_i(x,y)$ is the focus measure applied in a local window around the pixel $(x,y)$ for the i$^{th}$ image. Several focus measures exist in literature (Krotkov, 1987). We choose to use the variance of Tenenbaum gradient (Tenengrad) (Pech-Pacheco et al., 2000). Since a sharp and textured image has more pronounced edges, it seems natural to use an edge detector to calculate sharpness. The amplitude of gradient is calculated by equation 4 where $G(x,y)$ is the convolution between image $I(x,y)$ and Soble operators $S_x$ and $S_y$ (equation 3).

$$S_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \qquad S_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \qquad (3)$$

$$S(x,y) = \sqrt{[G_x(x,y)]^2 + [G_y(x,y)]^2} \qquad (4)$$

The variance of Tenengrad is given by:

$$FM_{ten\_var}(i,j) = \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} [S(x,y) - \bar{S}]^2 \qquad (5)$$

Where $\bar{S} = \frac{1}{NM} \sum_{x=i-N}^{i+N} \sum_{y=j-N}^{j+N} S(x,y)$ and $N$ the size of neighborhood.

With this measure, we obtain a curve (fig. 3) for each pixel where the position of maximum represents the image for which the current pixel is sharp.

Then, we use a Gaussian approximation to refine the result of position of sharpest pixel. In figure 3, the cross indicates the approximate position. We search $\bar{d}$ where the focus value is maximum ($F_{peak}$). Three measures are used: $F_{m-1}$, $F_m$ and $F_{m+1}$ (fig. 3).

$$\bar{d} = d_m + \frac{\log(F_{m-1}) - \log(F_{m+1})}{2\log(F_{m-1}) - 4\log(F_m) + 2\log(F_{m+1})} \qquad (6)$$

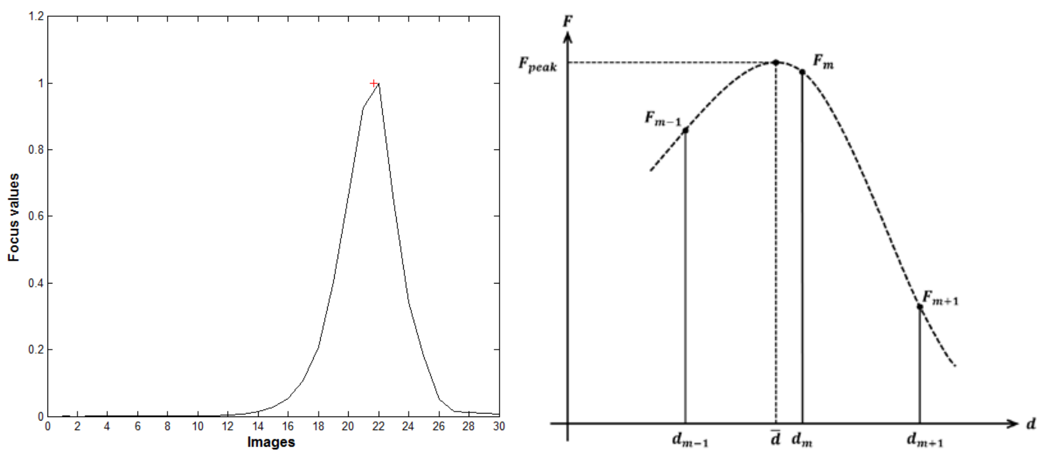


**Fig. 3. Focus measure curve and Gaussian approximation**

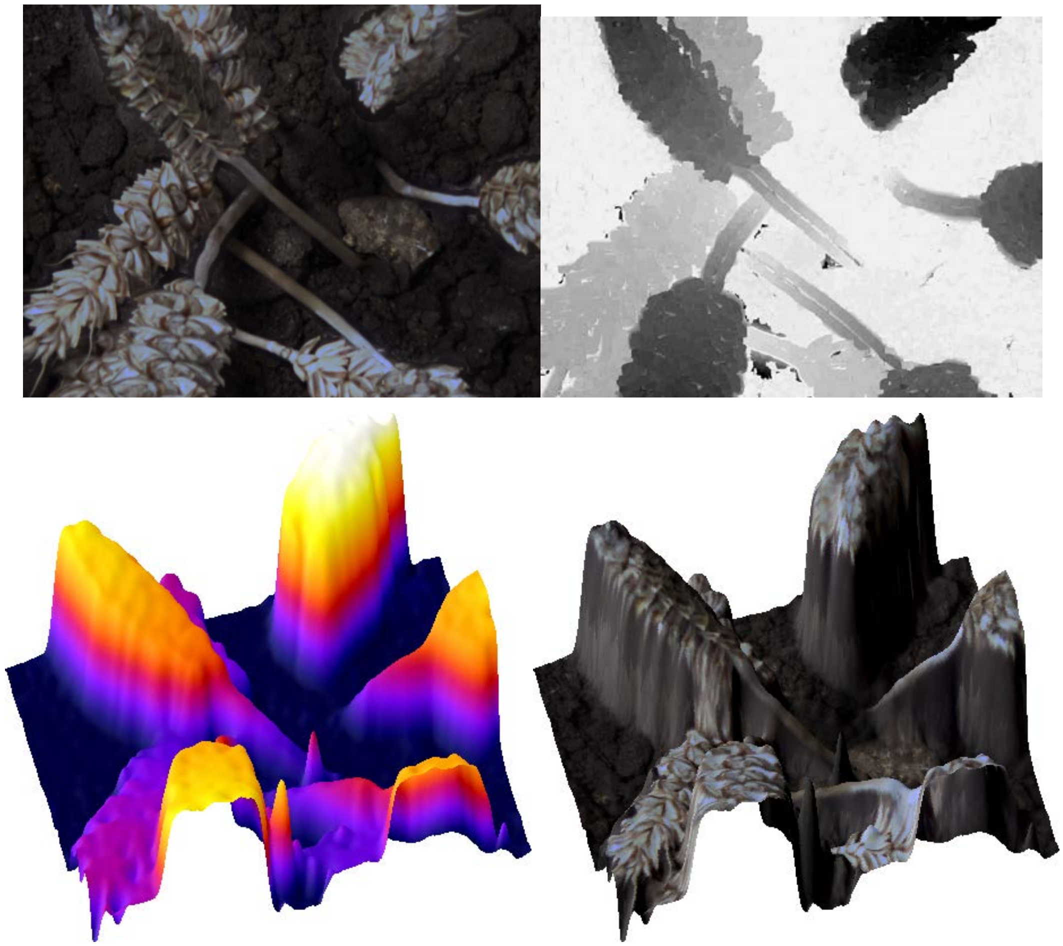

**Fig. 4. Merged image from the sequence, associated depth map and 3D visualization (with and without texture mapping)**

## Depth map and discussion

When the approximate position of each point is found, these results are used to retrieve the depth information of all points. We associate a grey level to each point according to the sharp position to obtain the depth map. In figure 4, darkest are the pixels, nearest are the corresponding points in the scene.

The primary purpose of this depth map is the association with our previous research based on automatic counting of wheat ears. Indeed, the depth information allows us to distinguish objects which are not located on the same spatial plane but overlapped in 2D image. The knowledge of spatial location of each point of scene can also eliminate unnecessary information such as soil, for example, during a classification of leaves and wheat ears.

Currently, we design another acquisition system to increase field of view available and keep a small depth of field. A new camera and new lens will be used to perform that. This future system will be always based on shape from focus technique.

## CONCLUSION AND FUTURE WORK

This paper describes a 3D acquisition system using shape from focus technique in a macroscopic level. The image registration problem is solved by phase correlation method. The depth map of scene based on the use of variance of Tenengrad is created. This system is designed to complete and open up perspectives about our previous research in evaluation of an early yield. We can easily use it in the field and it will allow us to start an acquisition campaign to obtain a database of images. This database will allow us to follow the evolution of crop and extract new parameters. An important point for any use of this acquisition method is the tradeoff between accuracy and available field of view. Indeed, accuracy is linked to depth of field and depends on the acquisition system.

The achievement of depth information allows going further in feature extraction necessary to characterize cultures. We will focus on the optimization of our system to obtain an accuracy which allows the evaluation of the number of grains in wheat ears.

Afterwards, a database of several kinds of crops will be built. This database allows us to test the robustness of our system according to different scenes. A direct agronomic application would be the assessment of winter losses of different crops.

## REFERENCES


Al Hiary, H., Bani Ahmad, S., Reyalat, M., Braik, M. and Z. ALRahamneh. 2011. Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications, Foundation of Computer Science. 17:31-38.

Barnard, S. and M. Fischler. 1982. Computational stereo. ACM Computing Surveys. 14:553-572.

Burks, T.F., Shearer, S.A., Gates, R.S. and K.D. Donohue. 2000. Backpropagation neural network design and evaluation for classifying weed species using color image texture. Transactions of the ASAE, American Society of Agricultural Engineers. 43:1029-1037.

Cointault, F., Guérin, D., Guillemin, J.P. and B. Chopinet. 2008. In-Field Wheat ears Counting Using Color-Texture Image Analysis, New Zealand Journal of Crop and Horticultural Science. 36: 117–130.

Cointault, F., Journaux, L., Destain, M.F. and P. Gouton. 2008. Wheat ear detection by textural analysis for improving the manual counting. International Conference on Signal Processing, Pattern Recognition and Applications.

Germain, C., Rousseau, R. and G. Grenier. 1995. Non destructive counting of wheatears with picture analysis. Image Processing and its application. 453-439.

Krotkov, E. 1987. Focusing. International Journal of Computer Vision. 1:223-237.

Nayar, S. K. and Y. Nakagawa. 1994. Shape from Focus. IEEE Transactions Pattern Analysis and Machine Intelligence. 16:824-821.

Pech-Pacheco, J., Cristobal, G., Chamorro-Martinez, J. and J. Fernandez-Valdivia. 2000. Diatom autofocusing in brightfield microscopy: a comparative study. International Conference on Pattern Recognition.

Raj, A. and R. Staunton. 2007. Estimation of image magnification using phase correlation. International Conference on Computational Intelligence and Multimedia Applications. 3:490-494.