Image processing to identify damage to soybean seeds

Rita de Cassia Mota Monteiro[1*] Gizele Ingrid Gadotti[2] Vanessa Maldaner[3]
Amanda Bento Jorge Curi[4] Michaela Bárbara Neto[5]

[1]Programa de Ciência e Tecnologia de Sementes, Faculdade de Agronomia Eliseu Maciel (FAEM), Universidade Federal de Pelotas (UFPel), 96160-000, Pelotas, RS, Brasil. E-mail: ritamonteiroo@gmail.com. *Corresponding author.
[2]Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brasil.
[3]Programa de Pós-graduação em Engenharia Agrícola (PPGEA), Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.
[4]Curso de Engenharia Agrícola, Universidade Federal de Pelotas (UFPel), Pelotas, RS, Brasil.
[5]Programa de Pós-graduação em Engenharia de Sistemas Agrícolas (PPGESA), Universidade de São Paulo (USP), Butantã, SP, Brasil.

ABSTRACT: *In Brazil and worldwide, commercialization of soybeans is of great importance to the economy, making their quality considered. The presence of damaged soybean seeds decreases the added value of the product. Businesses need fast and effective techniques to maintain the quality. The present research aimed to identify, through image processing, damage caused to soybean seeds, namely the presence of greenish seeds and wrinkled seeds due to variations of humidity and temperature, where it was possible to identify greenish and wrinkled soybean seeds from images. Results obtained for greenish seeds indicated that the red color scale is the most suitable for selection due to its more significant variation compared to the other color scales. For the separation of wrinkled seeds, it can be stated that it is possible to find a selection parameter with 74.3% accuracy in removing seeds with medium to high degrees of wrinkle damage.*
Key words: *greenish seeds, moisture damage, RGB, deterioration, wrinkled.*

Processamento de imagens para identificação de danos causados em sementes de soja

RESUMO: *No Brasil e no mundo, a comercialização da soja é de grande importância para a economia, sendo considerada sua qualidade. A presença de sementes danificadas de soja diminui o valor agregado do produto. As empresas precisam de técnicas rápidas e eficazes para manter a qualidade. O presente trabalho visa identificar, através do processamento de imagens, danos causados em sementes de soja, sendo eles: a presença de sementes esverdeadas e sementes enrugadas devido às variações de umidade e temperatura. Os resultados obtidos para sementes esverdeadas indicaram que a escala de cor vermelha é a mais adequada para seleção devido a sua variação mais significativa em relação às demais escalas de cores. Para a separação de sementes enrugadas, pode-se afirmar que é possível encontrar um parâmetro de seleção com 74,3% de acerto na remoção de sementes com médio a alto grau de danos por enrugamento.*
Palavras-chave: *sementes esverdeadas, dano por umidade, RGB, deterioração, enrugadas.*

INTRODUCTION

Soybean [*Glycine max* (L.) Merril.] is the most important *Fabaceae* currently grown globally, and is the main agricultural product exported (ESPÍNDOLA et al., 2015). Because its cultivation is distributed throughout the whole of Brazil, it is possible that seed production may occur under different environmental conditions.

Seeds of high physiological quality have characteristics such as a high vigor, germination, and health (KRZYZANOWSKI et al., 2018). In this scenario, damage such as the presence of greenish soybean seeds and deterioration due to variations in humidity and temperature, emphasizing the skin's wrinkling, cause the added value of the product to decrease.

High levels of wrinkling can be caused by a combination of two factors: dryness and high temperatures. These problems can damage the seeds' quality and may reduce their germination, viability, vigor, and seedling emergence index. Seed wrinkling causes expansions and contractions of the integument, and the reason for this damage is related to variations in humidity and exposure of seeds to high temperatures. Good seed quality is necessary to obtain products with high added value,

playing a fundamental role in the production system, thus meeting the need for excellent productivity (FRANÇA-NETO et al., 2016).

Also according to FRANÇA-NETO et al. (2016), several factors cause the appearance of green color in seeds, which varies according to the occurrence of environmental stresses and may even cause premature death or forced maturation of the plant, and depending on the intensity with which the greenish soy seeds occur in the lot, may result in problems in the lot produced due to low physiological quality. The appearance of greenish seeds in the soybean crop can occur due to several factors, such as diseases, inadequate selection of the pre-harvest desiccation point, and high temperatures during the maturation phase. The presence of green color in a seed indicates that it did not finish maturation efficiently and thus did not fully degrade the existing chlorophyll.

According to studies by JHAWAR (2016) & SANTIAGO et al. (2019), the use of digital image processing must be considered because it is an alternative that may contribute to the processing of seeds.

These techniques allow a quick, effective, and non-destructive analysis, and techniques based on the use of RGB (system based on primary colors red-green-blue) images are increasingly used to evaluate the color and shape of seeds, thus guaranteeing their high quality (LIU et al., 2015; MAHAJAN et al., 2018). Thus, in recent years, the use of algorithms with appropriate pre-processing, with the aim of developing a system with product characteristics with specific qualities, has become of great importance and utility for businesses.

Considering that the quality of soybean seed lots depends on the crop produced and the production site, among other things, moisture-damaged and greenish seeds can vary in volume by crop and location, which is a problem for the seed sector, causing a high discard rate. Thus, the possibility of improving efficiency through image processing to separate these damages seeds can make a great contribution to the sector.

Therefore, the present research aimed to identify, through image processing, damage to soybean seeds.

MATERIALS AND METHODS

The tests related to the present research were conducted in two consecutive years at the Federal University of Pelotas (UFPel) using three soybean lots. The greenish seeds evaluated came from the 2016/2017 harvest and the studied seeds with tegument wrinkling came from the 2017/2018 harvest.

Analysis of greenish soybean seeds
Sampling

For the preparation of the analysis of greenish soybean seeds, three lots containing greenish soybean seeds of different shades were received and were manually separated into samples of yellow, intermediate green I (greenish-yellow), intermediate green II (yellowish-green), and green seeds, making four types of samples. This separation was performed by the same person to avoid variations in tone.

The images were digitalized in a scanner with a black EVA (ethyl vinyl acetate) background with dimensions of 22 × 30 cm and an area of 11 × 11 cm, which was delimited due to the size of the samples. The images were captured in RGB and then processed.

The scanned images were processed using MATLAB software, where an algorithmic language was developed in C language, a compiled programming language, whose function was to verify which of the color bands allowed the easiest separation of the seeds, for future use in equipment for selection by color in processing units. For this, histograms were generated and with these were verified as major differences in shade and; therefore, which shade would be more efficient to separate the greenish seeds. Figure 1A illustrates the sequence used in the study from the arrival of the soybeans to obtain results.

Statistical analysis

Afterward, a descriptive statistical analysis was carried out with the aim of obtaining the largest amount of information possible from the data. The collected data were submitted to analysis of variance (P≤0.05) and then compared using the Tukey test with a 5% probability.

Analysis of wrinkled soybean seeds
Sampling

For the preparation of the analysis of wrinkled seeds, lots containing 90% damaged seeds lots were used in which 90% of seeds showed damage at different levels due to variations of humidity and temperature. For that, seven samples containing ten seeds each were separated, totaling 70 seeds with a high incidence of damage from the three lots.

Samples were separated manually by the same person to avoid variations; later, they were placed in a scanner. A black EVA background was used, with a grid of the same material and dimensions
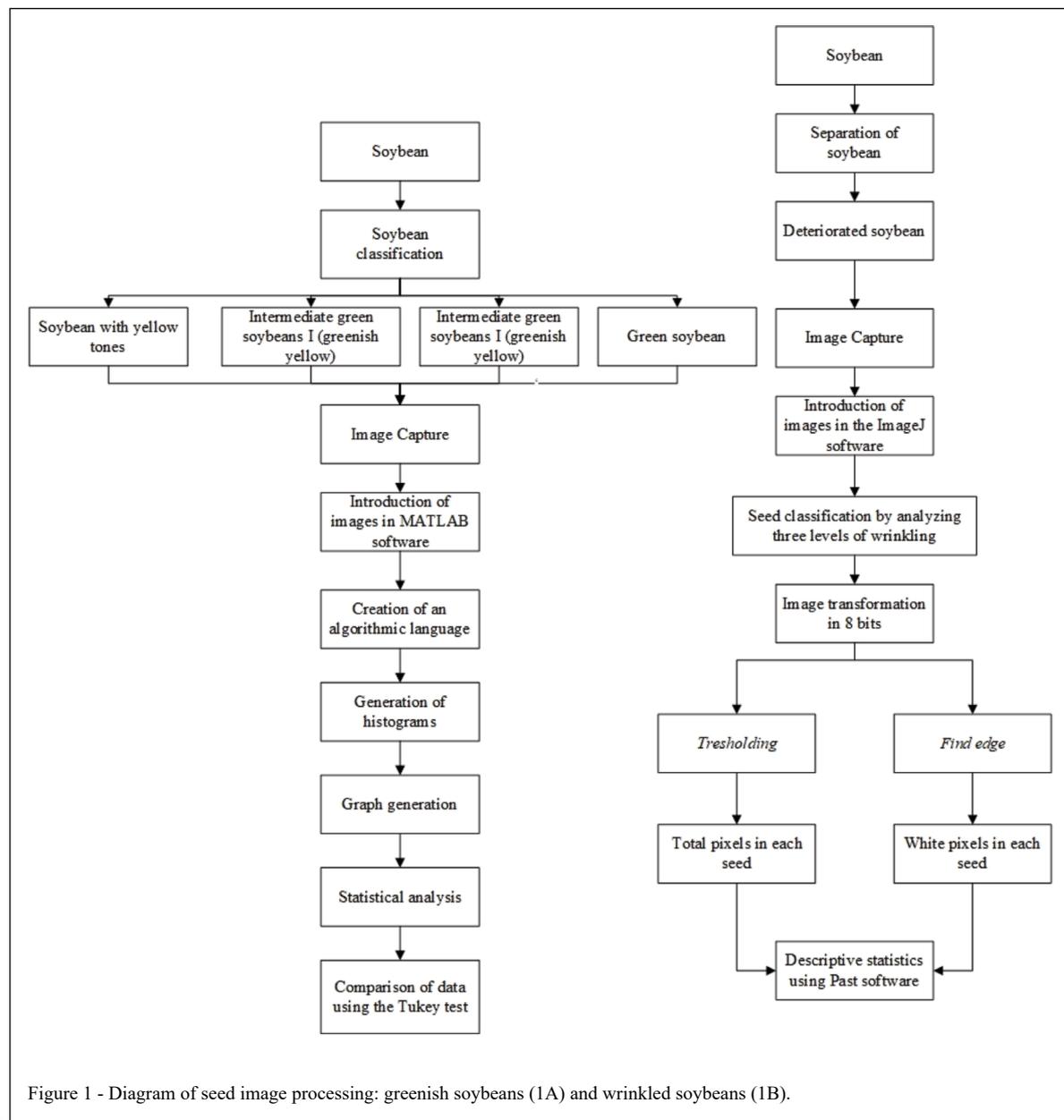
Figure 1 - Diagram of seed image processing: greenish soybeans (1A) and wrinkled soybeans (1B).

of 2 × 2 cm, to analyze the soybean seeds separately; these seeds, in turn, were arranged with the hilos facing one of the sides.

In sequence, the images were imported into ImageJ software, used for processing and extracting information from RGB images. Before extracting the information from the images, it was necessary to insert the analyzed individuals' color scales, using the 2 × 2 cm EVA grid measurement as a real measurement for the calibration tool. The reading of EVA pixels did not hinder the processing

of images in the software; the images were treated with the threshold tool, eliminating their variations.

Through the visualization of the enlarged images, classification of the seeds was achieved by analyzing three levels of wrinkling: high, when the damage was severe and evident; medium, in cases of small wrinkled areas; and low, when no wrinkles were visible (Figure 2).

With the aid of the Image J software and its pixel selection tool, multiple selections were worked out in the images, establishing the correct region of

interest (ROI) in the center of each image. For this purpose, frames of each image were cut using a known radius for the cropped image imported. This process was done to increase efficiency in performing the following steps.

In order to extract the information, it was necessary to transform the RGB image to eight bits, transforming it to grayscale, containing 256 possible gray tones ranging from zero (absolute black) to 255 (absolute white). The image binarization was necessary so that the seeds could be identified individually, represented by a contour and a number, helping to obtain several characteristics such as the projected area on a plane, perimeter, and pixel count in each of the identified regions, according to the interest of the work.

After transforming the image, two segmentation steps were performed: *threshold* and *find edge*. For the first, each image was divided into two or more classes of pixels (binary image). With that, the images were analyzed pixel by pixel, thus obtaining the total value of pixels in each seed, separating the background, classified as 255 (absolute white), of the object of interest. In the second stage, the filter was used as an edge detector, and, with its help, it was possible to check the wrinkles in the seed coat caused by variations in humidity and temperature. Figure 1B illustrates the sequence used in the study from the arrival of soybeans to obtain results.

Statistical analysis

Descriptive statistics of the data were obtained using PAST software, where the reality observed concerning the total number of white pixels in each lot could be described. With the aid of this software, traditional statistical data related to the physical properties of the pixel area of the images were obtained, in addition to the frequency histogram graphs, cumulated frequency, and the use of averages with confidence intervals.

RESULTS AND DISCUSSION

Green soybeans

With the results obtained from the three lots, averages were generated for each color scales, thus reaching a single frequency value. According to the values obtained, only the distribution of the component in the red scale showed a significant variation between the colors yellow, intermediate I, intermediate II, and green, in addition to highlighting the difference between all scales (Figure 3).

The analysis of variance (Table 1) was used to rectify the differences in the distribution of the greenish soy components, in which showed that, for the gray color scale, the average frequency did not show a difference between the soy seeds on any of the color scales. However, seeds differed from each other in color intensity.

The gray and green scales were statistically similar to each other for all treatments, except for the green-colored seeds. The red scale had the same averages for all treatments but differed from the other scales. The blue scale did not show a variation in the first three types of seeds; however, these data are of the same significance as those reported in the grayscale for green seeds.
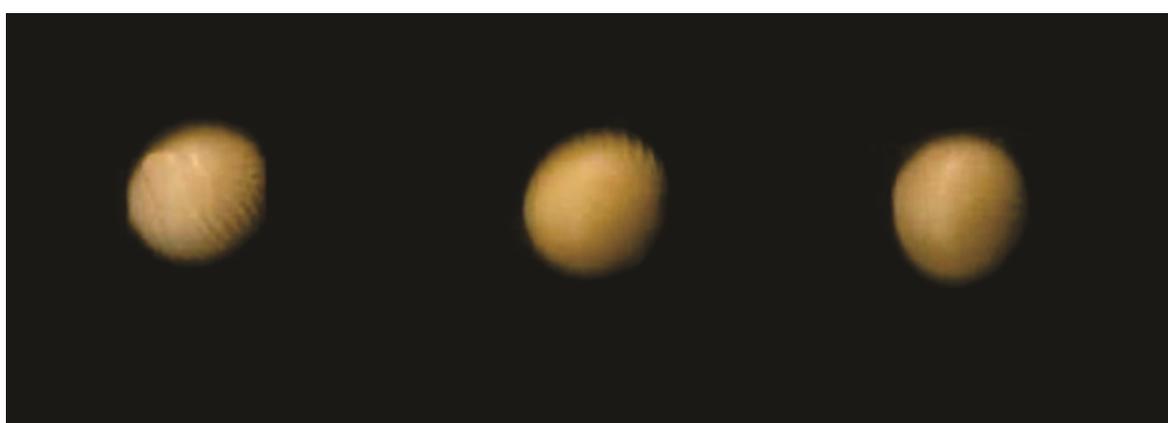
Figure 2 - Classification of high, medium and low damage in soybean seeds, respectively.

Figure 3 - Average pixel distribution of the color scale within each color scale (Average of yellow (A), yellow-green (AV1), yellow-green (AV2) and green (V)).
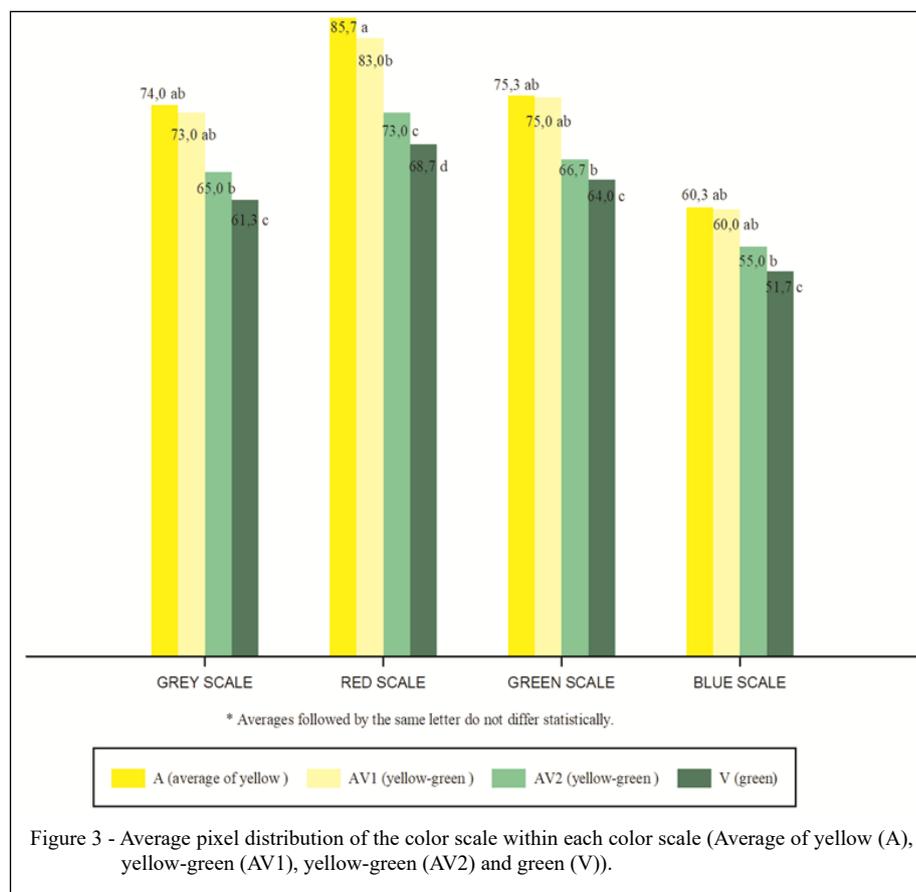
For rice, MONTEIRO, GADOTTI, and ARAÚJO (2019) obtained results that indicated that chalky rice grains could be selected on the blue scale in both white and parboiled rice and concluded that the red scale is not indicated for separating different types of rice. For white or parboiled rice, the green scale can be used to separate different types. Grayscale, widely used by the industry, presents intermediate values and is not the most suitable for

the selection process. Thus, in recent years, the use of algorithms with appropriate pre-processing, with the aim of developing a system that can determine specific qualities of product characteristics, has become of great importance and utility for the businesses and is a technique that could be used in seed selection.

MAHAJAN et al. (2018) evaluated the potential of using non-destructive image processing techniques to perform tests of physical purity,

Table 1 - Average of yellow (A), yellow-green (AV1), yellow-green (AV2) and green (V) soybean seeds depending on the color scale.

Treatments	--------Grey--------	--------Red--------	------------Green--------	------------Blue----------
A	74 aB	86 aA	75 aB	60 aC
AV1	73 aB	83 aA	75 aB	60 aC
AV2	65 bB	73 bA	67 bB	55 bC
V	61 cC	69 cA	64 cB	52 cD

*Averages followed by the same letter do not differ statistically, uppercase on the line and lowercase on the column, using the Tukey test at 5% probability.

viability, and vigor of soybean seeds through X-ray images. Results were obtained by correlating the standard germination rate in tests and indicated that the method is effective, fast, and non-destructive, making it a suitable alternative for seed quality tests, contributing to the use of structural tests and removing the limitations of personal inspections.

Wrinkled soybeans

Comparisons between the applied filters demonstrated that the process was efficient in identifying the wrinkling effect. It was possible to establish the total number of white pixels in each seed through the filters applied to the images and the histograms generated. Pixels not distinguished by the naked eye were considered very close to absolute black, whereas pixels considered white had values closer to 255.

Through the analysis of the results obtained by the PAST software, it was possible to remove low error values, standard deviation, and variance, demonstrating the quality of the data (Table 2). It was noted that the coefficient of variation showed that the lots with the highest coefficients could be considered to have the highest damage index.

The confidence interval analysis showed a higher amplitude between the lower and upper limits, showing greater variation in the data and less homogeneity of the pixel values (Table 1). The median lots 4 and 7 showed values close to and above the other values of the interquartile range of the others lots. This interval was chosen to become a selection

parameter, considering that seeds with values equal to or above would discard and be considered as damaged seeds. Lots 1, 4, and 7 had the largest amounts of damaged seeds. The visual classification confirmed that the highest incidence of seeds with a high level of wrinkling occurred in lots 1, 4, and 7, with 50% incidence in lots 1 and 4 and 60% in lot 7 (Table 2).

Analyzing the seven lots and using as a selection criterion a value of 2% white pixels, above which the seeds were discarded, the following results were obtained: 41.94% of the seeds were considered to have serious damage and were eliminated, 23.81% were considered to have medium damage, and 16.67% were considered to have low damage. This selection criterion was made by calculating the accuracy with a value of 74.3% using the following formula:

$$Efficiency = 1 - \frac{(Real\ data - Visible\ data)}{Real\ data} \times 100$$

From this formula, visible damage was accounted. Table 2 indicates the values in percentage of damage through accuracy.

When the default value of white pixels was reduced to 1.9%, the requirement became greater; thus, 48.39% of seeds were considered to have severe damage and were eliminated; 33.33% were considered to have medium damage, and 22.22% were considered to have a low incidence of damage.

Through the criterion adopted as rigid in the selection standard (less than 1.9%), the best lot was composed only of seeds that had low damage to the integument, while seeds with medium and high

Table 2 - Descriptive statistics: confidence interval and accuracy of seven lots of wrinkled soybean seeds.

--Descriptive statistics--						
	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5	Lot 6	Lot 7
Standard error	0.003	0.001	0.002	0.002	0.001	0.001	0.003
Variance	8.150E+0	1.045E-5	3.269E+0	4.484E+0	1.538E+0	1.417E+0	7.401E-5
Standard deviation	0.010	0.003	0.006	0.007	0.004	0.004	0.009
CV(%)	48.561	18.937	31,174	33.906	25.122	22.529	36.192
---Confidence interval---						
	Lot 1	Lot 2	Lot 3	Lot 4	Lot 5	Lot 6	Lot 7
deviation	0.009	0.003	0.006	0.009	0.004	0.004	0.009
Inferior limit	0.012	0.015	0.014	0.012	0.013	0.013	0.018
mean	0.018	0.017	0.018	0.018	0.016	0.017	0.024
Upper limit	0.025	0.019	0.022	0.024	0.018	0.018	0.030
Median	0.015	0.016	0.018	0.020	0.016	0.017	0.022
---Accuracy---						
High	50/100	30/100	40/100	50/100	40/100	40/100	60/100
Medium	20/100	40/100	30/100	20/100	50/100	30/100	20/100
Total	70/100	70/100	70/100	70/100	90/100	70/100	80/100

levels of damage, representing 42.31% of the lot, were discarded. The intention is to eliminate seeds with these levels of damage so that the best lot is obtained. Another standard used was the number of white pixels in the seed: using a parameter of 2%, an efficiency of 34.62% was reached in the detection of seeds that presented medium and high damage; when the parameter was reduced to the rigid one, the efficiency increased to 42.31%.

PESKE & BAUDET (2019) considered that the processing's equipments have low capacity, with that the reprocessing would be indicated so that a new separation takes place and thus higher efficiency of the whole process is achieved. In the processing units, one of the adjustment criteria is the discard of seeds, and in this work, it was 16.67% and 22.22%, respectively. So, the efficiency of the removal is a crucial factor for this type of data. In their study, PIAZ et al. (2018) presented an evaluation of seed quality after use of an electronic sorting machine and the results showed that the selection process was efficient.

In the research of SALIMI & BOELT (2019), alternative methods were used and the study concluded that by using multispectral images with VideometerLab3 and its software, it was possible to classify the mechanical damage to sugar beet seeds.

The techniques used in this research can be carried out quickly with small adjustments in machines that already exist on the market, thus bringing to the sector another tool for the seed processing. Based on the above experiments, the imaging process is feasible, being a step that can be used extensively post-harvest and that can increase the richness of the details of the analyzed damage.

CONCLUSION

It was possible to identify green and wrinkled soybean seeds by image processing. Results obtained for the greenish seeds indicated that the red color scale was the most suitable for selection due to its more significant variation compared to other scales. As for the separation of wrinkled seeds, it can be said that it is possible to find a selection parameter that gives 74.3% accuracy in the removal of seeds considered to have medium to serious wrinkle damage.

DECLARATION OF CONFLICT OF INTERESTS

The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

AUTHORS' CONTRIBUTIONS

Amanda, Gizele and Vanessa conceived, designed and carried out the experiments. Gizele, Michaela and Rita prepared and performed the statistics and prepare the manuscript. All authors critically reviewed the manuscript and approved the final version.

REFERENCES

ESPÍNDOLA, C. J.; CUNHA, R. C. C. A dinâmica geoeconômica recente da cadeia produtiva da soja no Brasil e no mundo. **GeoTextos**, Salvador, v.11, n.1, p.217-238, jul. 2015. Available from: <https://portalseer.ufba.br/index.php/geotextos/article/view/12692/9733>. Accessed: Jun. 10, 2019. doi: 10.9771/1984-5537geo.v11i1.12692.

FRANÇA-NETO, J. B. et al. **Tecnologia da produção de sementes de soja de alta qualidade**. Londrina: EMBRAPA, CNPSo, 2016. 82p. Available from: <https://ainfo.cnptia.embrapa.br/digital/bitstream/item/151223/1/Documentos-380-OL1.pdf>. Accessed: Jun. 10, 2019.

JHAWAR, J. Orange sorting by applying pattern recognition on colour image. **Procedia Computer Science**, Amsterdam v.78, p.691-697, 2016. Available from: <https://www.sciencedirect.com/science/article/pii/S1877050916001204>. Accessed: Aug. 1, 2019. doi: 10.1016/j.procs.2016.02.118.

KRZYZANOWSKI, F. C. et al. A alta qualidade da semente de soja: fator importante para a produção da cultura. **Circular técnica n. 136**. Embrapa Soja: Londrina, 2018.

LIU, D. et al. Discriminating and elimination of damaged soybean seeds based on image characteristics. **Journal Of Stored Products Research**, Amsterdam, v.60, p.67-74, jan. 2015. Elsevier BV. Available from: <https://www.sciencedirect.com/science/article/pii/S0022474X14000940>. Accessed: Jul. 15, 2019. doi: 10.1016/j.jspr.2014.10.001.

MAHAJAN, S. et al. Machine vision based alternative testing approach for physical purity, viability and vigour testing of soybean seeds (Glycine max). **Journal Of Food Science And Technology**, Berlim, v.55, n.10, p.3949-3959, jul. 2018. Springer Science and Business Media LLC. Available from: <https://link.springer.com/article/10.1007/s13197-018-3320-x>. Accessed: Jul. 22, 2019. doi: 10.1007/s13197-018-3320-x.

MONTEIRO, R. C. M. Processamento de imagens para identificação de defeitos no arroz. In: ZUFFO, A. M. (Org.). **A produção do Conhecimento nas Ciências Agrárias e Ambientais**. 4. ed. Ponta Grossa: Atena, 2019. p.298-206. Available from: <https://www.atenaeditora.com.br/arquivos/ebooks/a-producao-do-conhecimento-nas-ciencias-agrarias-e-ambientais-4>. Accessed: Apr. 29, 2019.

PESKE, S. T.; BAUDET LABBÉ, L. M. Beneficiamento de sementes. In: PESKE, S. T.; VILLELA, F. A.; MENEGHELLO, G.

E. **Sementes**: Fundamentos científicos e tecnológicos. 4ª. Edição. Pelotas: Ed. UFPel p.423-479. 2019.

PIAZ, G. D. et al. Uso de classificador óptico para seleção de soja. In: ALMEIDA, A.D. S. et al. (Org.). **Perspectivas da Produção de Sementes no Brasil**. Curitiba: Crv, 2018. p.245-250.

SALIMI Z.; BOELT, B. Classification of processing damage in sugar beet (*Beta vulgaris*) Seeds by Multispectral Image Analysis.

Sensors, Basel, v.19, n.10, p.2360-2369, 22 maio 2019. MDPI AG. Available from: <https://res.mdpi.com/d_attachment/sensors/sensors-19-02360/article_deploy/sensors-19-02360-v2.pdf>. Accessed: Jun. 11, 2019. doi: 10.3390/s1910236.

SANTIAGO, W. E. et al. Classification of oranges by digital image processing. **Acta Iguazu**, Cascavel, v.8, n.2, p.49-58, jun. 2019. Available from: <http://131.255.84.97/index.php/actaiguazu/article/view/19227/14390>. Accessed: Aug. 1, 2019.