

Scientific Electronic Archives

Issue ID: Sci. Elec. Arch. Vol. 14 (11)

November 2021

DOI: <http://dx.doi.org/10.36560/141120211467>

Article link: <https://sea.ufr.edu.br/SEA/article/view/1467>



Mathematical models for metric features extraction from RGB-D sensor

Elton Fernandes dos Santos

Universidade Federal de Mato Grosso

Corresponding author

Laurimar Gonçalves Vendrusculo

Embrapa Informática Agropecuária

laurimar.vendrusculo@embrapa.br

Luciano Bastos Lopes

Embrapa Agrossilvipastoril

Scheila Geiele Kamchen

Universidade Federal de Mato Grosso

Isabella C. F. S. Condotta

University of Illinois

Abstract. The use of the RGB-D camera has been applied in several fields of science. That popularization is due to the emergence of technologies such as the Intel[®] RealSense™ D400 series. However, despite the actual demand from some potential users, few studies concern the characterization of these sensors for object measurements. Our study sought to estimate models dealing with calculating the area and length between targets or points within RGB and depth images. An experiment was set up with white cardboard fixed on a flat surface with colored pins. We measured the distance between the camera and cardboard by calculating the average distance from the pixels belonging to the target area. The Information Criterion AIC and BIC associated with R^2 were performed to select the best models. Polynomial and power regression models reached the highest coefficient of determination and smallest values of AIC and BIC.

Keywords: image processing; depth camera; RealSense™.

Introduction

Even though RGB-D sensors were extensively employed in the past decade, mainly to promote human interaction through video game consoles, their use for object detection and metric measurements in productive activities has been in the early stages. Beyond entertainment use, these sensors were also used to solve object recognition complex problems, including human tracking activity (Wang et al., 2014; Basso et al., 2013; Nguyen et al., 2014). Nowadays, its characteristics of low-cost RGB-D sensor and stable 3D vision technologies

have been applied to detect animal and agricultural targets bringing promising outcomes to activities that are currently time-consuming, labor-intensive, and expensive work. Studies related to livestock production demonstrated a high correlation ($R^2 > 0.9$) between 3D cameras data and manual measurements (Hu et al., 2021, Pezzuolo et al., 2018; Condotta et al., 2018; Kongso, 2014). Most RGB-D information is analyzed through regression or machine learning models to estimate metric measurements, surface area, and volume.

Focusing on this prominent market, Intel® launched the RealSense™ D400 series of RGB-D cameras, providing colored image flow and depth maps. These cameras rely on an infrared light source, which brings improvements in data quality, allowing the most reliable outdoor applications, such as a tropical environment with a high luminous variation (Condotta et al., 2020). Despite its potentials, there is little information about performance and methodological standards for using it so far, which will be crucial for scientific applications. The approach implemented by Vit and Shani et al., 2018 used Euclidean distance in the 3D plane to measure artificial objects' size while Hu et al., 2021 extracted length and width from RGB-D images of non-restrained pigs using primary pixel count.

This work aimed to develop and validate mathematical models to perform area estimation and metric measurements using the depth and color images from an RGB-D camera.

Materials and Methods

The study was carried out at SIGEO Laboratory in Embrapa Agrosilvopastoral, Sinop - Mato Grosso State, Brazil. In order to obtain the relation between the target pixel area and depth images at the RGB-D camera, an experiment was set up: a white cardboard was fixed to a flat surface with colored pins (Fig. 1), perpendicular to the camera. The cardboard had 0.33 m² and was the planar target for further considerations. Pins were attached to form a spatial grid and support horizontal, vertical, and diagonal measurements.

An Intel® RealSense™ D435i camera captured images of the cardboard and its surroundings in the lab illuminated by fluorescent light of approximately 5 watts. The D435i model of RealSense technology was chosen because of its larger field of view (FOV) and superior global shutter performance compared to D415 models. These characteristics deal accordingly with blind spots

reduction and high-speed movements. Furthermore, the configuration of internal settings, including acquisition and post-processing, is supported by a cross-platform and open-source Intel RealSense SDK 2.0 (Cargagni et al., 2019). RGB-D data streams from the camera were stored in a rosbag file using in-house software written in C++. Later, depth and color frames were extracted from the rosbag file in comma-separated values (.csv) and Portable Network Graphics (.png) formats, respectively. The depth array was processed using the OpenCV library and Python language. Finally, the camera was configured following the recommendations of Intel RealSense (2018), composed of images with a resolution of 848 x 480 pixels, medium density, and presettable from the "HighResMedDensityPreset.json" file. Related to better resolution configuration, Study from Vit and Shani, et al., 2018 found that object size identification resulted in minor errors at 848 resolution with depth information while 640 resolution was able to identify with accuracy objects through RGB images. For distance measurements on the white cardboard, 13 reference dots formed 33 measurements equally distributed between vertical, horizontal, and diagonal alignments (Fig. 1a).

The camera was then placed at 17 increasing distances, from 0.9 m to 2.5 m, with a step of 0.1 m. Although this camera has a measuring range of up to 10 m, a maximum distance of 2.5 m was chosen because this range had satisfactory sensitivity to agriculture applications, similar to observed by Condotta et al. (2020). Both RGB and depth images were simultaneously acquired at each distance. Surface cardboard length was measured; however, only the diagonal was considered for distance measurements among predetermined dots for depth images. In this case, the distance between the camera and the cardboard was calculated through the depth image without performing stream alignment. One example of images captured from the RGB and depth sensors are shown in Figure 1.

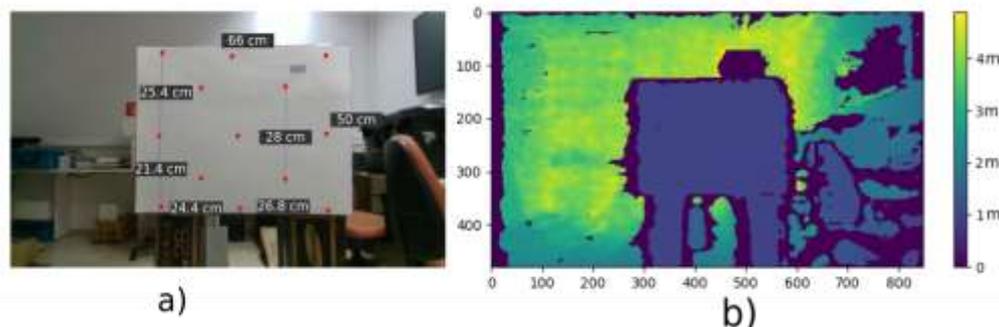


Figure 1. Sample images captured from RGB-D sensor. a) Color image (RGB) showing horizontal and vertical reference dots fixed on the cardboard at 0.9 m from the sensor, b) depth image showing the distance from the object to the sensor on a colored scale.

Mathematical modeling and accuracy assessment

To verify Intel® RealSense™ D435i camera's suitability for metric measurements as the best

mathematical model to correlate digital and manual measurements of distance and area, four distinct mathematical models were compared: linear,

logarithmic, polynomial, and power. For this reason, statistical analysis was performed to assess the accuracy of models where the best fit model should emphasize the robustness criteria and computational simplicity. This study has employed three statistical analyses: coefficient of determination (R^2), Akaike Information Criterion (AIC) (Akaike, 1974), and the Bayesian Information Criterion (BIC) (Stone, 1979). AIC and BIC compare candidate models. These techniques are based on an in-sample fit to estimate a model's log-likelihood and its complexity to predict future values, as shown by equations 4 and 5, respectively. Therefore, the models achieving the lowest values of AIC and BIC are considered the best models. The coefficient of determination (R^2) is a well-known technique that quantifies the proportion of variance explained by a statistical model through measured (A_m) and estimated (A_e) pixels area or distance.

$$R^2 = 1 - \frac{SSE}{SST} \quad (1)$$

$$\sum \text{of squares error}(SSE) = \sum_{i=1}^n (A_m - A_e)^2 \quad (2)$$

$$\sum \text{of squares total}(SST) = \sum_{i=1}^n (A_m - \bar{A}_m)^2 \quad (3)$$

where:

A_m = measured values;

A_e = estimated values of A_m ;

\bar{A}_m = mean of the A_m values;

n = number of images.

$$AIC = 2k - 2 * \quad (4)$$

where:

LL = maximum value of log-likelihood function of the model;

k = number of estimated parameters.

$$BIC = k * \ln(n) - 2 * \quad (5)$$

where:

LL = maximum value of the log-likelihood function of the model;

n = number of recorded measurements or sample size;

k = number of estimated parameters.

where:

LL = maximum value of the log-likelihood function of the model;

n = number of recorded measurements or sample size;

k = number of estimated parameters.

Model validation

For model validation purposes, area and length best-fit equations were applied to assess measured and predicted length, width, and area at the three corrugated paper boxes with a defined size.

We categorized boxes sizes as Large (73x46 cm), Medium (4 x32 cm), and Small (24x22cm). The objects were positioned perpendicularly to the camera and sampled individually over three distances, 1.01 m, 1.54 m, and 2 m.

The difference of the actual boxes' length and boxes' width was computed from the same measurements captured by the sensor using depth and RGB images dimensions. The goal in the validation step was to evaluate RGB-D sensor capability to register sizes on best-fit equations. Lately, a paired t-test was applied to assess significant differences in box surface area and sides measurements.

The individual area was evaluated for each image acquired between the camera sensor's and the studied scenario at increasing distances. The boxes' area variations were evaluated by increasing the distance between the camera and the studied scenario. Similarly, the differences in actual area and acquired by sensor were compared. The area validation was only applied on depth images at three size boxes over different distances.

Results and discussion

Models Development

From 0.9 m to 2.5 m, the cardboard area ratio (m^2/pixel) was recorded and translated from pixels to metric units for each distance range. Figures 2 and 3 show the trend lines and best equations adjusted by linear, logarithmic, polynomial, and power models. Equations 6 and 7 were rewritten because they obtained the highest coefficient of determination associated with model simplicity to calculate area in depth and RGB images. As can be observed from these figures, for both depth and RGB images, power and polynomial models were the best to fit experimental data ($R^2 = 1$, p-value < 0.001)

$$A_m = A_p * 5.48 * 10^{-0.6} * x^{1.93} \quad (6)$$

where:

A_m = area in square meters,

A_p = area of one pixel assuming $A_p = 1$;

x = distance from the sensor to the object varying from 0.8 m to 2.5m.

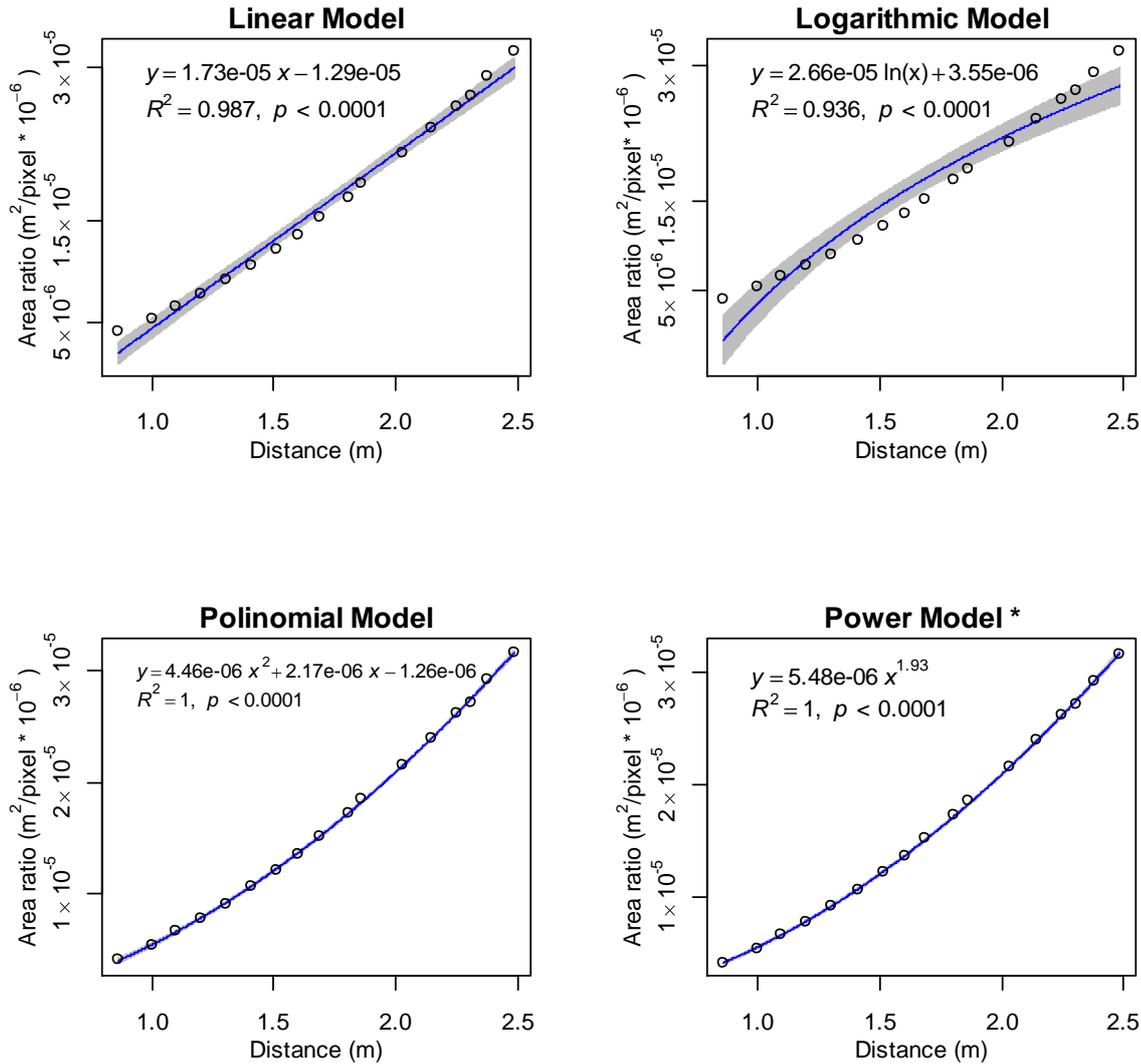


Figure 2. Relationship of area ratio (m^2/px) from RGB image over different distances fitted with different models and distances.

$$A_m = A_p * 5.68 * 10^{-06} * x^{1.89} \quad (7)$$

where:

A_m = area in square meters;

x = distance of the camera to the object;

A_p = area in pixels that should be replaced by 1 and multiplied by the number of pixels associated with a specific distance from the sensor. ($A_p = 1$ in eq. 7), especially when the target has an irregular surface.

Up to 2 meters, there is significant variation of length in both types of images, as observed in figures 4 and 5. In this case, the sensor recorded a few segments with the supposed same length with

slightly different lengths. Also, the difference increases as the sensor is moved away from the target. The errors of vertical and horizontal measurements are likely different, and this difference increases with the distance (Intel RealSense, 2019). The study of Carfagni et al. (2017) corroborates this finding and describes it as systematic errors on the Intel SR300 sensor. In homogeneous distance errors cause both length and area measurement inconsistencies. These errors are mainly due to the sensor's limitations to perceive the target as planar and depth offset. Choo (2015) and Condotta et al. (2020) reported a similar behavior with a Microsoft Kinect[®] sensor. The stochastic nature of the error is mainly affected by the depth of the object in the scene. Furthermore, radial errors cause little distortion, which explains greater variation at depth images compared to colored.

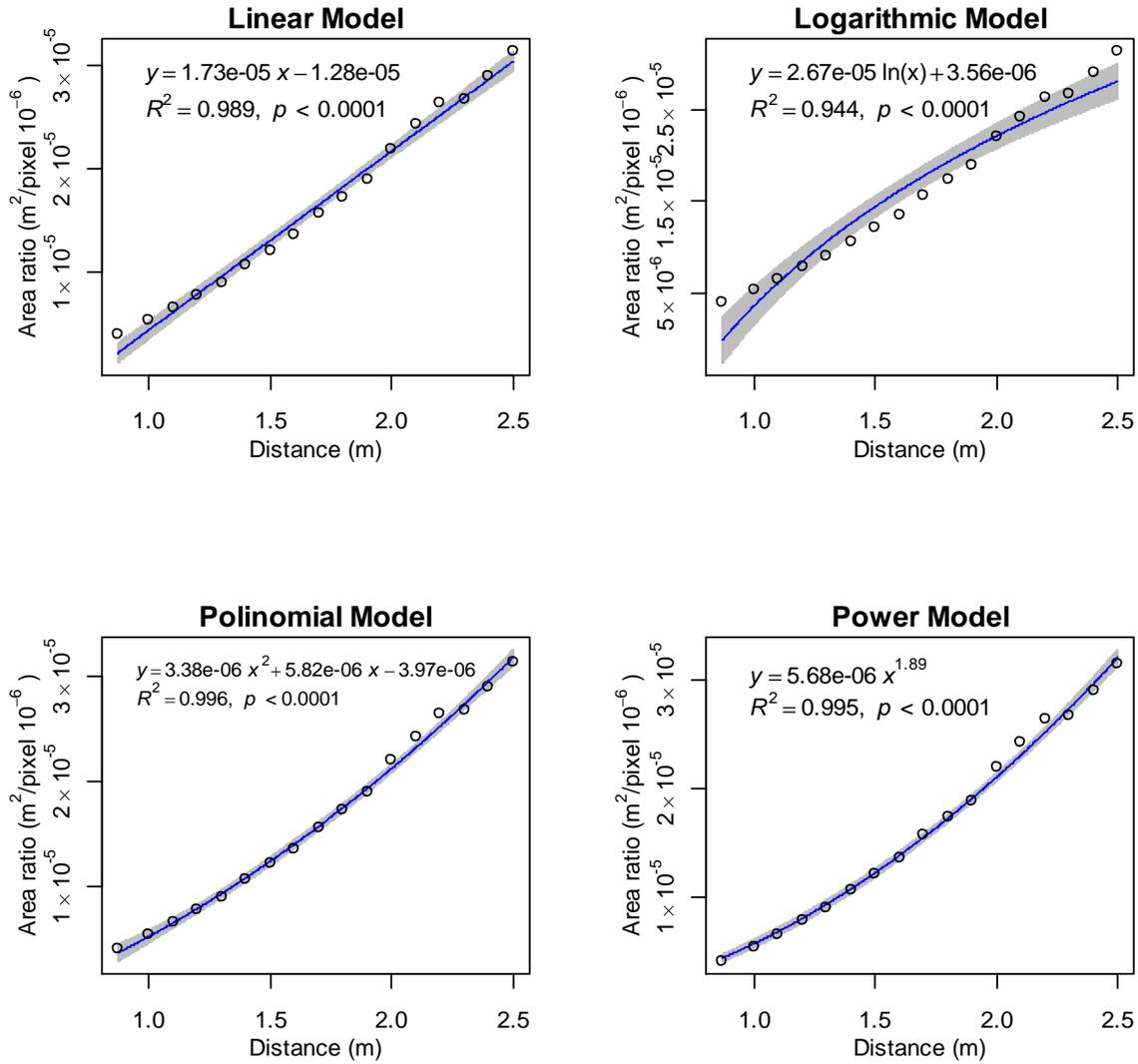


Figure 3. Relationship of area ratio (m²/px) from RGB image over different distances fitted with different models and distances.

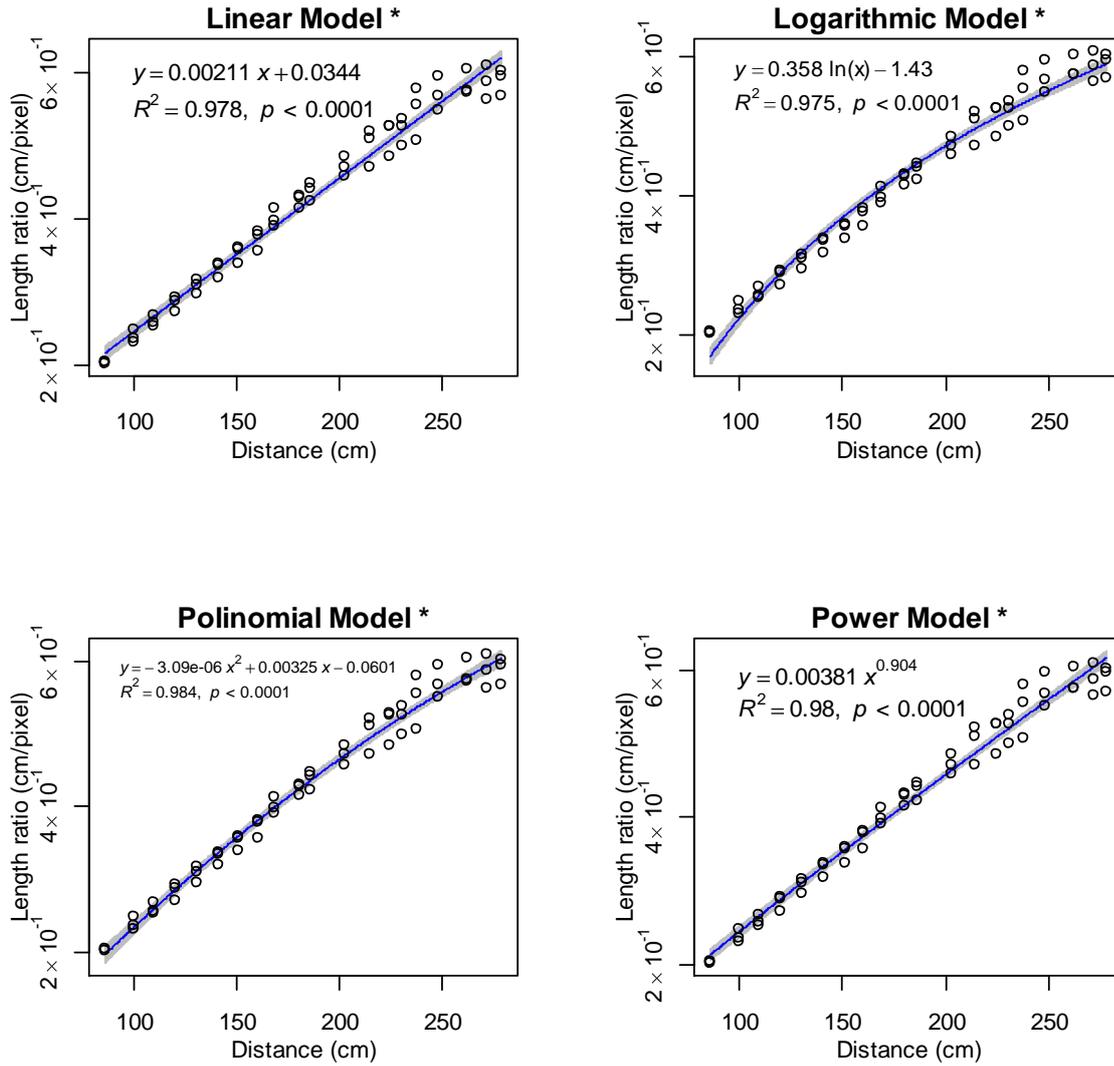


Figure 4. Relationship between length measurements obtained within depth image and distance from the sensor fitted by different models.

$$D_{cm} = D_p * 3.81 * 10^{-03} * x^{0.904} \quad (8)$$

where:

D_{cm} = distance in centimeters within cardboard reference points;

D_p = distance in pixels;

X = distance of the camera to the object.

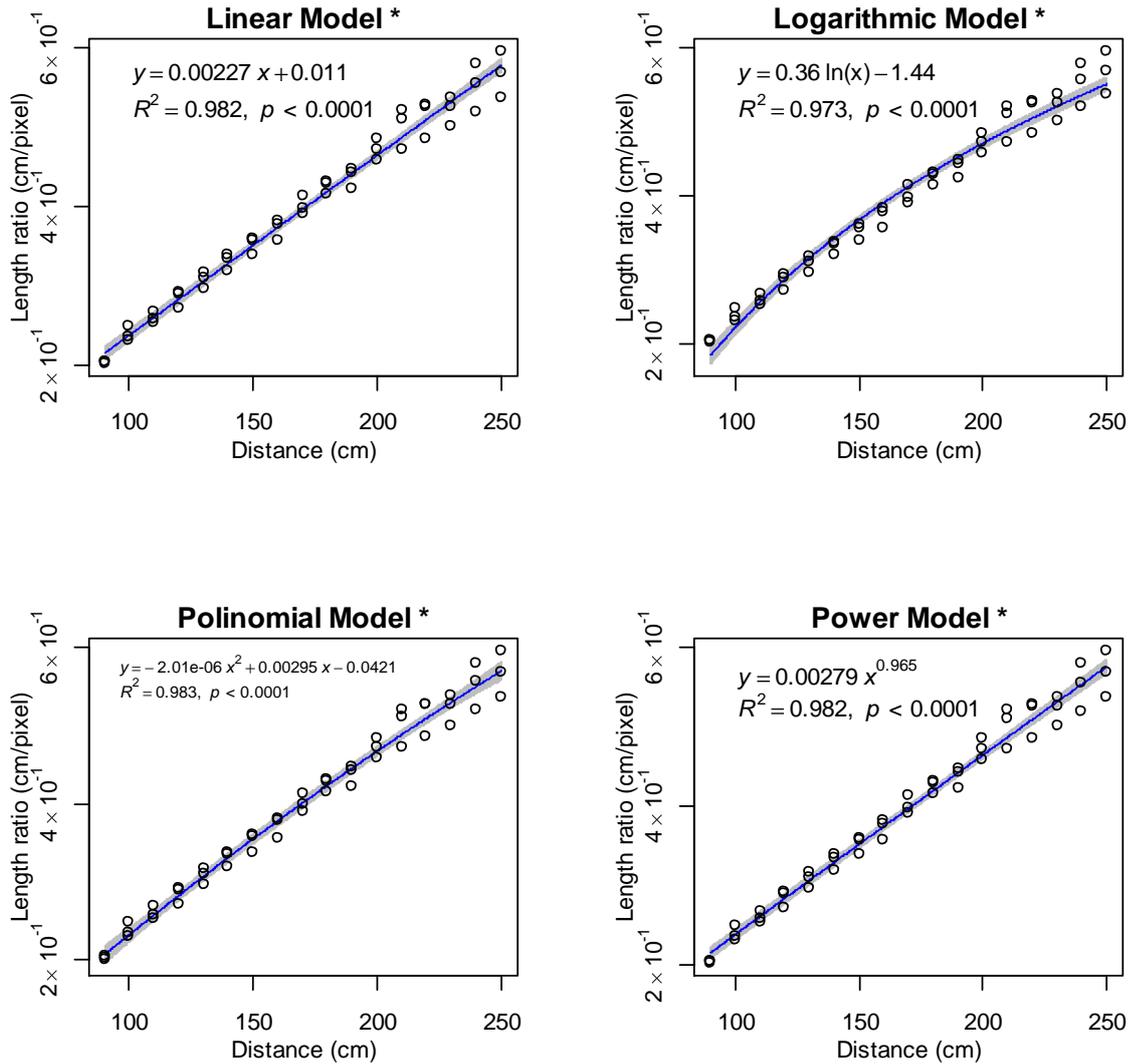


Figure 5. Relationship between length measurements obtained within the color image (RGB) and distance from the sensor -fitted by different models.

$$D_{cm} = D_p * 2.79 * 10^{-3} * Z^{0.965} \quad (9)$$

where:

D_{cm} = distance in centimeters within cardboard reference points;

D_p = distance in pixels;

x = distance of the camera to the object using data from an associated colored image.

Similarly, from Figures 4 and 5 it can be inferred that polynomial and power models obtained a high coefficient of determination ($R^2 > 0.98$).

Applying the criterion that the model with the minimum values of AIC and BIC linked with the highest R^2 was the optimal model chosen, the polynomial approach was the best for both models of area and length from depth or color images, as shown in Table 1. Power model were the second-best option.

Table 1. Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) results of four mathematical models tested for measuring area and length in depth and RGB images. The minimum values of AIC and BIC, which correspond to the most suitable model, are bold.

Measurement/Image Type		Models			
		Linear	Logarithmic	Polynomial	Power
Area/Depth		-416	-389	-481	-480
Area/RGB	AIC	-419	-391	-436	-434
Length/Depth		-304	-295	-320	-309
Length/RGB		-277	-257	-279	-278
Area/Depth		-414	-387	-478	-478
Area/RGB	BIC	-417	-389	-433	-432
Length/Depth		-298	-289	-312	-303
Length/RGB		-271	-251	-271	-272

Models Validation

Paired t-test compared box's length or width measurements from RGB and depth images resulted in no statistically significant difference ($p < 0.05$) in any box size and distance from the sensor. The mean difference at length estimates suggests an increment of 1.62 cm for RGB images compared to depth images. Also, the mean difference in box's width estimations showed a decrement of 1.3 cm for RGB images compared to depth images. This result agrees with the satisfactory performance of equations generated in this study to estimate linear measurements (e.g., length and width) from both image formats.

For validation purposes, we computed the power mathematical regression models (eq. 6, 7, 8, and 9) at three different box's sizes dealing with estimating area (cm²), width and length (cm). The highest errors measurements (Digital estimated

value – Analogic measured value) were found in depth images compared to RGB images values. This means that real box dimensions were over and underestimated in all observed distances (Fig. 6 a). In particular, at a 1.5 m distance from the camera sensor, errors from depth images were closer to zero, except length measurement at the large box. For RGB images, in general, measurements at the closest distance (1 m) had the lowest errors compared to further distances (2m). Also, estimated values were underestimated at closer distances and overestimated at farther distances for small and medium boxes.

Considering the boxes' width, Figure 6b shows that depth and RGB images from small and medium boxes present a lesser difference than the larger box at three distances. Generally, those results agree with Khoshelam (2012) results using a Microsoft Kinect® sensor.

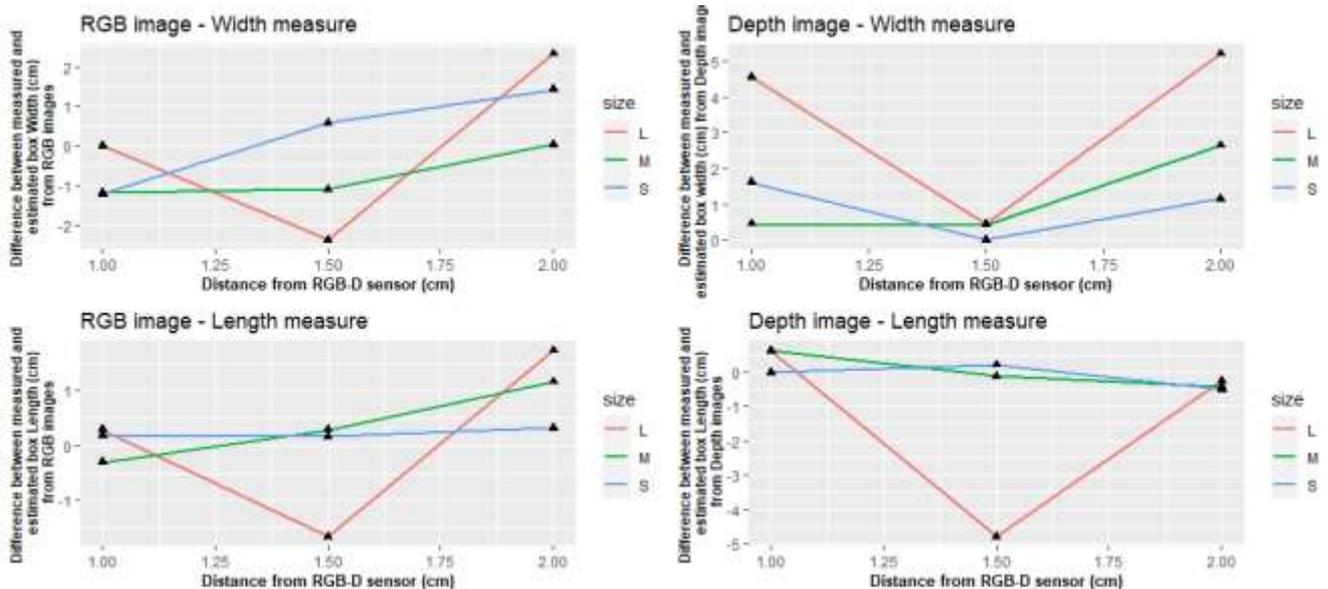


Figure 6. The validation process of distance estimation comparing difference box's width length and, in cm, from Large (L), Medium (M) and, Small (S) boxes obtained from the depth and RGB images using equations 8 and 9, respectively.

Figure 7 shows that the highest absolute area error was found in large box considering all distances and both types of images (RGB image = 238 cm² and Depth image=259 cm²). However, when we calculate the proportional error, taking into account the actual area box as a unit, we found out that, generally, in RGB images, the estimated areas of the three types of boxes were lower than actual areas up to 1.5 m, varying from 0.64 % to 6.7 %. On the other hand, at 2 meters of distance from the

camera, the area values were greater and varied from 3.2 % to 7.4 %. Nevertheless, considering the depth image, the area estimated overall from depth images was greater than the actual boxes' area for each distance. The only exception was a large box distant 1.5 m, where the measured area was underestimated by 10%.

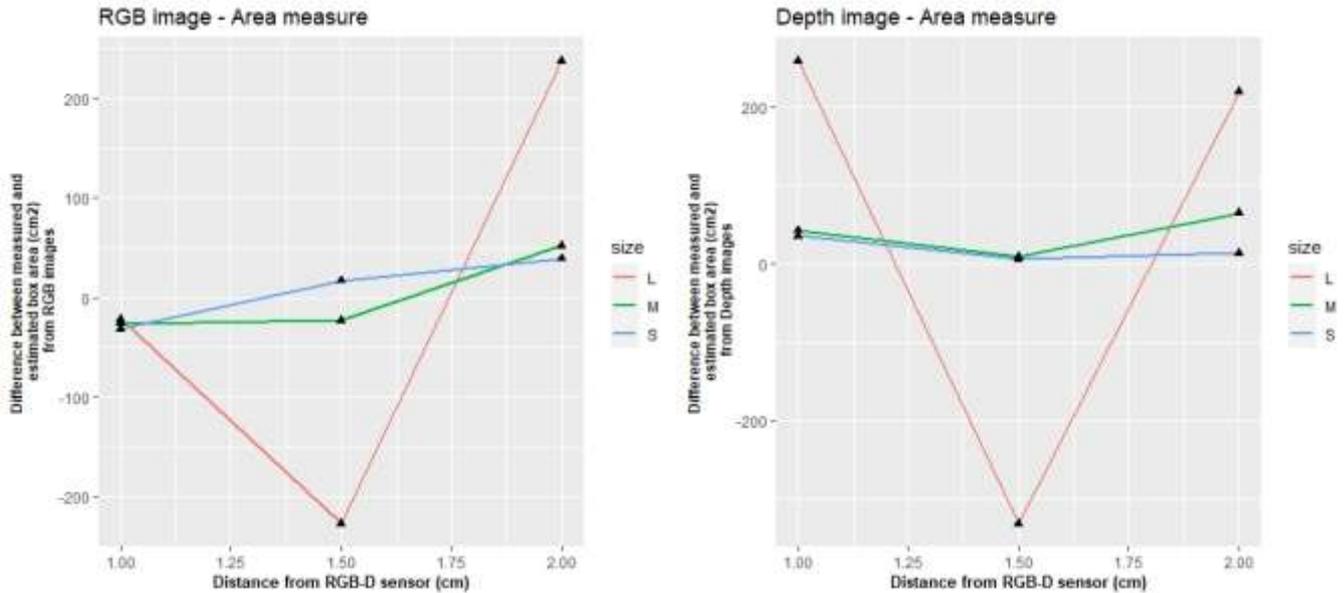


Figure 7. Area error (cm²) between measured paper boxes area in RGB and depth images obtained at different distances with equation 6.

Conclusion

RGB-D sensor is becoming essential to diverse industrial and agricultural applications that need to measure the within area and distance to specific targets. This paper presents best-fit models for area and length using Intel® RealSense™ D345i sensor in RGB and depth images. We estimated mathematical models from recorded, both RGB and depth, images at incremental distances. This study suggested that the polynomial and power regression model obtained the combination of a high coefficient of determination ($R^2 > 0.9$) and minimum values of AIC and BIC. However, considering model simplicity, We might say the power regression model has fewer computational demands, which in the on-the-fly automatic application can be the best option. Those models are either recommended in Condotta et al. (2020) study. The validation step has shown that objects with large areas, despite their distance from RGB-D sensor, demonstrated larger errors than actual areas. Until 1.5 m the area on both RBG and depth images was underestimated, and at 2 meters from the sensor, they were overestimated. The area at medium and small targets had fewer errors than large objects, reaching 64 cm² of error. The main limitations of this experimental setup are: i) the small scale of the area target (0.33 m²), which for targets much greater than can be challenging; (ii) the

restricted range of the sensor (0.8 - 2.5 m). We will plan to tackle these constraints in future studies and targets with other complex morphological features.

Acknowledgment

Embrapa and Fundo Amazônia Project funded this research. We want to thank technical support from Embrapa Agrosilvopastoral.

References

BANGLI, L.; HAIBIN, C.; ZHAOJIE, J.; HONGHAI, L. Bangli. L.; Haibin. C.; Zhaojie. J.; Honghai. L. RGB-D sensing-based human action and interaction analysis: a survey. *Pattern Recognition*, vol., 94, p. 1–12, 2019.

BASSO, F. MUNARO, M.; MICHIELETTO, S. Basso, F.; Munaro, M.; Michieletto, S.; *et al.* Fast and robust multi-people tracking from RGB-D data for a mobile robot. *Advances in Intelligent Systems and Computing*, vol. 193, p. 265–276, 2013.

CARFAGNI, M.; FURFERI, R.; GOVERNI, L.; SERVI, M.; UCCHEDDUU, F.; VOLPE, Y. On the Performance of the Intel SR300 Depth Camera: Metrological and Critical Characterization. In: *IEEE Sensors Journal*, vol. 17, p. 4508-4519, 2017. doi: 10.1109/JSEN.2017.2703829.

- CHOO, B.; DeVORE, M.D.; BELING, P.A. Statistical models of horizontal and vertical stochastic noise for the Microsoft Kinect™ sensor. In: IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, 2014, p. 2624-2630, doi: 10.1109/IECON.2014.7048876.
- CONDOTTA, I.C.F.S.; BROWN-BRANDL, T.M.; SILVA-MIRANDA, K.O. Evaluation of using a depth sensor to estimate the weight of finishing pigs. In: *8th European Conference on Precision Livestock Farming*, 2017, Nantes, 2017. vol. 1. p. 495-502.
- CONDOTTA, I.C.F.S.; BROWN-BRANDL, T.M.; SOUZA, R.V.; SILVA-MIRANDA, K.O. Using an artificial neural network to predict pig mass from depth images. In: *10th International Livestock Environment Symposium (ILES X)*, 2018, Omaha, NE. 2018.
- CONDOTTA, I.C.F.S.; BROWN-BRANDL, T.M.; SOUZA, R.V.; SILVA-MIRANDA, K.O.; STINN, J.P. Evaluation of a depth sensor for mass estimation of growing and finishing pigs. *Biosyst. Eng.*, vol. 173, p. 11–18, 2018a. doi: 10.1016/j.biosystemseng.2018.03.002.
- CONDOTTA, I.C.F.S.; BROWN-BRANDL, T.M.; PITLA, S.K.; STINN, J.P.; SILVA-MIRANDA, K.O. Evaluation of low-cost depth cameras for agricultural applications. *Computers and Electronics in Agriculture*, vol.173, 2020
- Intel® Realsense™. User guide Intel Realsense D400 series/SR300 viewer. Revision 002, (2018).
- Intel® RealSense™. Lib. Realsense: D400 Series visual presets. 2.0 [San Francisco]: GitHub, 2019. 1 Phone code library. Available in: < <https://github.com/IntelRealSense/librealsense/tree/master/examples/align> < Access in: 02 mar 2020.
- KHOSHELHAM, K.; ELBERINK, S.O. Accuracy and resolution of Kinect depth data for indoor mapping applications. *Sensors*, vol. 12, p. 1437-1454, 2012.
- KONGSO, J. Estimation of pig weight using a Microsoft Kinect prototype imaging system. *Computers and Electronics in Agriculture*, vol. 39, p. 32–35, 2014.
- LAI, K.; BO, L.; REN, X.; FOX, D. A large-scale hierarchical multi-view RGB-D object dataset. In: *IEEE International Conference on Robotics and Automation*, Shanghai, 2011, p. 1817-1824. doi: 10.1109/ICRA.2011.5980382.
- NGUYEN, T.V.; FENG, J.; YAN, S. Seeing human weight from a single RGB-D image. *Journal of Computer Science and Technology*, vol. 29, p. 777–784, 2014.
- PEZZUOLO, A.; GUARINO, M.; SARTORI, L.; GONZÁLEZ, L.A. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. *Computers and Electronics in Agriculture*, vol. 148, p. 29-36, 2018.
- VIT, A.; SHANI, G. Comparing RGB-D Sensors for Close Range Outdoor Agricultural Phenotyping. *Sensors*. 2018, 18, 4413.
- YU, H.; LEE, K.; MOROTA, G. Forecasting dynamic body weight of non-restrained pigs from images using an RGB-D sensor camera. *Translational Animal Science*. 2021, vol.5:1, txab006, <https://doi.org/10.1093/tas/txab006>
- WANG, S.; PAN, H.; ZHANG, C.; TIAN, Y. RGB-D image-based detection of stairs, pedestrian crosswalks and traffic signs. *J. Vis. Commun. Image Representation*, vol. 25, p. 263-272, 2014.