Volume estimation of unbroken soybeans samples using digital image processing techniques

Jorge E. Villaverde[a,*], Mario Cleva[a]

[a] Universidad Tecnológica Nacional – Facultad Regional, French 414 - Resistencia - Chaco – Argentina

[*] Corresponding Author: jorgevillaverde@frre.utn.edu.ar

Abstract

The calculation of volume of different oilseed grains, through computational models, has demonstrated effectiveness and efficiency. In the present work, the model has been extended to allow calculations of the volume of soybeans. In the model is proposed each grain of the sample is assimilated to a parallelepiped with main axes L (length), W (width) and T (thickness). The L and W values are determined from the Feret distances of the image, and the thickness is assumed to be proportional to width of the grain. The proportionality constant k is calculated using the formula of the model and validated against the experimental volume of the samples, fielding a confidence or percentual relative deviation. The model developed approximates soybean volume with confidence of a 1.25% using low-cost hardware for image acquisition and moderate computational resources.

Keywords: grain morphology; Feret distance; ImageJ

1. Introduction

Soybean production has been relevant, during the last decades, for most agro-industrial countries (Hamza et al., 2024; Pagano and Miransari, 2016; da Costa et al., 2011). Therefore, food quality is a paramount feature in agriculture technology. Recent advances in computer vision have made technology ubiquitous to processing vast amounts of data in cloud environments (Alharbi and Aldossary (2021); Van der Merwe et al., 2010).

Experimental object volume determination can be performed using toluene displacement method (Mohsenin, 1986). Volume determination is a rutinary task and is related with other quality indicators, such as density. This method is based on Archimedes' principle and uses a pycnometer with toluene for volume determinations. Toluene is used instead of water because, having a surface tension 25 times lower, it is not absorbed by the grain. However, the use of toluene requires adequate laboratory configuration and special techniques due to its high toxicity. Likewise, this determination causes each processed sample to be modified, making it impossible to reuse it for other purposes.

In the case of soybeans, there are studies that have already experimentally determined their physical properties (Deshpande et al., 1993), which present us with a framework to validate the veracity and accuracy of our results. Generally, for grains and seeds a magnitude similar to density known as hectoliter weight is used, which is defined as the weight in kilograms of a level container of grains with a volume of 100 liters or its proportional value. For example, for rice grains, the hectoliter weight is a good estimation of both the physical quality of the grain and milling quality (Garnero, 2012). However, this magnitude, also called bulk density, is not a good measure of grain density, which could have been used to approximate volume. Its determination is carried out with a Schopper-type balance, and the measurement of the magnitude takes into account the inter-grain spaces that affect the volume measurement, giving a lower density value. The difficulty in determining the density of a grain sample on a routine basis could be due to the complexity of determining the volume of a grain sample in a simple and non-destructive manner (Ferrari et al., 2021; Cleva et al., 2017).

Previous works have proposed using this method to approximate the volume of different grains through digital image processing techniques (Ferrari et al., 2021; Cleva et al., 2017). In this sense, this work aims to expand the spectrum applied to different grains and to validate the method adjustment/correction, as well as to verify its correct application through digital image processing techniques.

Recently, mathematical models similar to the one presented in this work have been used to calculate the volume of soybeans (Miranda et al., 2022; Nevavuori et al., 2019; Sadeghi-Tehran et al., 2019). Moreover, (Zhao et al., 2021) present a model for detecting soybeans by means of deep convolutional neural networks (Deep Convolutional Neural-Networks), which although they are capable of detecting, in real time, seed defects, unlike the present works, do not focus on the calculation of seed properties.

Kaliniewicz et al. (2022) have also presented a similar approach for volume determination of seeds, using seeds dimension and approximation of a constant (or volume coefficient) that best fits different spiciest with complex shape and size. They have also used pycnometer for determining the real volume of seeds, and as a parameter to calculate the volume coefficient. Our model differs in the determination of the dimensions of the seeds (soybeans in this case) and how we calculate the approximation.

The aim of this study is to introduce a methodology for volume approximation of a soybean sample using soybean grain geometric model and simple software calculations. It requires a digital scale, a desktop scanner and Digital Image Processing techniques that are applied to the digital image of the sample. The methodology uses a geometric model of the soybean that allows its volume to be determined proportionally to the product of the square of the width and the length of the grain. The proportionality constant is obtained experimentally, the width and length values are obtained by Digital Image Processing algorithms.

2. Materials and method

2.1. Model description

Figure 1

Figure 1 shows a three-dimensional representation of a soybean grain. It shows the way in which a parallelepiped whose length L_i (length), W_i (width) and T_i (thickness) coincide with the magnitudes of the grain can be approximated. Its volume can be described by equation (1).

$$V_i = k_1 T_i W_i L_i$$

Equation 1: Parallelepiped volume approximation.

Where soybean grain volume (V_i) can be expressed as a total volume fraction of the parallelepiped, and k_i being this fraction. Therefore, $k_i \leq 1$. On the other hand, taking into account the shape of the soybeans, equation (2) can be derived.

$$T_i < W_i < L_i$$

Equation 2: Soybean observation.

When the grain rests on a flat surface, a plane is generated by lengths W_i and L_i that are parallel to the plane of support, which validates the assumption that W_i is greater than T_i. Thus, when a scanner is used and the grains are placed in the tray, it records an image for each grain in which W_i and L_i are always visible. Taking this into consideration, we can yield Equation 3 as a proposed thickness approximation.

$$T_i = k_2 W_i$$

Equation 3: Proposed thickness approximation

Then, replacing (3) in (1) we get (4).

$$V_i = \mathrm{k_1 T W_i L_i} = \mathrm{k_1 k_2 W_i W_i L_i} = k\mathrm{W_i^2} L_i$$

Equation 4: Replacement of T_i.

Where,

$$k = k_1 k_2$$

Equation 5: k unification.

Therefore, for a sample of *n* grains, the total sample volume obtained by digital image processing, V_{pdi}, can be presented as:

$$\mathrm{V_{pdi}} = k \sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{W_i^2 \, L_i}$$

Equation 6: Volume approximation using image processing technique.

In this model it is assumed that this constant does not vary for different grain dimensions, as long as they are unbroken.

2.2. Determination of the constant *k* of the model

For a sample of *n* grains, the experimental volume V_{exp} is determined using the toluene displacement method (Mohsenin, 1986). Similarly, between the experimental volume V_{exp} and the sample volume V_{pdi} the adjustment constant k can be determined, as shown in equation 7.

$$k = \frac{\mathrm{V_{exp}}}{\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{W_i^2\, L_i}}$$

Equation 7: k determination by volume approximation.

The length L_i for each grain is obtained as the maximum length of what is called the Feret diameter, which is defined as the distance between two parallel tangent lines on opposite sides of the grain boundary (Merkus, 2009). The maximum distance perpendicular to the previous one between opposite points of the contour is taken as width W_i. The determination of these distances is carried out using image processing techniques.

With the value of the constant *k* determined, the expression in equation 6 is used to calculate the volume of the sample. Since the volume of the grain is smaller than that of the parallelepiped that contains it, the value of the constant *k* must be less than unity.

2.3. Samples preparation

The soybean grain samples were extracted from a commercial package. A total of 9 samples, each of which has 100 whole grains without visible cracks were prepared (Figure 2), making a total of 900 grains. The selection of these grains was carried out at random for all cases and they were extracted from a single package. The samples were visually inspected, with the aid of a magnifying glass and the number of grains in the sample were selected to fit on the document table of the scanner.

Figure 2

Each sample consisted of 100 grains. Their masses were determined with a Denver Instrument model MXX 612 digital lab scale with a precision of 0.01 grams. This determination is necessary for the calculation of the experimental volume required in the toluene displacement method. From the total number of samples, 5 were used to determine the proportionality constant k and the remaining 4 were used to validate the proposed model.

2.4. Image acquisition

To acquire soybean images, a Hewlett Packard G3110 scanner was used. Images were scanned using 300 dots per inches (DPI) and afterwards processed using ImageJ®[1] software (Broeke et al., 2015). ImageJ® is a Java based open-source software developed by the National Institutes of Health for scientific. This software was designed with an extensible architecture via Java plugin, which allows the incorporation of new functionality. It has also the ability of providing macros to record repeatable use actions that can be later applied to other images, simulating a program to process images (Haeri and Haeri, 2015).

Figure 3

Using this software, the first step was separating the images into RGB channels and selecting the one that generate more contrast between the grains and the background of the image. As shown in Figure 3, the red channel was selected for all images, since this was a better fit. We can see in Figure 3(b) that its histogram presents a better contrast, and it will yield a cleaner image binarization.

[1] https://imagej.net/

Afterwards, image binarization was perform in order to reduce the image quality to an 8-bit image using the threshold of 109 over 255. This process produces images 4(a) and 4(b) where we can see the red channel binarization process and the contour detection afterwards.

Figure 4

Once the image has been binarized we can apply the analysis of particles in order to calculate the Feret maximum and minimum distances, which will be traduced to length and width of the grain (Broeke et al., 2015; Cleva et al., 2017).

3. Results

Table 1 presents, for each sample, the total number of grains that make it up, its experimentally determined volume V_{exp}, the sum of the square of the width W_i times the length L_i and the corresponding constant *k*. Finally, the value of the constant *k* and its standard deviation are calculated.

Table 1

As it is shown in Table 1, the value of the constant *k* has a standard deviation of the order of 1.2% of the average value. In all cases the value of the constant *k* was less than unity as previously mentioned.

$$\mathrm{PRD} = \frac{\mathrm{V_{exp}} - \mathrm{V_{pdi}}}{\mathrm{V_{exp}}} 100$$

Equation 8: Percentual Relative Deviation.

Table 2

Table 2 shows the results of the model using the control samples for validation purposes. This table shows that the Percentual Relative Deviation of the control samples yields 1.25%, resulting a good approximation by the control group.

4. Discussion and Conclusions

This work presents a method for calculating the volume of soybeans using digital image processing techniques, a desktop scanner, and minimal computational requirements.

The proposed approximation model yields result with a 1.2% stand deviation, making it accurate enough to be used in the industry. The method is safe, fast, non-destructive, robust, operator-independent, low-cost and easy to implement. Another characteristic of the proposed method is that it does not require complex hardware configurations nor calibration processes, so its implementation does not present operational difficulties.

The development of geometric models for determining volume in grains has already been used by the authors of this work. In contrast to previous work, this one only works with whole soybeans. However, it is estimated that the work methodology is applicable to other types of grains and seeds, a procedure that is being carried out.

References

Alharbi, H. A. and Aldossary, M. (2021). Energy-efficient edge-fog-cloud architecture for iot-based smart agriculture environment. *IEEE Access*, 9:110480–110492.

Broeke, J., Pérez, J. M. M., and Pascau, J. 2015. *Image processing with ImageJ*. Packt Publishing Ltd.

Cleva, M. S., Sampallo, G. M., Rodrigues Da Fonseca, C., Aquino, D. C., Liska, D. O., and Ventre, F. R. (2017). Método para la determinación de la densidad de una muestra de granos enteros de alubia por procesamiento digital de imágenes. *RIA. Revista de investigaciones agropecuarias*, 43(2):144–148.

da Costa, M. L., Dorr, A. C., dos Reys, M. A., and Strassburger, R. 2011. Strategies adopted by agro-industry complex of soybean players in paraguay. *CEP*, 97015:900.

Deshpande, S., Bal, S., and Ojha, T. 1993. Physical properties of soybean. *Journal of Agricultural Engineering Research*, 56(2):89–98.

Ferrari, E., Quiriban, A., Gili, A., Picca, A., and Repollo, R. (2021). Caracterización morfológica del grano de triticale (x triticosecale wittmack) mediante procesamiento y análisis de imágenes digitales. *RIA. Revista de investigaciones agropecuarias*, 47(2):232–239.

Garnero, S. 2012. Calidad intrínseca de los granos en la poscosecha. Technical report, Universidad Tecnológica Nacional.

Gomes, J. F. S. and Leta, F. R. 2012. Applications of computer vision techniques in the agriculture and food industry: a review. *European Food Research and Technology*, 235(6):989–1000.

Gómez-Ríos, A., Tabik, S., Luengo, J., Shihavuddin, A., Krawczyk, B., and Herrera, F. 2019. Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Systems with Applications*, 118:315–328.

Haeri, M. and Haeri, M. 2015. ImageJ plugin for analysis of porous scaffolds used in tissue engineering. *Journal of Open Research Software*, 3(1).

Hamza, M., Basit, A., Shehzadi, I., Tufail, U., Hassan, A., Hussain, T., Siddique, M., and Hayat, H. (2024). Global impact of soybean production: A review. *Asian Journal of Biochemistry, Genetics and Molecular Biology*, 16:12–20.

Kaliniewicz, Z., Choszcz, D., and Lipiński, A. (2022). Determination of seed volume based on selected seed dimensions. *Applied Sciences*, 12(18).

Merkus, H. (2009). *Particle Size Measurements: Fundamentals, Practice, Quality*. Particle Technology Series. Springer Netherlands.

Miranda, M. C. d. C., Aono, A. H., and Pinheiro, J. B. 2022. A novel image-based approach for soybean seed phenotyping using machine learning techniques. *Crop Science*.

Mohsenin, N. 1986. *Physical Properties of Plant and Animal Materials: Structure, Physical Characteristics, and Mechanical Properties*. Structure, Physical Characteristics, and Mechanical Properti. Gordon and Breach.

Nevavuori, P., Narra, N., and Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Computers and Electronics in Agriculture*, 163:104859.

Pagano, M. C. and Miransari, M. (2016). 1 - the importance of soybean production worldwide. In Miransari, M., editor, *Abiotic and Biotic Stresses in Soybean Production*, pages 1–26. Academic Press, San Diego.

Sadeghi-Tehran, P., Virlet, N., Ampe, E. M., Reyns, P., and Hawkesford, M. J. (2019). Deepcount: In-field automatic quantification of wheat spikes using simple linear iterative clustering and deep convolutional neural networks. *Frontiers in Plant Science*, 10.

Van der Merwe, J., Ramakrishnan, K., Fairchild, M., Flavel, A., Houle, J., Lagar-Cavilla, H. A., and Mulligan, J. 2010. Towards a ubiquitous cloud computing infrastructure. In *2010 17th IEEE Workshop on Local & Metropolitan Area Networks (LANMAN)*, pages 1–6.

Zhao, G., Quan, L., Li, H., Feng, H., Li, S., Zhang, S., and Liu, R. 2021. Real-time recognition system of soybean seed full-surface defects based on deep learning. *Computers and Electronics in Agriculture*, 187:106230.

Samples	$V_{exp\ (cm^3)}$	$\sum_{i=1}^{n} W_i^2 \, L_i(cm^3)$	**k**
1-100	13.5414	23.2793	0.5817
101-200	13.3603	22.2727	0.5998
201-300	13.1506	22.2254	0.5917
301-400	13.9454	23.3865	0.5963
401-500	13.616	22.7214	0.5993
Mean			0.5938
Standard Deviation			0.0128

Table 1: Measured and calculated parameters for the analyzed samples.

Control sample	**Experimental Volume** (cm^3)	$V_{pdi}\ (cm^3)$	**PRD**
501-600	13,7077	13.5234	1.34%
601-700	14.0237	14.8612	5.97%
701-800	13.7511	13.7035	0.35%
801-900	13.7381	13.5785	1.16%
Mean			1.25%

Table 2: Volume results obtained by image processing and by toluene displacement for control samples.

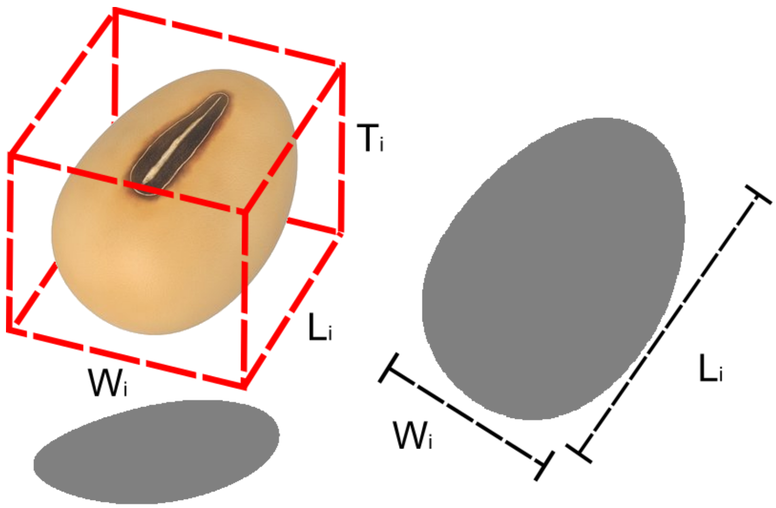


Figure 1. Three-dimensional representation of a soybean.

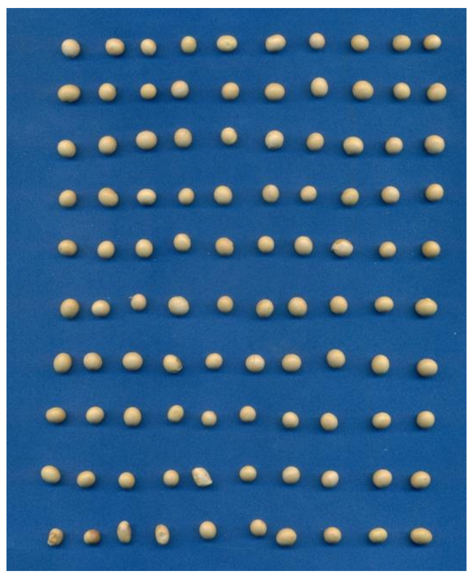
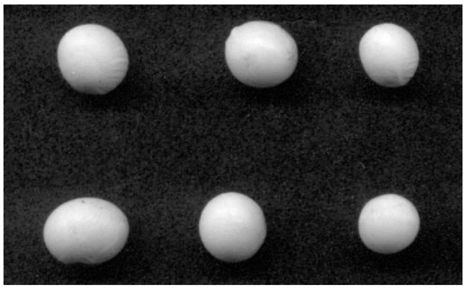
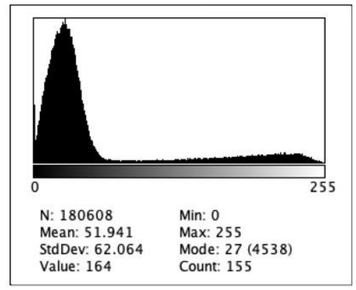


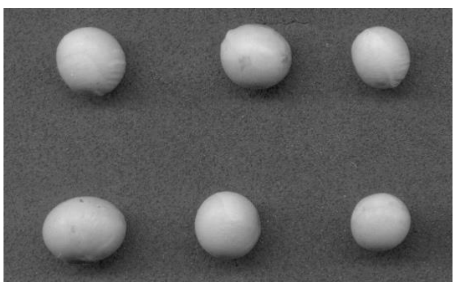
Figure 2. Sample being processed (100 grains).



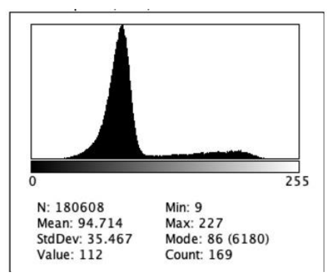
(a) red channel



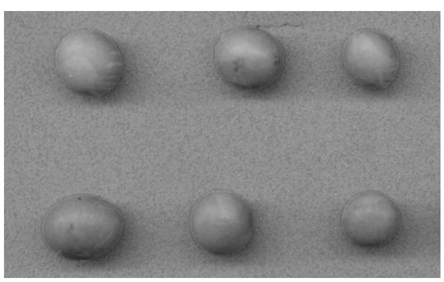
(b) red channel histogram



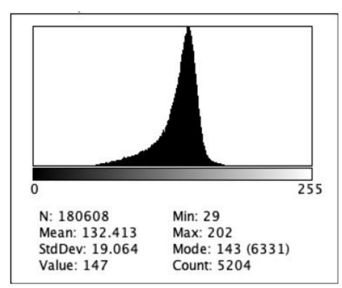
(c) blue channel



(d) blue channel histogram

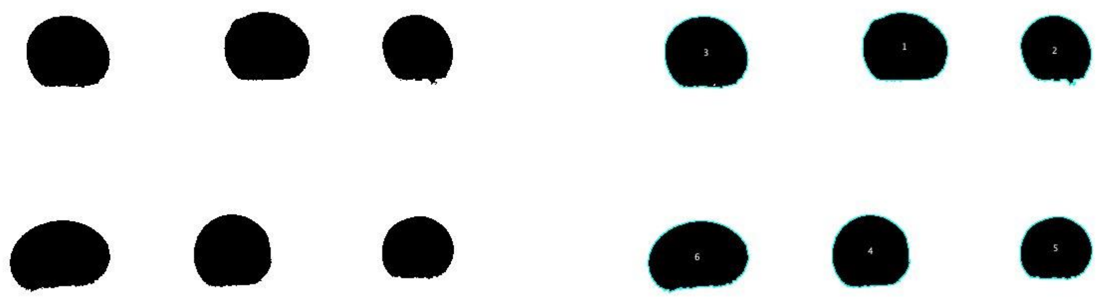


(e) blue channel



(f) blue channel

Figure 3. Image Channel separation.



(a) Binary image (b) Contour of binary images

Figure 4. Image pre-processing.