Biological Systems Engineering: Papers and Publications

Biological Systems Engineering

4-27-2020

# Evaluation of low-cost depth cameras for agricultural applications

Isabella C.F.S. Condotta

Tami M. Brown-Bandl
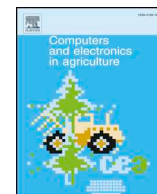
Santosh Pitla

John P. Stinn

Késia O. Silva-Miranda

# Evaluation of low-cost depth cameras for agricultural applications

Isabella C.F.S. Condotta[a,b,1], Tami M. Brown-Brandl[b,c,*], Santosh K. Pitla[b], John P. Stinn[d], Késia O. Silva-Miranda[a]

[a] Biosystems Engineering Department, São Paulo University, "Luiz de Queiroz" College of Agriculture, Av. Pádua Dias 11, Piracicaba, São Paulo 13418-900, Brazil
[b] Biological Systems Engineering Department, University of Nebraska-Lincoln, 3605 Fair St Lincoln, NE 68503, United States
[c] USDA, Agricultural Research Service, U.S. Meat Animal Research Center, P.O. 166 Clay Center, NE 68933 United States
[d] Iowa Select Farms, 811 South Oak, Iowa Falls, IA 50126, United States

ABSTRACT

Low-cost depth-cameras have been used in many agricultural applications with reported advantages of low cost, reliability and speed of measurement. However, some problems were also reported and seem to be technology-related, so understanding the limitations of each type of depth camera technology could provide a basis for technology selection and the development of research involving its use. The cameras use one or a combination of two of the three available technologies: structured light, time-of-flight (ToF), and stereoscopy. The objectives were to evaluate these different technologies for depth sensing, including measuring accuracy and repeatability of distance data and measurements at different positions within the image, and cameras usefulness in indoor and outdoor settings. Then, cameras were tested in a swine facility and in a corn field. Five different cameras were used: (1) Microsoft Kinect v.1, (2) Microsoft Kinect v.2, (3) Intel® RealSense™ Depth Camera D435, (4) ZED Stereo Camera (StereoLabs), and (5) CamBoard Pico Flexx (PMD Technologies). Results indicate that there were significant camera to camera differences for ZED Stereo Camera and Kinect v.1 camera (p < 0.05). All cameras showed an increase in the standard deviation as the distance between camera and object increased; however, the Intel RealSense camera had a larger increase. Time-of-flight cameras had the smallest error between different sizes of objects. Time-of-flight cameras had non-readable zones on the corners of the images. The results indicate that the ToF technology is the best to be used for indoor applications and stereoscopy is the best technology for outdoor applications.

## 1. Introduction

Low-cost depth cameras have been used as an alternative to expensive laser scanners (Khoshelham & Elberink, 2012) in various areas, such as mapping and 3D reconstruction (Izadi et al., 2011), indoor robotics (Benavidez and Jamshidi, 2011; Correa et al., 2012; Ganganath and Leung, 2012), objects' detection and recognition (Hernandez-Lopez et al., 2012), and gesture recognition (Chang et al., 2011a, 2011b). Most of these research areas were originally approached using standard digital image processing and analysis, but problems such as lighting, color distinction and excess of equipment; led to the use of depth cameras for image acquisition.

There are several technologies used for depth acquisition. Stereoscopy (SC) was the first one used to acquire information on the objects' geometry. The structured light technology (SL) was introduced to overcome some of the problems with the stereoscopy and, after that,

the time-of flight (ToF) cameras started being used.

The depth images provided by these cameras are composed of a numeric map containing the distances, most commonly in metric units, between the sensor and each pixel that makes up the image. Khoshelham and Elberink (2012) examined the accuracy and precision of the depth data provided by a structure-light sensor (Microsoft Kinect v.1) and provided an explanation of various errors. The random error of measurements increased with distance between sensor and object and fluctuated between a few millimeters to 4 cm. In addition, the errors in the distance data originated from three sources: (1) calibration errors, (2) configuration of the measuring area (improper lighting or image geometry), and (3) smooth or bright surfaces. While lighting is generally not a problem with depth cameras for indoor applications, intense lighting can generate low contrast in the infrared image and, therefore, result in gaps on the depth image. These gaps can also occur when the distance from the object to the sensor is outside of the

operating range of the camera, or when the orientation of the object surface is such that the emitter does not illuminate some regions, or the camera fails to capture information. Surfaces that are too bright or smooth are very reflective and can also prevent measurement.

Generally, the depth cameras' error is low (Dutta, 2012); however, the standard deviation of the distance data increases with increasing distance between sensor and object, and is greater on the corners of the image. In addition, depth data has been reported to be unusable or inaccurate on the object's edges because, in these areas, the depth map is obtained through interpolation of the projections of the reflected infrared light on two different regions, the edge and the background (Gottfried et al., 2011). For studies that use more than one depth camera there is a need to understand the need for calibration between cameras of the same type.

Depth cameras have been used in a variety of agricultural applications (Condotta et al., 2018a; Guo et al., 2017; Hao et al., 2014; Kongsro, 2014; Kulikov et al., 2014; Lao et al., 2016; Lee et al., 2016; Stavrakakis et al., 2015; Wang et al., 2018; Zhu et al., 2015), with reported advantages of low cost, reliability and speed of measurement. However, some problems were also reported, such as occlusion; difficulty of acquiring reliable data on environments with excess light, preventing its use for outdoor applications; shape distortion; motion blur; and noise. Some of these issues seem to be technology-related, so understanding the limitations of each type of depth camera technology could provide a basis for technology selection and the development of research involving its use.

### 1.1. Technologies' principles

There is a variety of commercial depth cameras that have being used in agricultural applications (Table 1). Currently, Kinect (v.1 and v.2) and Xtion PRO Live are not available on market. Technology used by these depth cameras can be divided into three different principles: stereoscopy (SC), structured light (SL), and Time-of-flight (ToF).

To form the depth image, stereoscopy cameras use two RGB cameras to acquire images from the same scene at slightly different positions and the 3-dimensional position of a point on the scene is calculated by triangulation between correspondent points on both images (Berkovic & Shafir, 2012; Keselman et al., 2017). Structured light cameras form the depth image by using the IR emitter to project a beam of light that is divided into multiple beams when reflected on the objects, forming a pattern of points which are captured by the IR camera. This pattern is compared with a standard of predetermined distance from the camera. The distance from each pixel to the camera is calculated by triangulation (Andersen et al., 2012; Berkovic and Shafir, 2012; Sarbolandi et al., 2015; Zhang, 2018). The time-of-flight (ToF) cameras use a technology that is based on measuring the time that the light emitted by the IR emitter requires to travel to a scene and back to the camera. The IR light is, then, captured by the IR camera, and, knowing the fixed speed of light, it is possible to calculate the distance traveled (Sarbolandi et al., 2015).

### 1.2. Objectives

The objective of this study was to evaluate different technologies for depth sensing in agricultural applications, including (1) measure the accuracy and repeatability of distance data, (2) measure the accuracy and repeatability of dimensions data with respect to positions within the image, size of objects, and distance from the camera, (3) determine the maximum and minimum useful distance and the associated errors for each camera in indoor and outdoor settings, (4) test each camera to evaluate its usefulness in collecting plant and animal phenotypic data.

## 2. Materials and methods

Five different depth sensing cameras were tested: (1) Microsoft Kinect v.1, (2) Microsoft Kinect v.2, (3) Intel® RealSense™ Depth Camera D435, (4) ZED Stereo Camera (StereoLabs), and (5) CamBoard Pico Flexx (PMD Technologies) (Fig. 1). Each of these cameras represent one of three different technologies, or a combination of technologies (ToF, SC combined with SL, SL, ToF and SC, respectively). Two ToF cameras were tested because Microsoft Kinect v.2 has being discontinued but has been one of the most used depth cameras in agricultural research, so another commercially available ToF camera was also evaluated (CamBoard Pico Flexx) along with Kinect v.2.
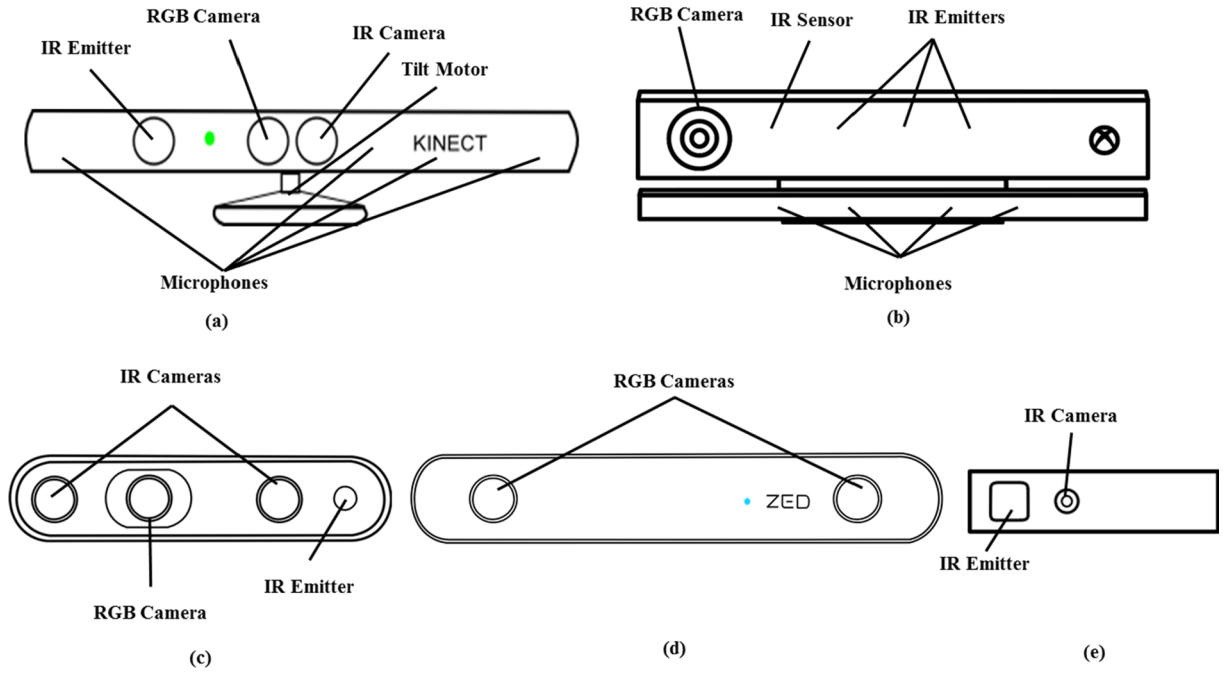
Image acquisition and image processing programs were developed. For image acquisition, different programing environments were used for each camera. For CamBoard Pico Flexx and the Kinect cameras v.1 and v.2, a numerical computing software (MATLAB, R2018a) on a Windows computer was used. For Intel® RealSense™ camera, a C++ program was developed on an UP-Core board with an Ubuntu kernel using Intel® RealSense™ SDK, and, for ZED camera, a C++ program was developed on a NVIDIA Jetson TX2 board, also with an Ubuntu kernel, using ZED SDK. All the images were processed using the same image processing program to ensure consistency of analysis; this program was developed using MATLAB, R2018a.

### 2.1. Distance accuracy and repeatability

A total of fifteen depth cameras, three of each type, were tested. Each group of three sensors of the same type was compared amongst each other. Different cameras were not compared against each other. Each camera was used to collect images of a wall at five distances (from 1.0 to 3.0 m, every 0.5 m) (Fig. 2). Five depth images and five RGB images (except for Pico Flexx, that doesn't have an RGB camera) from each camera were collected.

Depth images were processed with an algorithm developed to extract a fixed area of $11 \times 11$ pixels at the center of the wall. These points were recorded and, then, the average and the standard deviation were calculated.

To evaluate the repeatability of the depth data provided by different cameras of the same type, a multiple linear regression model was developed in Excel® software, using *dummy* variables (Draper and Smith, 2014) to include the effects of all three the cameras of each type in the

**Table 1**
Comparison of commercial depth cameras currently used in agricultural research. SL = structured light, ToF = time-of-flight, SC = stereoscopy.

| Camera | Principle | Measuring Range (m) | Depth Resolution | RGB Max Resolution | Frame Rate (FPS) | FoV Depth (H × V) | Price (US$) |
|---|---|---|---|---|---|---|---|
| Kinect v.1 | SL | 0.4–3.5 | 640 × 480 | 640 × 480 | 15/30 | 57° × 43° | 250 |
| Xtion PRO Live | SL | 0.8–3.5 | 640 × 480 | 1280 × 1024 | 30/60 | 58° × 45° | 140 |
| Xtion 2 | SL | 0.8–3.5 | 640 × 480 | 2592 × 1944 | 30 | 74° × 52° | 236 |
| Kinect v.2 | ToF | 0.5–4.5 | 512 × 424 | 1920 × 1080 | 15/30 | 70° × 60° | 140 |
| CamBoard Pico Flexx | ToF | 0.1–4.0 | 224 × 171 | – | 45 | 62° × 45° | 390 |
| CamBoard Pico Monstar | ToF | 0.5–6.0 | 352 × 287 | – | 60 | 100° × 85° | 1930 |
| ZED | SC | 0.5–20 | 2208 × 621 | 2208 × 621 | 15/30/60/100 | 90° × 60° | 450 |
| RealSense D415 | SL + SC | 0.16–10 | 1280 × 720 | 1920 × 1080 | 90 | 63° × 40° | 150 |
| RealSense D435 | SL + SC | 0.11–10 | 1280 × 720 | 1920 × 1080 | 90 | 85.2° × 58° | 180 |

**Fig. 1.** Components of commercial depth cameras being used in this study (out of scale). (a) Microsoft Kinect v.1, (b) Microsoft Kinect v.2, (c) Intel® RealSense™ Depth Camera D435, (d) ZED Stereo camera (StereoLabs) and (e) CamBoard Pico Flexx (PMD Technologies).

equation. Zeros and ones are attributed for the two dummy variables (IsCamera1 and IsCamera2) generated. By analyzing the three possible combinations, all three cameras are included on the regression, like illustrated bellow:

If isCamera1 = 1 and IsCamera2 = 0, then Camera = 1
If isCamera1 = 0 and IsCamera2 = 1, then Camera = 2
If isCamera1 = 0 and isCamera2 = 0, then Camera = 3

This model was compared with a reduced model (simple linear regression between actual distances and camera-acquired distances) that did not include the individual effects of each camera.

This comparison was made by using the Efroymson's algorithm ("stepwise" regression) (Efroymson, 1960) for comparing two regression models, with null hypothesis given the reduced model equivalent to the global model and with alternative hypothesis, considering the models non-equivalent. The test statistic is given in Eq. (1).
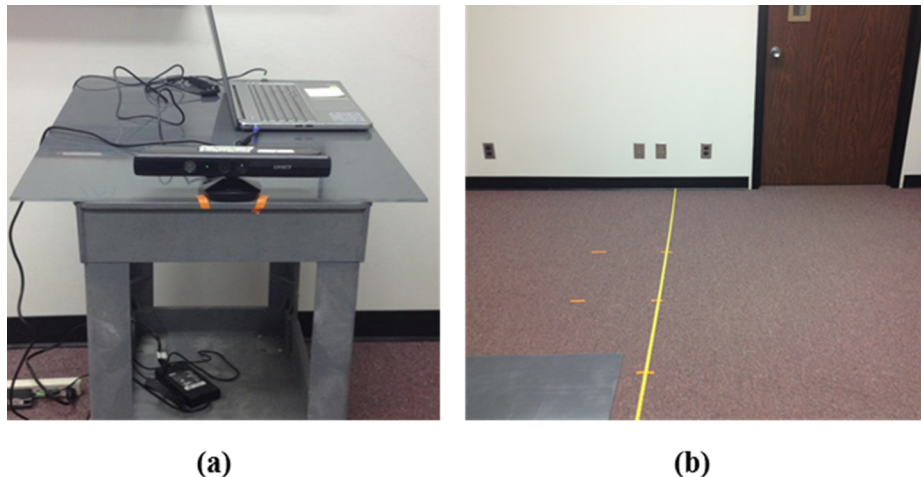
$$F(n, d) = \frac{(SS_r - SS_g)/(DF_r - DF_g)}{SS_g/DF_g}$$

(1)

where

$SS_r$ = sum of the squares of the residue of the reduced model;
$SS_g$ = sum of the squares of the residue of the global model;
$DF_r$ = degrees of freedom of the residue of the global model;
$DF_g$ = degrees of freedom of the residue of the reduced model.

### 2.2. Dimension accuracy and repeatability

Three cameras of each type were used, totaling fifteen cameras. Three sizes of poster board squares (10 × 10, 20 × 20, and 30 × 30 cm) were recorded at five distances (1.0–3.0 m; every 0.5 m) and four different positions on the image (center, edge on the horizontal axis, edge on the vertical axis, and corner; Fig. 3). A total of five depth and RGB images (except CamBoard Pico Flexx) from each camera were



**Fig. 2.** Positioning of the components of the experiment: (a) table with depth cameras and computer for data acquisition and (b) markings on the floor (0.5 m from each other) to capture images of the wall at different distances (from 0.5 m to 3.0 m, every 0.5 m).
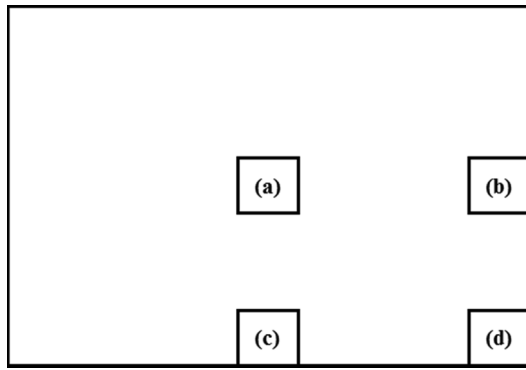
**Fig. 3.** Positions used for images acquisition: (a) center of image, (b) edge on the horizontal axis of image, (c) edge on the vertical axis of image and (d) corner of image.

collected.

The data was analyzed to obtain three different parameters: length, area, and projected volume of the poster board squares. The length of the square (in pixels) was obtained both for the RGB and the depth images. For the RGB images, a manual measure process was performed, using the Image Viewer application from MATLAB R2018a. An algorithm for pre-processing the depth image was developed in order to eliminate the background and keep only the foam board square. This was done by automatic selecting a region of interest on the image and applying a threshold based on distances.

After the pre-processing step, the dimensions of the foam board square were automatically acquired. To measure both the length and the area of the square, the image was first transformed into binary. The maximum sum of the columns on the image was collected as being the length in pixels and, then, the total sum of the image was acquired, obtaining the area of the square, in pixels.

To measure the projected volume (cube) of the square being analyzed (e.g. a cube of 10 × 10 × 10 cm for the 10 × 10 cm square), the pre-processed image was used to obtain values of the third dimension of the square. This was performed by subtracting the image from the theoretical distance from the base of the cube to the camera (Fig. 4). After obtaining the values of depth for each pixel of the theoretical cube, these values were added to obtain the volume of the cube, in pixel cm.

The length ratio, in pixel cm$^{-1}$, was calculated dividing the length, in pixels, obtained for both RGB (when available) and depth images, by the actual length of the foam board square (either 10, 20 or 30 cm). Furthermore, the area ratio, in px cm$^{-2}$, was also calculated by dividing the area obtained on the depth image, in pixels, by the actual area of the

squares (either 100, 400 or 900 cm$^2$).

These ratios were analyzed using the General Linear Procedure (proc GLM) of SAS software, testing the effects of the use of different positions (center, edge on the horizontal axis of the image, edge on the vertical axis of the image, and corner) and different sizes (10 × 10 cm, 20 × 20 cm and 30 × 30 cm) of foam board squares used. Then, regression models were generated for length ratio (px cm$^{-1}$) versus distance, and for area ratio (px cm$^{-2}$) versus distance. The models for length ratio (px cm$^{-1}$) were compared with equations generated from each of the cameras' specifications for field of view (FoV) and resolution, Eq. (2).

$$\frac{l_{px}}{l_{cm}} = 2 \times Res_{px} \times \cot\left(\frac{FoV}{2}\right) \times Z^{-1} \qquad (2)$$

where

$l_{px}$ = length, in pixels;
$l_{cm}$ = length, in cm;
$Res_{px}$ = resolution, in pixel – either vertical or horizontal;
$FoV$ = field of view, in degrees – either vertical or horizontal;
$Z$ = distance between camera and scene, in cm.

Unit transformation equations were assessed in order to eliminate the need for the presence of an object with predetermined size to acquire dimensions on an image, as has been used by several authors (Phillips and Dawson, 1936; Zaragoza, 2009). To obtain these equations, the metric unit (either cm or cm$^2$) from the regression models developed were isolated.

### 2.3. Maximum useful distance

With one of each type of camera tested, five depth and RGB (except CamBoard Pico Flexx) images of foam board squares were collected at each tested distance, ranging from 1.0 to 20.0 m, every 1.0 m, both indoors and outdoors. For distances 1.0–10.0 m, a 30 × 30 cm square was used, while a 60 × 60 cm square was used from 11.0 to 20.0 m. The squares were placed at the center of the image with a tripod.

The foam board square on each image was selected using the same algorithm used on the second experiment. The values of distance from camera to square were averaged. The residuals between actual distance and camera-provided distance were calculated and a regression model of residuals versus actual distance was calculated when possible (e.g. when the square was within the acquisition range of the camera). The maximum distance of useful data (image with no missing regions and reliable distance date) for each camera was recorded.

### 2.4. Minimum useful distance

Similarly, five depth images of foam board squares were collected at each tested distance, ranging from 0.1 m to 1.0 m, every 0.1 m. For distances 0.1 m to 0.5 m, a 5 × 5 cm square was used, while a 10 × 10 cm square was used from 0.6 to 1.0 m. The squares were placed at the center of the image with a small tripod. When camera had a close-range mode available, the same was selected for these measurements. Images were processed in the same way used for determining the maximum useful distance. Finally, the minimum distance of useful data (image with no missing regions and reliable distance date) for each camera was recorded.

### 2.5. Field test of cameras

#### 2.5.1. Animal phenotypic data evaluation

The livestock experiment was conducted in a grow-finish building at the USDA-ARS Meat Animal Research Center (USMARC) in Clay Center, Nebraska (40°31′20″N 98°3′18″W). Digital RGB color images, depth images, and masses were collected on a population of grow-finish pigs
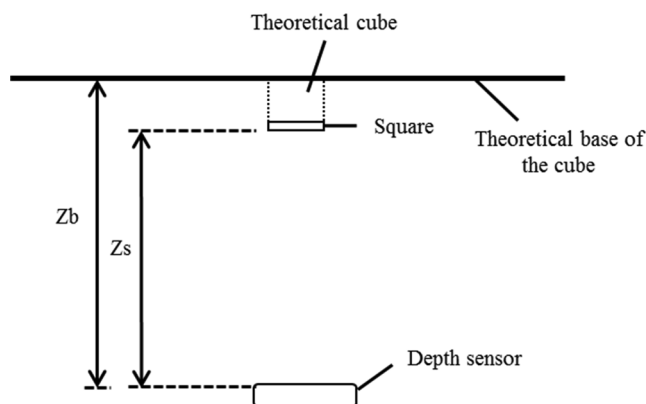


**Fig. 4.** Top-view diagram of the positioning of the depth cameras and the object being analyzed for the calculation of the volume of the theoretical cube with the same size of the square being analyzed.

**(a)**                                        **(b)**

**Fig. 5.** Mass and images were captured on individual pigs using a standard swine scale (b) and five different depth cameras. The cameras were mounted to the wall directly above the center of the scale (a).

at three different time-points.

Top-view images were acquired of fifteen grow-finish pigs randomly sampled at three time-points, five different animals at each time. Animals were approximately 8, 12, and 16 weeks old. A balance of barrows and gilts (Landrace and Yorkshire cross) during each measurement period was evaluated. A total of five depth and RGB (when available) images from each type of camera was collected. The cameras were mounted on a bracket above the scale and images were collected as the animals were being weighed (Fig. 5). Weights and pigs' identification numbers were manually recorded.

The images were processed and analyzed using methods proposed by Condotta et al. (2018a) and Condotta et al. (2018b) to acquire the pig's projected volume (px cm) (Fig. 6). The unit was transformed (from px cm to $cm^3$) using equations developed on the second experiment. Linear regressions of mass (kg) versus volume ($cm^3$) were generated for each of the cameras used.

### 2.5.2. Camera evaluation in simulated and crop field environments

The plant experiment was conducted at an 8093 $m^2$ (2 acres) set of maize yield test plots in a research field on the east campus of the University of Nebraska-Lincoln, in Lincoln, Nebraska (40°48′24″N and 96°40′54″W).

All five cameras tested were evaluated for crop sensing applications aimed at autonomous vehicle operation or plant phenotyping applications. The cameras were tested in an experimental corn field planted on June 10, 2019, with scheduled harvesting date in October. Fig. 7 shows the corn field setup, with corn row spacing at approximately 80 cm (30 in.) and 15 cm (6 in.) between plants. Both RGB and depth images were collected for all five cameras (except Pico Flexx that does not provide RGB images). Distance, in pixels, between the crop rows on both RGB and depth images, was determined using the Image Viewer application
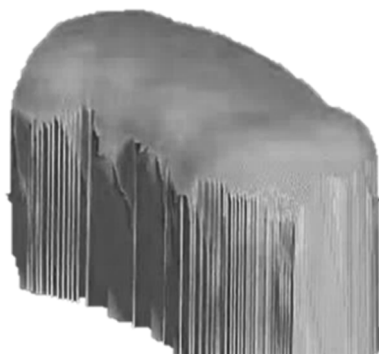


**Fig. 6.** Projected volume of a pig.

of MATLAB R2018a to manually select two points, one on the left-side corn row and one on the right-side corn row. A red tape was used to mark plants at 1.0 m distance from the cameras on both left and right rows for validation purposes. The actual spacing between the marked plants was manually measured to be 75 cm). After that, unit transformation of the measurements, from pixels to meters, was performed. Absolute percentage residuals were calculated between actual and image-acquired spacing in between rows.

### 3. Results and discussion

#### 3.1. Distance accuracy and repeatability

The average pixel values of distances obtained for the region of the wall in the picture, as well as their average standard deviations, for three cameras of each type tested, are shown in Table 2. The result of the Efroymson's algorithm showed that the behavior of different cameras of the same type is the same ($p < 0.05$) for cameras that use the time-of-flight (ToF) principle, CamBoard Pico Flexx (P = 0.18710) and Microsoft Kinect v.2 (P = 0.70697), but different for the other cameras, Intel® RealSense™ D435 (P = 0.02538), Microsoft Kinect v.1 (P = 0.00002), and ZED Stereo Camera (P = 0.00007). From Table 2 it can be noticed that there is an increase in the standard deviation of pixels with increasing distance from camera to wall, corroborating with the data obtained by Khoshelham & Elberink (2012), that, when developing a theoretical random error model for depth data, found that the random error of depth measurements increased with increasing distance to the sensor, ranging from a few millimeters to 4 cm.

These results indicate that Intel RealSense cameras had a higher average standard deviation of pixel values when compared with the other cameras. This is probably due to the lack of consistency between cameras, given that the standard errors for the three cameras analyzed were 0.000, 0.002, and 0.004 at 1.00 m; 0.000, 0.000, and 0.037 at 1.50 m; 0.000, 0.022, and 0.044 at 2.00 m; 0.044, 0.000, and 0.062 at 2.50 m; and 0.024, 0.038, and 0.046 at 3.00 m, with no visible best camera.

#### 3.2. Dimension accuracy and repeatability

Unit transformation equations were generated from regression models between the area (px $cm^{-2}$) and length (px $cm^{-1}$) ratios and the distance (m) from the camera to the foam board square. To transform px to cm, the equations are on the form shown by Eq. (3), and, to transform px to $cm^2$, the equations are on the form shown by Eq. (4). Equations for transformation from depth images and from color images (when available) were also generated. Tables 3 and 4 contain the

5
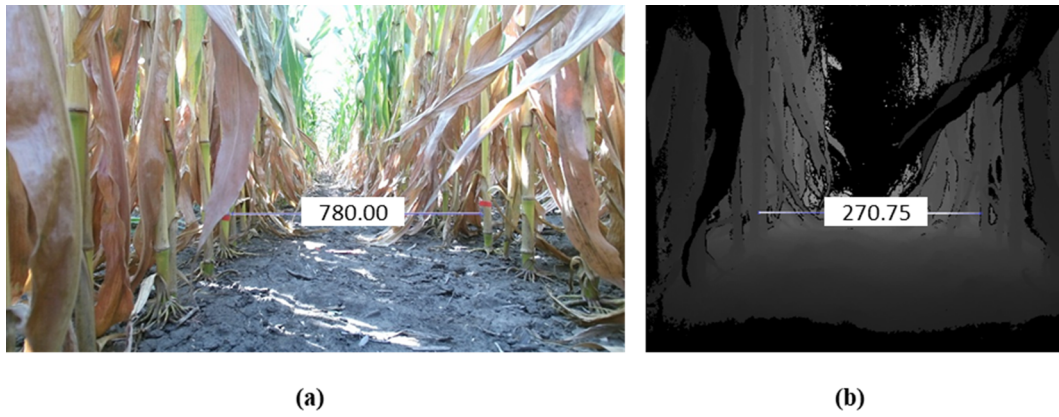
**Fig. 7.** Field setup for acquisition of (a) RGB and (b) depth images. Corn rows spacing, in pixels, were manually collected on both RGB and depth images for all cameras tested.

coefficients for Eq. (3) generated with both experimental data and theoretical data (from camera specifications and Eq. (2)), for depth and color images, respectively. Table 3 also contains the coefficients for Eq. (4) for the experimental data. Because depth cameras provide the distance between camera and objects being measured, units' transformations can be automatically performed, without the need to use a scale on the image.

$$l_{cm} = l_{px} \times a \times Z^b \tag{3}$$

where

 $l_{cm}$ = length, in cm;
 $l_{px}$ = length, in pixels;
 $Z$ = distance from depth camera to object being measured, in m;
 $a$ = coefficient;
 $b$ = coefficient.

$$S_{cm2} = S_{px} \times a \times Z^b \tag{4}$$

where

 $S_{cm2}$ = area, in cm$^2$;
 $S_{px}$ = area, in pixels;
 $Z$ = distance from depth camera to object being measured, in m;
 $a$ = coefficient;
 $b$ = coefficient.

It is possible to verify that, the coefficients obtained for theoretical and experimental data are different. This difference is either larger or smaller depending on the camera used. The errors between theoretical and experimental data are presented on Table 5 for each of the positions on the image tested. Fig. 8 presents the length ratio (px cm$^{-1}$) for both experimental and theoretical data. From these, it is possible to note that for all cameras, except ZED stereo camera, the deviation between experimental and theoretical data is smaller for objects located on the center of the image (position 1). This is an indication that all cameras

need to be calibrated for image distortion when performing dimensions acquisition, especially if objects are not positioned on the center of the image.

Besides the image distortion errors, other sources of errors also exist. Camera-to-camera FoV variation is one of these sources, e.g. Intel® specifications for RealSense™ D435's FoV for both depth and color images has ± 3° camera-to-camera variation, and an assessment of the true FoV of the camera should be performed in other to obtain more precise dimension acquisitions. Another source is the distortion on the object's border, e.g. the foam board square sides definition is dependent on the depth technology used and, therefore, can generate a smaller/greater number of pixels on the object's edges. This error doesn't occur for the measurements on the color images and it is possible to observe on Table 4 that the deviation between coefficients for experimental and theoretical coefficients for these images is smaller than for depth images.

In order to calculate true volume from a depth map, unit conversions from pixels to measurement units is needed, since the volume is calculated by the sum of the distance data from object to its support surface ('height of the object') for the whole object area. As the distance data provided by the sensors are in cm and the area of the object is given in number of pixels, the volume is retrieved in an unwanted unit (px cm). Another problem is the fact that the area of the object in the image varies with its distance from the sensor. This generates the need for correction of the value obtained to perform any comparison between volumes. In other words, for the same object, different values of volumes, lengths and areas can be acquired if the distance from the sensor varies. As what changes to the calculation of the volume is the area of the object and not its depth (distance between the square and the wall), this value can be adjusted using Eq. (4), obtained for correcting the area unit.

Testing the effects of using different sizes of foam board squares on the length (px cm$^{-1}$) and area (px cm$^{-2}$) ratios, showed (Table 6) that the 10 cm × 10 cm square is different ($p < 0.05$) from the others for all cameras, except for the acquisition of the area ratio with CamBoard

**Table 2**
Average distances (m) and standard deviation (m) obtained by five depth cameras, for the five analyzed distances with their respective standard deviations. One hundred and twenty-one points (N = 121) were used for each image to gather each value, and three cameras of each type were used.

| Camera | Distances from sensor to wall (m) | | | | | P-value |
|---|---|---|---|---|---|---|
| | 1.00 | 1.50 | 2.00 | 2.50 | 3.00 | |
| CamBoard Pico Flexx | 0.99 ± 0.001 | 1.48 ± 0.002 | 1.99 ± 0.003 | 2.50 ± 0.004 | 3.00 ± 0.006 | 0.18710 |
| Intel® RealSense™ D435 | 0.99 ± 0.002 | 1.47 ± 0.012 | 1.97 ± 0.022 | 2.46 ± 0.035 | 2.97 ± 0.036 | 0.02538 |
| Microsoft Kinect v.1 | 0.99 ± 0.002 | 1.49 ± 0.003 | 1.99 ± 0.006 | 2.49 ± 0.008 | 3.00 ± 0.013 | 0.00002 |
| Microsoft Kinect v.2 | 1.00 ± 0.001 | 1.50 ± 0.001 | 2.01 ± 0.001 | 2.51 ± 0.002 | 3.01 ± 0.003 | 0.70697 |
| ZED Stereo Camera | 0.99 ± 0.001 | 1.48 ± 0.001 | 1.99 ± 0.002 | 2.48 ± 0.004 | 2.99 ± 0.003 | 0.00007 |

**Table 3**

Coefficients for unit transformation equations from px to cm (length) and from px to cm$^2$ (area) on depth images provided by five different depth cameras.

| | Theoretical length coeficientes[1] | | $R^2$ | Experimental length coefficients[1] | | $R^2$ | Area coefficients[2] | | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Camera | a | b | | a | b | | a | b | |
| CamBoard Pico Flexx | 0.484 | 1.000 | 1.000 | 0.505 | 0.927 | 0.950 | 0.257 | 1.934 | 0.977 |
| Intel® RealSense™ D435 | 0.154 | 1.000 | 1.000 | 0.143 | 0.862 | 0.807 | 0.023 | 1.866 | 0.824 |
| Microsoft Kinect v.1 | 0.164 | 1.000 | 1.000 | 0.163 | 0.937 | 0.965 | 0.029 | 1.941 | 0.984 |
| Microsoft Kinect v.2 | 0.268 | 1.000 | 1.000 | 0.268 | 0.927 | 0.928 | 0.074 | 1.950 | 0.972 |
| ZED Stereo Camera | 0.071 | 1.000 | 1.000 | 0.061 | 0.999 | 0.956 | 0.004 | 2.033 | 0.955 |

[1] $length_{cm} = length_{px} \times a \times (Distance\ from\ the\ camera)^{b}$.

[2] $area_{cm^2} = area_{px} \times a \times (Distance\ from\ the\ camera)^{b}$.

**Table 4**

Coefficients for unit transformation equations from px to cm (length) on RGB images provided by four depth cameras.

| | Theoretical length coefficients[1] | | $R^2$ | Experimental length coefficients[1] | | $R^2$ |
|---|---|---|---|---|---|---|
| Camera[2] | a | b | | a | b | |
| Intel® RealSense™ D435 | 0.162 | 1.000 | 1.000 | 0.166 | 1.011 | 0.983 |
| Microsoft Kinect v.1 | 0.188 | 1.000 | 1.000 | 0.194 | 0.988 | 0.987 |
| Microsoft Kinect v.2 | 0.094 | 1.000 | 1.000 | 0.092 | 1.020 | 0.990 |
| ZED Stereo Camera | 0.071 | 1.000 | 1.000 | 0.068 | 1.030 | 0.978 |

[1] $length_{cm} = length_{px} \times a \times (Distance\ from\ the\ camera)^{b}$.

[2] CamBoard Pico Flexx is not equipped with an RGB camera.

**Table 5**

Averages and standard errors (N = 45) obtained for length ratio (px cm$^{-1}$) deviation, for the four positions on image (1 - center, 2 - edge on the horizontal axis, 3 - edge on the vertical axis, 4 - corner) and fifteen depth cameras (three of each type).

| Camera | Position | Length ratio deviation (px cm$^{-1}$) |
|---|---|---|
| CamBoard Pico Flexx | 1 | 0.04 ± 0.03 |
| | 2 | 0.10 ± 0.08 |
| | 3 | 0.08 ± 0.06 |
| | 4 | 0.09 ± 0.08 |
| Intel® RealSense™ D435 | 1 | 0.47 ± 0.36 |
| | 2 | 0.69 ± 0.57 |
| | 3 | 0.69 ± 0.56 |
| | 4 | 0.75 ± 0.60 |
| Microsoft Kinect v.1 | 1 | 0.13 ± 0.13 |
| | 2 | 0.13 ± 0.11 |
| | 3 | 0.21 ± 0.20 |
| | 4 | 0.23 ± 0.27 |
| Microsoft Kinect v.2 | 1 | 0.06 ± 0.06 |
| | 2 | 0.08 ± 0.05 |
| | 3 | 0.28 ± 0.18 |
| | 4 | 0.19 ± 0.19 |
| ZED Stereo Camera | 1 | 1.28 ± 0.89 |
| | 2 | 1.22 ± 0.83 |
| | 3 | 1.34 ± 0.78 |
| | 4 | 1.43 ± 1.10 |

Pico Flexx (PMD Technologies). That could be explained by the fact that very small objects suffer from edge deformation effect, as indicated by Gottfried et al. (2011). Microsoft Kinect v.1 showed differences (p < 0.05) between all sizes of squares used for both length ratio and area ratio, and ZED Stereo Camera (StereoLabs) showed differences (p < 0.05) between all sizes of square used when measuring length ratio. This could indicate that a square of 20 × 20 cm is still too small for these cameras and suffer from edge effect.

The test of the effect of the objects' positions in the image (center, edge on the horizontal axis, edge on the vertical axis, and corner) on the length (px cm$^{-1}$) and area (px cm$^{-2}$) ratios showed (Table 7) a different behavior for each type of camera used. For the ToF cameras (CamBoard Pico Flexx and Microsoft Kinect v.2), all positions presented different (p < 0.05) length ratios amongst the four positions tested. Additionally, for these cameras, the depth images presented unusable regions on the corner of the images, so the foam board squares had to be moved to the area closest to the corner where the camera would acquire useful distance information. The area ratio for CamBoard Pico Flexx camera is the same (p < 0.05) between positions 1 and 3 (center and edge on vertical axis) and between positions 2 and 4 (edge on horizontal axis and corner). The area ratio for Microsoft Kinect v.2 is the same (p < 0.05) between positions 1 and 2, and between 1 and 4; position 3 (edge on vertical axis) differs (p < 0.05) from the others.

For acquiring length ratio with Intel® RealSense™ D435, position 1 (center) differs (p < 0.05) from the others, position 3 (edge on vertical axis) has the same effect as positions 2 (edge on horizontal axis) and 4 (corner). The effect of positions on the area ratio for this camera was the same presented by Microsoft Kinect v.1 for both length and area ratios: positions 1 and 2 have the same (p < 0.05) behavior and positions 3 and 4 have the same behavior.

Dutta (2012), showed that the standard deviation of the distance data increases on the corners of the image. Thus, the ideal for data comparison of length and area of objects acquired with depth cameras is positioning them at a fixed region of the image, preferably in the center. If offsets need to be made, the ideal is to make them on the horizontal direction of the image for all cameras, except for CamBoard Pico Flexx (PMD Technologies); in which the differences are smaller, reducing distortion of values and enabling data comparison. ZED Stereo Camera (StereoLabs) did not present any effect of positions on the length and area ratios acquisition and, thus, the objects can be positioned in any place on the image. Additionally, for ToF cameras, objects should not be positioned on the corner of images, as these regions do not present useful distance data.
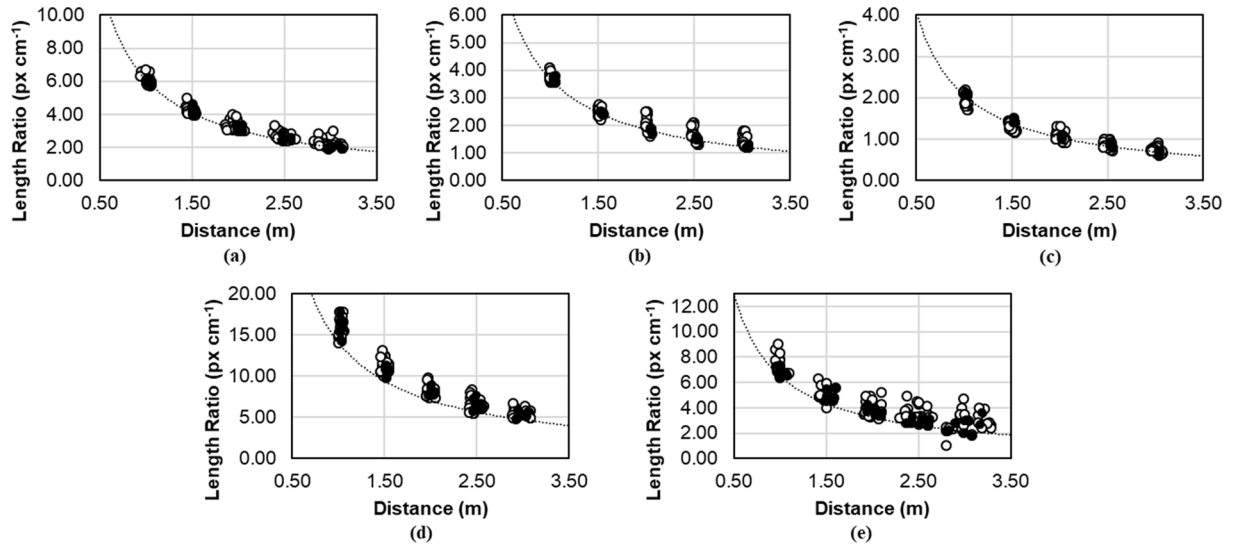
CamBoard Pico Flexx (PMD Technologies), Intel® RealSense™ D435, and Microsoft Kinect v.2 showed no significant effect (p < 0.05) on the length ratio (px cm$^{-1}$) and on the area ratio (px cm$^{-2}$) when using different cameras to acquire those dimensions (Table 8).

Microsoft Kinect v.1 camera showed significant (p < 0.05) interaction between position and size of the foam board square for length ratio. Microsoft Kinect v.2 showed significant interaction between position and size of the foam board square for both area ratio and length ratio estimation. ZED Stereo Camera showed significant (p < 0.05) interaction between camera and size of the foam board square for length ratio estimation. For both Kinect cameras, it appears that the source of interaction is the 10 × 10 cm foam board square and for the ZED camera it appears that the source is the 30 × 30 cm square.

*3.3. Maximum useful distance*

Average distance from camera to the foam board square (m) was plotted against actual distance (Fig. 10a, Fig. 11a, Fig. 12a, Fig. 13a, and Fig. 14a) for both indoors and outdoors for all cameras. Table 9

**Fig. 8.** Length ratio (px cm$^{-1}$) varying with distance (m) from camera to objects for all cameras used: (a) Microsoft Kinect v.1, (b) Microsoft Kinect v.2, (c) CamBoard Pico Flexx (PMD Technologies), (d) ZED Stereo camera (StereoLabs) and (e) Intel® RealSense™ Depth Camera D435. Dots represent the experimental data and the dashed line represents the theoretical data. Black markers are the squares on the center of the image (position 1).

shows the maximum distance of useful depth data on both environments and classifies the usability of the data for performing measurements outdoors.

The cameras that use stereoscopy (Intel® RealSense™ and ZED Stereo Camera) could provide useful data on an outdoor environment up to 12 m (Intel, Fig. 11a) or 20 m (ZED, Fig. 14a). For the ToF cameras, some data can be acquired outdoors if the distance from the objects being analyzed is small (1 m for CamBoard Pico Flexx, Fig. 12a and 2 m for Microsoft Kinect v.2, Fig. 15a), which prevents their use for autonomous robotic navigation, for example. Kinect v.1 did not present useful data on an outside environment at any distance (Fig. 12a).

The residuals between actual distance and camera-provided distance were calculated and plotted (Fig. 9b, Fig. 10b, Fig. 11b, Fig. 12b, and Fig. 13b) against distance. Intel® RealSense™ D435 (Fig. 10b) and ZED Stereo Camera (Fig. 13b) presented the highest residuals, up to 3.0 and 5.0 m, respectively, while the other cameras (Flexx, Fig. 9b; Kinect v.1, Fig. 11b, and Kinect v.2, Fig. 12b) presented errors ranging from −0.2 to 0.2 m. Also, the residuals for those two cameras increased with increasing distances, differently from the other cameras, that presented a more constant residual with no visible pattern. A power regression model of residuals versus actual distance was calculated for both Intel® RealSense™ D435 and ZED Stereo Camera, which presented residuals with a visually non-random behavior. The power regression equations

present form shown in Eq. (5), and the coefficients for these equations are presented on Table 10.

$$Res = a \times Z^b \tag{5}$$

where

$Res$ = residuals, in m;
$Z$ = distance from camera to object being measured, in m;
$a$ = coefficient;
$b$ = coefficient.

### 3.4. Minimum useful distance

Average distance from camera to the foam board square (m) was acquired and is presented on Table 11. Intel® RealSense™ D435 and CamBoard Pico Flexx performed the best for close range applications (as close as 0.20 m for Intel and 0.30 m for Flexx). Both Kinect cameras could sense objects as close as 0.80 m, but in order to have the correct shape of objects on the scene, the 1.0 m distance is more advisable. ZED Stereo camera showed the same shape problems as encountered on the previous tests, but with the square as close as 0.70 m the region where the object should be had too many holes, so the shape got even more compromised.
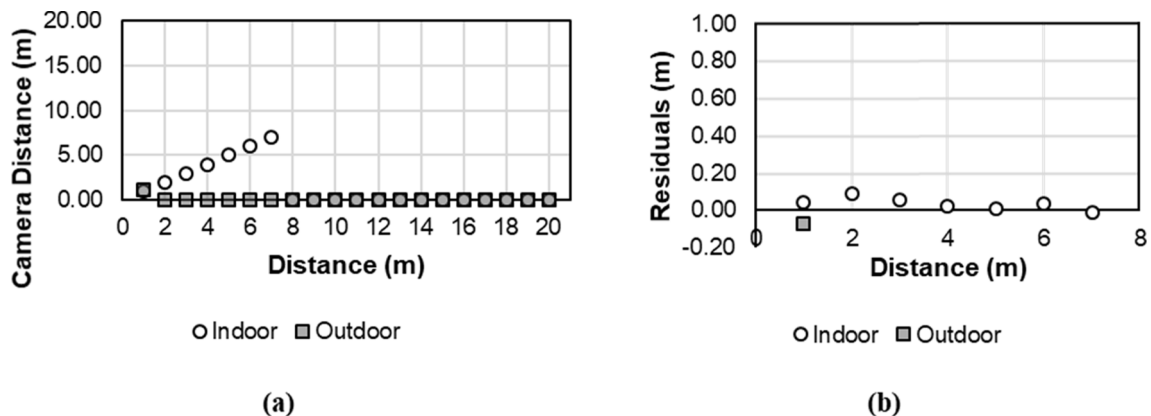


**Fig. 9.** Average distance from camera to the foam board square (m) versus actual distance (m) and residuals of the measurements for CamBoard Pico Flexx (PMD Technologies).

**Table 6**

Averages and standard errors (N = 60) obtained for length ratio (px cm$^{-1}$) and area ratio (px cm$^{-2}$), for the three sizes of square used (10 × 10 cm, 20 × 20 cm and 30 × 30 cm) and fifteen depth cameras (three of each type).

| Camera | Square Size (cm) | Length Ratio (px cm$^{-1}$) | Area Ratio (px cm$^{-2}$) |
|---|---|---|---|
| CamBoard Pico Flexx | 10 × 10 | 1.20 ± 0.01[a] | 1.54 ± 0.02 |
| | 20 × 20 | 1.17 ± 0.01[b] | 1.54 ± 0.02 |
| | 30 × 30 | 1.17 ± 0.01[b] | 1.55 ± 0.02 |
| Intel® RealSense™ D435 | 10 × 10 | 4.84 ± 0.06[a] | 21.58 ± 0.39[a] |
| | 20 × 20 | 4.24 ± 0.06[b] | 17.65 ± 0.39[b] |
| | 30 × 30 | 4.11 ± 0.06[b] | 16.85 ± 0.39[b] |
| Microsoft Kinect v.1* | 10 × 10 | 3.92 ± 0.17[a] | 15.08 ± 0.14[a] |
| | 20 × 20 | 3.61 ± 0.17[b] | 13.84 ± 0.14[b] |
| | 30 × 30 | 3.50 ± 0.17[c] | 13.16 ± 0.14[c] |
| Microsoft Kinect v.2* | 10 × 10 | 2.33 ± 0.15[a] | 5.56 ± 0.05[a] |
| | 20 × 20 | 2.19 ± 0.15[b] | 5.25 ± 0.05[b] |
| | 30 × 30 | 2.15 ± 0.15[b] | 5.11 ± 0.05[b] |
| ZED Stereo Camera* | 10 × 10 | 9.87 ± 0.07[a] | 90.50 ± 1.77[a] |
| | 20 × 20 | 9.34 ± 0.07[b] | 85.43 ± 1.77[b] |
| | 30 × 30 | 9.04 ± 0.07[c] | 83.18 ± 1.77[b] |

[a,b,c]Rows for each column of each camera section, with different superscripts, are significantly different (p < 0.05).

\* Microsoft Kinect v.1 showed significant interaction between position and size of the foam board square for length ratio. Microsoft Kinect v.2 showed significant interaction between position and size of the foam board square for both area ratio and length ratio. ZED Stereo Camera showed significant interaction between camera and size of the foam board square for length ratio.

**Table 7**

Averages and standard errors obtained (N = 45) for length ratio (px cm$^{-1}$) and area ratio (px cm$^{-2}$), for the four positions on image (1 – center, 2 - edge on the horizontal axis, 3 - edge on the vertical axis, 4- corner) and fifteen depth cameras (three of each type).

| Camera | Position | Length Ratio (px cm$^{-1}$) | Area Ratio (px cm$^{-2}$) |
|---|---|---|---|
| CamBoard Pico Flexx | 1 | 1.21 ± 0.01[a] | 1.69 ± 0.03[a] |
| | 2 | 1.08 ± 0.01[b] | 1.41 ± 0.03[b] |
| | 3 | 1.28 ± 0.01[c] | 1.68 ± 0.03[a] |
| | 4 | 1.14 ± 0.01[d] | 1.40 ± 0.03[b] |
| Intel® RealSense™ D435 | 1 | 4.16 ± 0.07[a] | 16.98 ± 0.45[a] |
| | 2 | 4.38 ± 0.07[b] | 17.85 ± 0.45[a] |
| | 3 | 4.46 ± 0.07[bc] | 19.60 ± 0.45[b] |
| | 4 | 4.58 ± 0.07[c] | 20.34 ± 0.45[b] |
| Microsoft Kinect v.1* | 1 | 3.56 ± 0.02[a] | 13.44 ± 0.17[a] |
| | 2 | 3.61 ± 0.02[a] | 13.82 ± 0.17[a] |
| | 3 | 3.78 ± 0.02[b] | 14.48 ± 0.17[b] |
| | 4 | 3.76 ± 0.02[b] | 14.37 ± 0.17[b] |
| Microsoft Kinect v.2* | 1 | 2.14 ± 0.02[a] | 5.13 ± 0.06[ab] |
| | 2 | 2.09 ± 0.02[b] | 5.03 ± 0.06[a] |
| | 3 | 2.42 ± 0.02[c] | 5.86 ± 0.06[c] |
| | 4 | 2.27 ± 0.02[d] | 5.21 ± 0.06[b] |
| ZED Stereo Camera | 1 | 9.35 ± 3.89 | 86.70 ± 75.07 |
| | 2 | 9.28 ± 3.77 | 82.41 ± 67.40 |
| | 3 | 9.46 ± 3.64 | 86.81 ± 67.09 |
| | 4 | 9.59 ± 3.93 | 89.56 ± 73.07 |

[a,b,c,d]Rows for each column of each camera section, with different superscripts are significantly different (p < 0.05).

\* Microsoft Kinect v.1 showed significant interaction between position and size of the foam board square for length ratio. Microsoft Kinect v.2 showed significant interaction between position and size of the foam board square for both area ratio and length ratio.

**Table 8**

Averages and standard errors obtained (N = 60) for length ratio (px cm$^{-1}$) and area ratio (px cm$^{-2}$), for the three sizes of square used (10 × 10 cm, 20 × 20 cm and 30 × 30 cm) and fifteen depth cameras (three of each type).

| Camera | Camera # | Length Ratio (px cm$^{-1}$) | Area Ratio (px cm$^{-2}$) |
|---|---|---|---|
| CamBoard Pico Flexx | 1 | 1.18 ± 0.01 | 1.55 ± 0.02 |
| | 2 | 1.18 ± 0.01 | 1.55 ± 0.02 |
| | 3 | 1.18 ± 0.01 | 1.53 ± 0.02 |
| Intel® RealSense™ D435 | 1 | 4.44 ± 0.06 | 19.29 ± 0.40 |
| | 2 | 4.44 ± 0.06 | 18.82 ± 0.40 |
| | 3 | 4.30 ± 0.06 | 17.96 ± 0.40 |
| Microsoft Kinect v.1 | 1 | 3.65 ± 0.02[a] | 13.79 ± 0.14[a] |
| | 2 | 3.66 ± 0.02[a] | 13.99 ± 0.14[ab] |
| | 3 | 3.72 ± 0.02[b] | 14.30 ± 0.14[b] |
| Microsoft Kinect v.2 | 1 | 2.22 ± 0.01 | 5.29 ± 0.05 |
| | 2 | 2.23 ± 0.01 | 5.31 ± 0.05 |
| | 3 | 2.23 ± 0.01 | 5.32 ± 0.05 |
| ZED Stereo Camera* | 1 | 9.84 ± 0.07[a] | 94.89 ± 1.77[a] |
| | 2 | 9.41 ± 0.07[b] | 87.97 ± 1.77[b] |
| | 3 | 9.01 ± 0.07[c] | 76.24 ± 1.77[c] |

[a,b,c]Rows for each column of each camera section, with different superscripts are significantly different (p < 0.05).

\* ZED Stereo Camera showed significant interaction between camera and size of the foam board square for length ratio.

### 3.5. Field test of cameras

#### 3.5.1. Animal phenotypic data evaluation

Top view images were taken on a group of pigs as they were being weighed. Fig. 14 shows the sample of the images captured with all five depth cameras used. As the images are evaluated, all the pigs' shapes are easily visible in all the images except the image captured from the ZED Stereo Camera. For this camera, the main body of the animal is easily seen, but the head and the tail are difficult to discern from the background. The image taken with Kinect v.1 has rough edges on the pig. The slightly rough edges are also observed on the image taken with the RealSense™ camera. Both Time-of-Flight cameras had a clear image, and Kinect v.2 have the smoothest outline of the pigs, probably due to its higher resolution.

The projected volumes (cm$^3$) acquired from the body of the animals, without the head and tail regions, were plotted against the mass (kg) of the animals (Fig. 15). Linear regressions were fitted, and equations were generated with the form presented on Eq. (6) and coefficients presented on Table 11. The best correlations between projected volume and mass was obtained for the Time-of-Flight technology (Microsoft Kinect v.2 and CamBoard Pico Flexx). Cameras that use the structured light technology (Intel® RealSense™ and Microsoft Kinect v.1) also had a good correlation (R$^2$ higher than 0.9) between those two variables. The camera that uses stereoscopy-only (StereoLabs ZED) had the lowest correlation (R$^2$ = 0.798) between projected volume and mass. Considering both the shapes provided and the depth information, the cameras that use ToF technology would be more advisable to be used for indoor applications, such as animal phenotyping, followed by cameras that use structured light technology. Stereoscopy-based technology presented the worst performance for this application.

$$V = a \times M + b \tag{6}$$

where

$V$ = volume, in cm$^3$;
$M$ = mass, in kg;
$a$ = coefficient;
$b$ = coefficient.
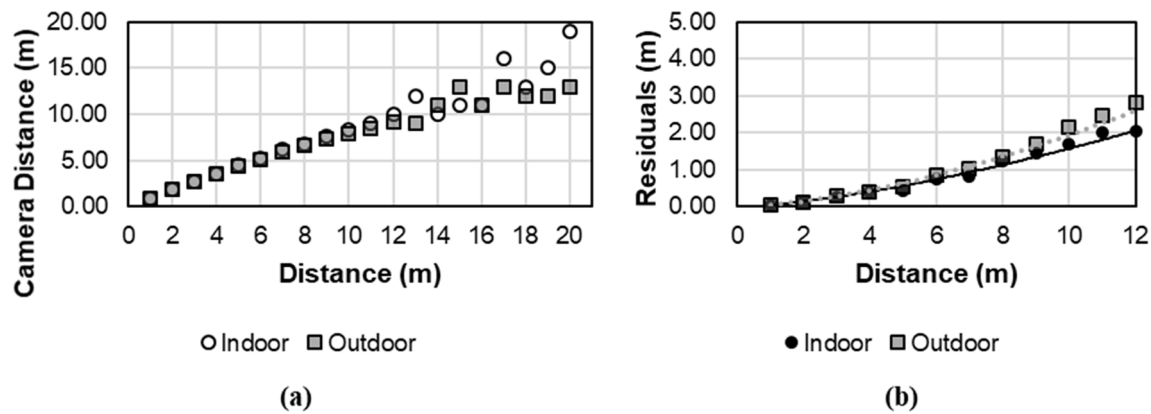
**Fig. 10.** Average distance from camera to the foam board square (m) versus actual distance (m) and residuals of the measurements for Intel® RealSense™ D435. Note the predictable and repeatable pattern in the residuals.

### 3.5.2. Camera evaluation in simulated and crop field environments

Depth images of corn rows obtained with all cameras can be seen in Fig. 16. The structured light camera (Microsoft Kinect v.1) was the only camera that did not provide useful information in the field. Because the plants were tall enough to provide shade, the ToF cameras (CamBoard Pico Flexx and Microsoft Kinect v.2) generated useful information. However, taking in to consideration their performance on the outdoor section of experiment 3 of this paper, it is likely these cameras would have difficulties generating images in certain situations in the field.

The corn rows and the gap between them were discernable on the images, especially for Microsoft Kinect v.2 and Intel® RealSense™ Depth Camera D435. The manually acquired row spacing, in pixels, for both the RGB and depth images were transformed to metric units using Eq. (3) and theoretical coefficients presented in Tables 3 and 4 of the second experiment. These values are presented on Fig. 17 and the absolute percentage residuals of these measurements are provided on Table 12. All cameras, but Pico Flexx presented acceptable residuals. The smallest residuals for depth images were obtained for ZED Stereo Camera, followed by Microsoft Kinect v.2 and Intel® RealSense™. For the RGB images, ZED Stereo Camera also presented the smallest residuals, followed by Kinect v.1 and Intel® RealSense™. Because ZED did not show any differences between the different positions for dimensions' acquisition, camera calibration concerning image distortion on the borders of the image is less important and, therefore, it was expected that the unit transformation equations for this camera would show smaller errors without a previous calibration. Because the spacings were manually selected in the images, further investigation is needed in how these points could be automatically selected for row spacing acquisition on depth images. When looking into automatic image processing and segmentation, it is visible (Fig. 17) that Kinect v.2

presented the most discernible shapes, followed by RealSense™. Taking into consideration that the ToF cameras would likely have difficulties in certain field situations, i.e. with less shadows and in a more open space (younger plants), RealSense™ would be the best camera for this application. This preliminary investigation indicated the potential uses of off-the-shelf depth cameras for autonomous applications given that estimating the distance between the crop rows could allow a small robotic machine to autonomously navigate for under the canopy sensing (see Table 13).

### 3.6. Final considerations

In this section, overall observations about the cameras are pointed out:

i. For outdoor use, both ZED Stereo camera (StereoLabs) and Microsoft Kinect v.1 generate overexposed RGB initial images (Fig. 18).

ii. When acquiring single images, Intel® RealSense™ Depth Camera D435 can produce non-usable images after camera is triggered a few times (Fig. 19). With this, the camera has to be restarted in order to produce useful images again.

iii. Time-of-flight cameras produce unreadable zones on the corner of images (Fig. 20), so positioning of objects should be carefully considered.

i. Some of these cameras are not available for commercialization across the globe. User has to consider costs for international shipping if camera is not available in their country. Microsoft Kinect v.2 is not being produced anymore.
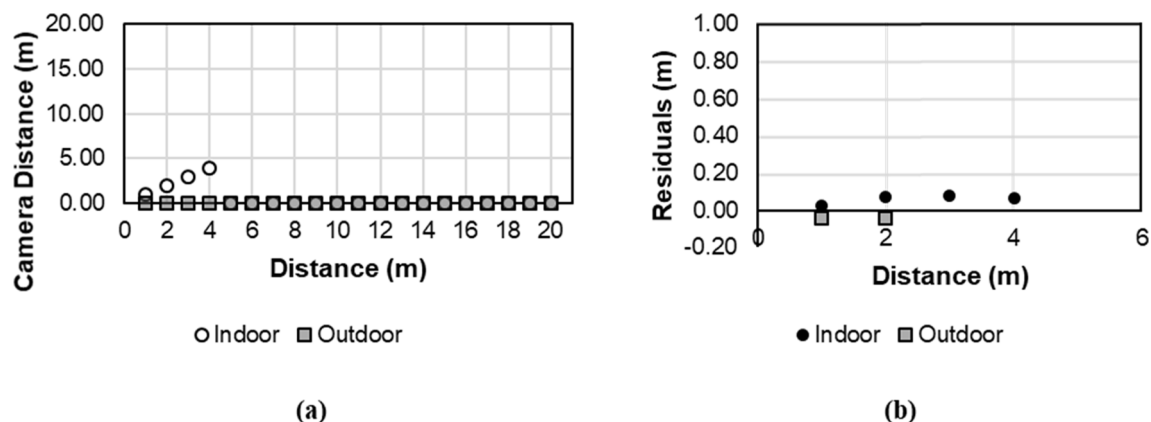


**Fig. 11.** Average distance from camera to the foam board square (m) versus actual distance (m) and residuals of the measurements for Microsoft Kinect v.1.
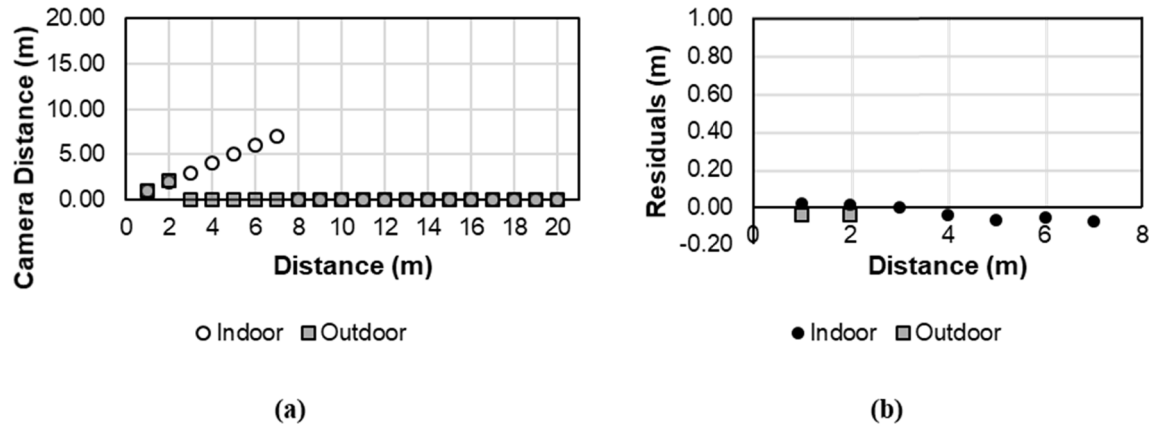
**Fig. 12.** Average distance from camera to the foam board square (m) versus actual distance (m) and residuals of the measurements for Microsoft Kinect v.2.
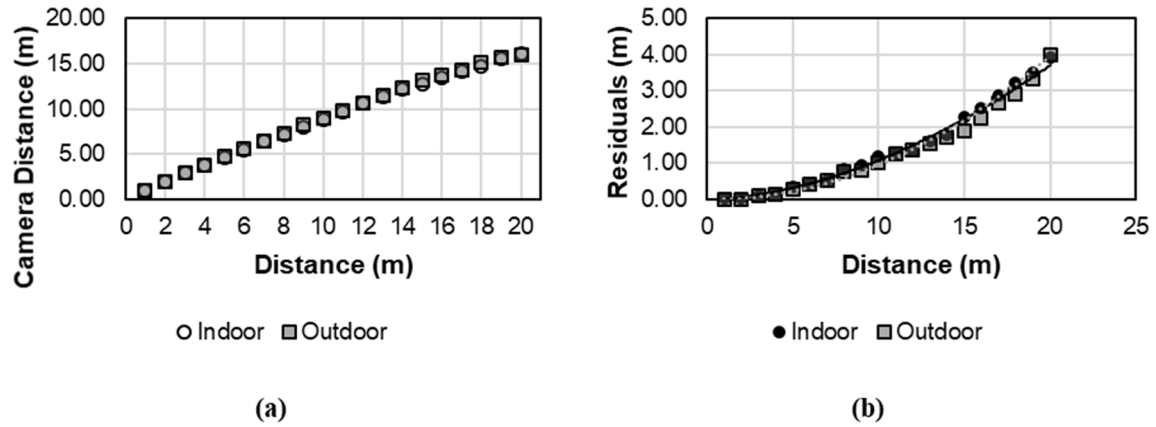


**Fig. 13.** Average distance from camera to the foam board square (m) versus actual distance (m) and residuals of the measurements for ZED Stereo Camera (StereoLabs). Note the predictable and repeatable nature of the residuals.
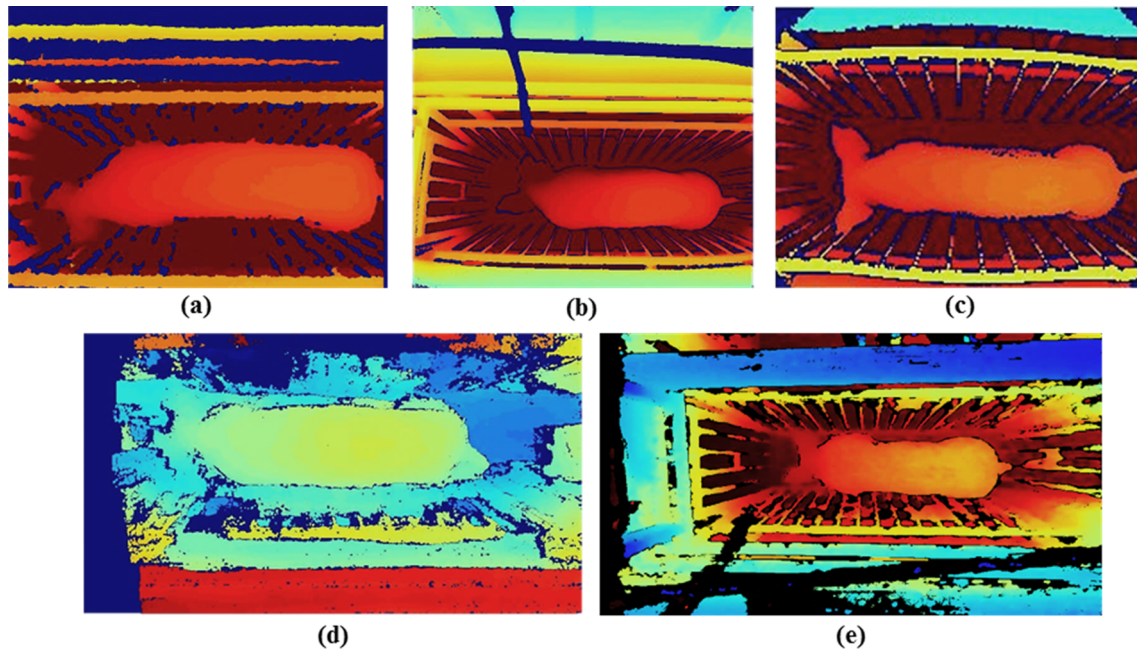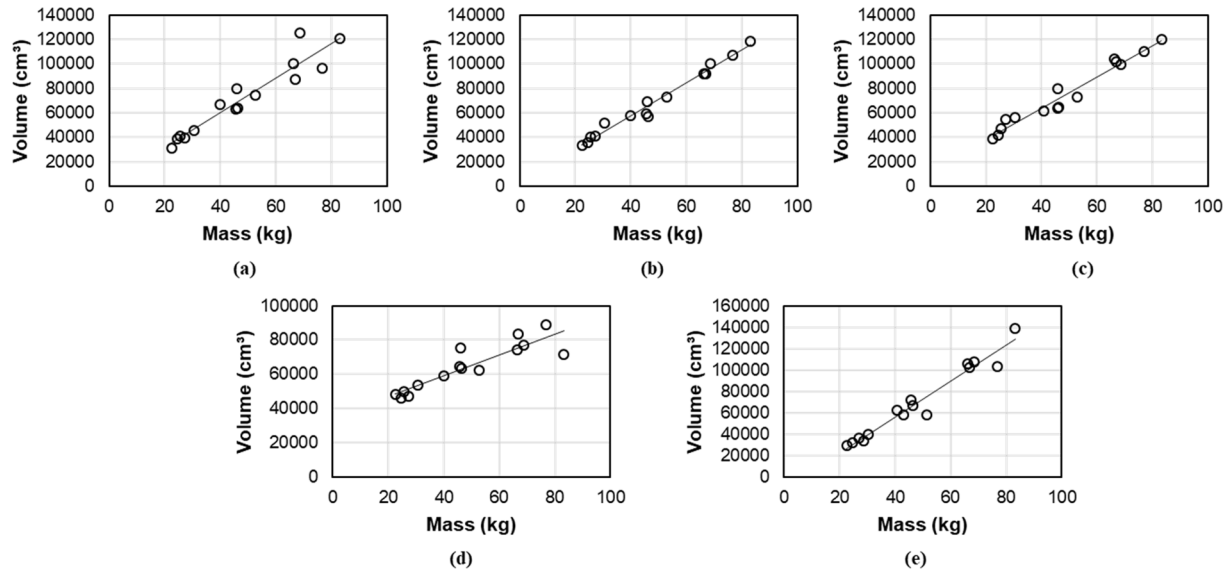


**Fig. 14.** Sample depth images acquired from the grow-finish pigs for all five cameras used: (a) Microsoft Kinect v.1, (b) Microsoft Kinect v.2, (c) CamBoard Pico Flexx (PMD Technologies), (d) ZED Stereo camera (StereoLabs) and (e) Intel® RealSense™ Depth Camera D435. Cameras were firmly mounted at a distance of 1.70 m above the floor.

**Table 9**

Maximum distance (m) that depth cameras can acquire data for both indoor and outdoor environments and possibility of performing dimensions' measures outdoor.

| Camera | CamBoard Pico Flexx | Intel® RealSense™ D435 | Microsoft Kinect v.1 | Microsoft Kinect v.2 | ZED Stereo Camera |
|---|---|---|---|---|---|
| **Max. Distance Indoor (m)** | 7.0 | 20.0 | 4.0 | 7.0 | 20.0 |
| **Max. Distance Outdoor (m)** | 1.0 | 20.0 | 0.0 | 2.0 | 20.0 |
| **Measurement Outdoor** | No | Yes | No | No | Yes |



**Fig. 15.** Linear regressions of volume versus pig mass obtained with 5 different depth cameras: (a) Microsoft Kinect v.1, (b) Microsoft Kinect v.2, (c) CamBoard Pico Flexx (PMD Technologies), (d) ZED Stereo camera (StereoLabs) and (e) Intel® RealSense™ Depth Camera D435. A total of 15 pigs were used develop equations; coefficients are shown in Table 11.

**Table 10**

Coefficients of residuals power curves for Intel® RealSense™ D435 and ZED Stereo Camera, indoor and outdoor.

| Camera | Coefficients indoor[1] | | $R^2$ | Coefficients outdoor[1] | | $R^2$ |
|---|---|---|---|---|---|---|
| | a | b | | a | b | |
| Intel® RealSense™ D435 | 0.054 | 1.466 | 0.991 | 0.049 | 1.597 | 0.991 |
| ZED Stereo Camera | 0.019 | 1.766 | 0.993 | 0.008 | 1.049 | 0.978 |

[1] $residuals_m = a \times (Distance\ from\ the\ camera)^{b.}$

**Table 11**

Average distances (m) and standard deviation (m) of points within foam board square obtained by five depth cameras, for the ten analyzed distances. "-" indicates that the camera did not generate an image of the square at the specific distance.

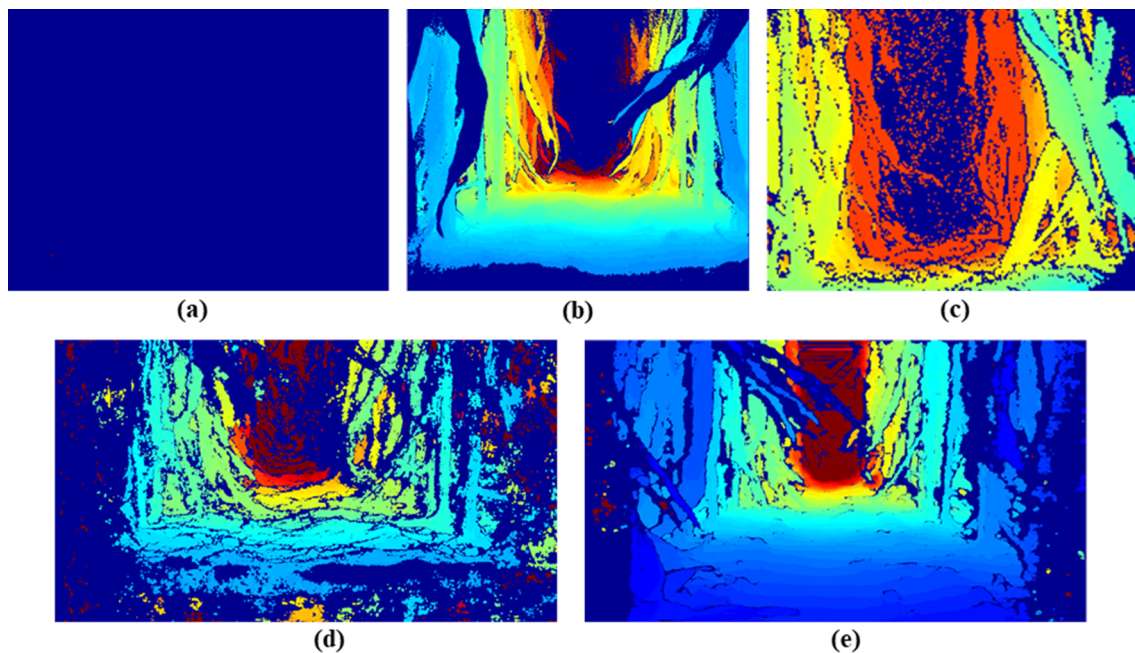| Actual Distance (m) | Camera Distance (m) | | | | |
|---|---|---|---|---|---|
| | CamBoard Pico Flexx | Intel® RealSense™ D435 | Microsoft Kinect v.1 | Microsoft Kinect v.2 | ZED Stereo Camera |
| **0.1** | – | – | – | – | – |
| **0.2** | – | 0.20 ± 0.03 | – | – | – |
| **0.3** | 0.28 ± 0.04 | 0.30 ± 0.03 | – | – | – |
| **0.4** | 0.39 ± 0.05 | 0.41 ± 0.03 | – | – | – |
| **0.5** | 0.49 ± 0.06 | 0.51 ± 0.04 | – | – | – |
| **0.6** | 0.59 ± 0.11 | 0.61 ± 0.07 | – | – | – |
| **0.7** | 0.68 ± 0.11 | 0.71 ± 0.07 | – | – | 0.71 ± 0.08* |
| **0.8** | 0.78 ± 0.11 | 0.81 ± 0.08 | 0.80 ± 0.09* | 0.81 ± 0.05* | 0.80 ± 0.10 |
| **0.9** | 0.88 ± 0.11 | 0.91 ± 0.07 | 0.90 ± 0.10* | 0.91 ± 0.06* | 0.90 ± 0.09 |
| **1.0** | 0.98 ± 0.11 | 1.02 ± 0.08 | 1.00 ± 0.11 | 1.01 ± 0.07 | 0.99 ± 0.09 |

\* These images, although showed some parts of the foam board square with reliable distance information, did not contain reliable shape information, i.e. square was partially on the picture.

## 4. Conclusions

Low-cost depth-cameras use one or a combination of three technologies: structured light, time-of-flight (ToF), and stereoscopy. Five different cameras were tested for their suitability to be used in agriculture applications. Significant, but small camera to camera differences were found for ZED Stereo Camera and Kinect v.1 camera ($p < 0.05$). Increases in measurements standard deviation were found

**Fig. 16.** Depth images acquired from the corn field for all five cameras used: (a) Microsoft Kinect v.1, (b) Microsoft Kinect v.2, (c) CamBoard Pico Flexx (PMD Technologies), (d) ZED Stereo camera (StereoLabs) and (e) Intel® RealSense™ Depth Camera D435. Note figure a has no useful data because Kinect v.1 does not provide depth images outdoors.
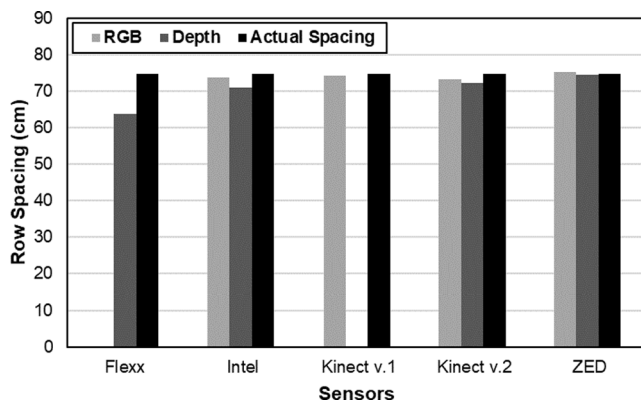


**Fig. 17.** Corn rows' spacing acquired with all five cameras tested: CamBoard Pico Flexx (PMD Technologies), Intel® RealSense™ Depth Camera D435, Microsoft Kinect v.1, Microsoft Kinect v.2, and ZED Stereo camera (StereoLabs); for both the RGB images and depth images. Distance information for the RGB data unit transformation was manually acquired. CamBoard Pico Flexx does not provide RGB images, and Kinect v.1 did not provide useful depth data on the corn field.

**Table 12**
Coefficients and determination coefficient ($R^2$) of linear regression equations for predicting mass (kg) from volume ($cm^3$).

| Camera | Coefficients[1] | | $R^2$ |
|---|---|---|---|
| | a | b | |
| CamBoard Pico Flexx | 1281.5 | 12599.0 | 0.960 |
| Intel® RealSense™ D435 | 1679.1 | 11146.0 | 0.953 |
| Microsoft Kinect v.1 | 1411.0 | 3497.6 | 0.901 |
| Microsoft Kinect v.2 | 1347.0 | 3643.9 | 0.979 |
| ZED Stereo Camera | 606.5 | 35054.0 | 0.798 |

[1] $volume_{cm3} = a \times mass_{kg} + b$.

**Table 13**
Absolute percentage residual (%) for corn rows' spacing collected, in pixels, on RGB and Depth images provided by five cameras tested - CamBoard Pico Flexx (PMD Technologies), Intel® RealSense™ Depth Camera D435, Microsoft Kinect v.1, Microsoft Kinect v.2, and ZED Stereo camera (StereoLabs) - and transformed to cm using Eq. (2) developed on this paper. CamBoard Pico Flexx does not provide RGB images, and Kinect v.1 did not provide useful depth data on the corn field.

| Camera | Absolute Percentage Residual RGB | Absolute Percentage Residual Depth |
|---|---|---|
| CamBoard Pico Flexx | – | 14.81 |
| Intel® RealSense™ D435 | 1.50 | 5.12 |
| Microsoft Kinect v.1 | 0.82 | – |
| Microsoft Kinect v.2 | 2.20 | 3.56 |
| ZED Stereo Camera | 0.57 | 0.47 |

for all cameras as the distance between camera and object increased; however, Intel® RealSense™ camera had a much larger increase. Time-of-flight cameras had the smallest error between different sizes of objects but had non-readable zones on the corners of the images. Errors in area and length measurements were found to differ at different positions in the image for all cameras except ZED. This distortion is greater in the vertical axis of the image for all cameras, except CamBoard Pico Flexx. Cameras that use stereoscopy-based technology can be used for outdoors applications. Cameras that use ToF technology, although provide some data on outside environments, should be used outside only if necessary and in a close range (up to 1.0–2.0 m). All cameras provide useful distance data indoors, and the stereoscopy cameras have both a larger range of operation and larger errors, that increase with increasing distances in a predictable and repeatable way. Considering both the smoothness of the shape provided and the error associated with the depth information, the ToF cameras will work best for housed-animal phenotyping, followed by cameras that use structured light technology. For outdoor applications, stereoscopic cameras provided the widest range of operation and more reliable data on extremely lit environments, i.e. the sun light. The errors associated with those cameras were predictable and repeatable, which is an indication that a calibration is possible to guarantee the most accurate measurements.
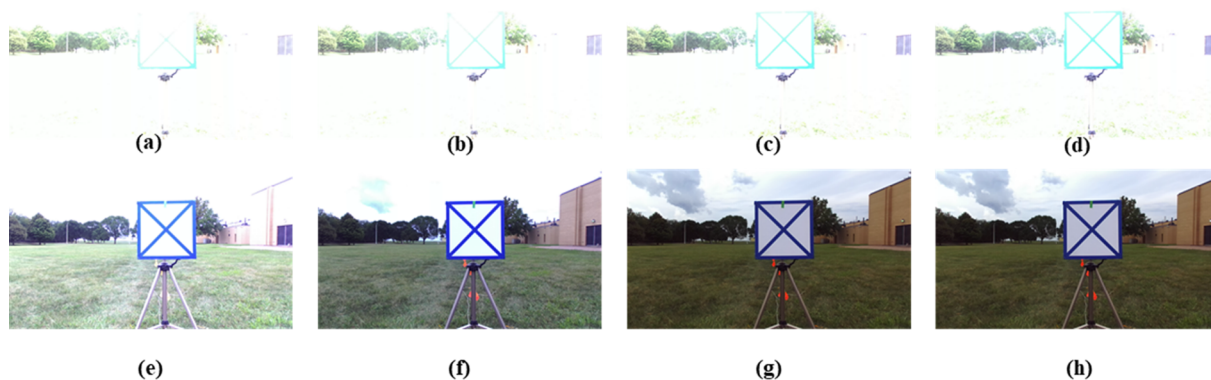
**Fig. 18.** Overexposed initial RGB images (a–f) before stabilization (g and h).
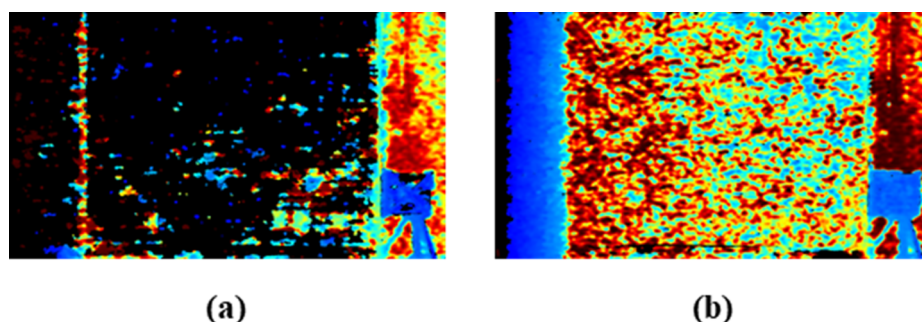


**Fig. 19.** Intel® RealSense™ Depth Camera D435 can produce non-usable images after camera is triggered a few times for single pictures. (a) shows an unusable image and (b) shows an image of the same scene after the camera was restarted.
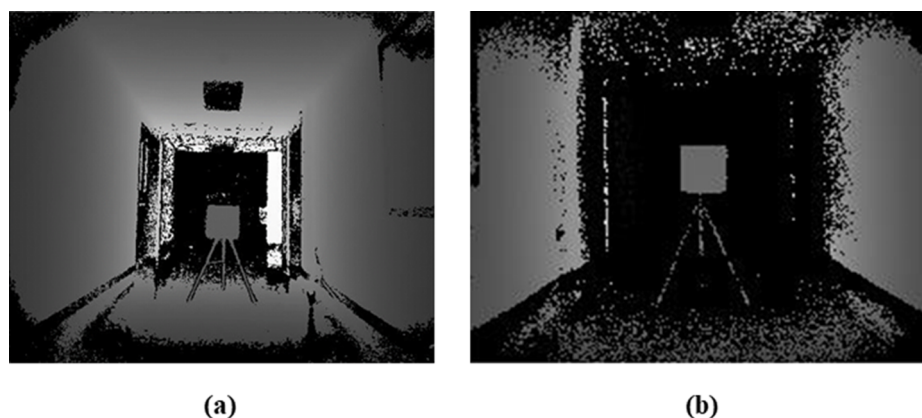


**Fig. 20.** Unreadable zones on the corner of images generated by time-of-flight cameras: (a) Microsoft Kinect v.2, (b) Camboard Pico Flexx (PMD Technologies).

Some depth cameras have added features that allow for position of the camera to be recorded. These added features are helpful for autonomous applications to capture plant phenotypic data or crop growth data; in addition, these cameras can be used to provide input for autonomous robot navigation.

### CRediT authorship contribution statement

**Isabella C.F.S. Condotta:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Funding acquisition. **Tami M. Brown-Brandl:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Santosh K. Pitla:** Resources, Writing - review & editing. **John P. Stinn:** Conceptualization, Methodology.

**Késia O. Silva-Miranda:** Supervision, Funding acquisition.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

experimental corn field, the Meat Animal Research Center for providing access to their facilities; and John Holman, Nathan Olsufka, Brooke Compton, and Hessan Sedaghat for helping with data acquisition.

## References

Andersen, M.R., Jensen, T., Lisouski, P., Mortensen, A.K., Hansen, M.K., Gregersen, T., Ahrendt, P.J.A.U., 2012. Kinect Depth Sensor Evaluation for Computer Vision Applications. Aarhus University, pp. 1–37 https://doi.org/Technical Report ECE-TR-6.

Benavidez, P., Jamshidi, M., 2011. Mobile robot navigation and target tracking system. In: 2011 6th International Conference on System of Systems Engineering. IEEE, pp. 299–304.

Berkovic, G., Shafir, E., 2012. Optical methods for distance and displacement measurements. Adv. Opt. Photonics 4 (4), 441–471. https://doi.org/10.1364/AOP.4.000441.

Chang, Y.J., Chen, S.F., Huang, J.D., 2011a. A Kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res. Dev. Disabil. 32 (6), 2566–2570. https://doi.org/10.1016/j.ridd.2011.07.002.

Chang, Y.J., Wang, T.Y., Chen, Y.R., 2011b. A location-based prompting system to transition autonomously through vocational tasks for individuals with cognitive impairments. Res. Dev. Disabil. 32 (6), 2669–2673. https://doi.org/10.1016/j.ridd.2011.06.006.

Condotta, I.C., Brown-Brandl, T.M., Silva-Miranda, K.O., Stinn, J.P., 2018a. Evaluation of a depth sensor for mass estimation of growing and finishing pigs. Biosyst. Eng. 173, 11–18. https://doi.org/10.1016/j.biosystemseng.2018.03.002.

Condotta, I.C., Brown-Brandl, T.M., Stinn, J.P., Rohrer, G.A., Davis, J.D., Silva-Miranda, K.O., 2018b. Dimensions of the modern pig. Trans. ASABE 61 (5), 1729–1739. https://doi.org/10.13031/trans.12826.

Correa, D.S.O., Sciotti, D.F., Prado, M.G., Sales, D.O., Wolf, D.F., Osorio, F.S., 2012. Mobile robots navigation in indoor environments using kinect sensor. In: 2012 Second Brazilian Conference on Critical Embedded Systems. IEEE, pp. 36–41. https://doi.org/10.1109/CBSEC.2012.18.

Draper, N.R., Smith, H., 2014. Applied Regression Analysis Vol. 326 John Wiley & Sons.

Dutta, T., 2012. Evaluation of the Kinect™ sensor for 3-D kinematic measurement in the workplace. Appl. Ergon. 43 (4), 645–649. https://doi.org/10.1016/j.apergo.2011.09.011.

Efroymson, M.A., 1960. Multiple regression analysis. In: Ralston, A., Wilf, H.S. (Eds.) Mathematical Methods for Digital Computers.

Ganganath, N., Leung, H., 2012. Mobile robot localization using odometry and kinect sensor. In: 2012 IEEE International Conference on Emerging Signal Processing Applications. IEEE, pp. 91–94.

Gottfried, J.M., Fehr, J., Garbe, C.S., 2011. Computing range flow from multi-modal kinect data. In: International Symposium on Visual Computing. Springer, Berlin, Heidelberg, pp. 758–767.

Guo, H., Ma, X., Ma, Q., Wang, K., Su, W., Zhu, D., 2017. LSSA_CAU: An interactive 3d point clouds analysis software for body measurement of livestock with similar forms of cows or pigs. Comput. Electron. Agric. 138, 60–68. https://doi.org/10.1016/j.compag.2017.04.014.

Hao, Guo, Shengli, Zhang, Qin, Ma., Wang Peng, Su., Wei, Zhu Dehai, Bing, Qi, 2014. Measurement of body size of dairy cows based on point cloud acquisition equipment. J. Agric. Eng. 30 (5), 116–122.

Hernandez-Lopez, J.J., Quintanilla-Olvera, A.L., López-Ramírez, J.L., Rangel-Butanda, F.J., Ibarra-Manzano, M.A., Almanza-Ojeda, D.L., 2012. Detecting objects using color and depth segmentation with Kinect sensor. Procedia Technol. 3, 196–204. https://doi.org/10.1016/j.protcy.2012.03.021.

Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Fitzgibbon, A., 2011. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp. 559–568.

Keselman, L., Iselin Woodfill, J., Grunnet-Jepsen, A., Bhowmik, A., 2017. Intel realsense stereoscopic depth cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–10. https://doi.org/10.1109/CVPRW.2017.167.

Khoshelham, K., Elberink, S.O., 2012. Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 12 (2), 1437–1454. https://doi.org/10.3390/s120201437.

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008.

Kulikov, V.A., Khotskin, N.V., Nikitin, S.V., Lankin, V.S., Kulikov, A.V., Trapezov, O.V., 2014. Application of 3-D imaging sensor for tracking minipigs in the open field test. J. Neurosci. Methods 235, 219–225. https://doi.org/10.1016/j.jneumeth.2014.07.012.

Lao, F., Brown-Brandl, T., Stinn, J.P., Liu, K., Teng, G., Xin, H., 2016. Automatic recognition of lactating sow behaviors through depth image processing. Comput. Electron. Agric. 125, 56–62. https://doi.org/10.1016/j.compag.2016.04.026.

Lee, J., Jin, L., Park, D., Chung, Y., 2016. Automatic recognition of aggressive behavior in pigs using a kinect depth sensor. Sensors 16 (5), 631. https://doi.org/10.3390/s16050631.

Phillips, R.W., Dawson, W.M., 1936. A study of methods for obtaining measurements of swine. J. Anim. Sci. 1936 (1), 93–99.

Sarbolandi, H., Lefloch, D., Kolb, A., 2015. Kinect range sensing: structured-light versus time-of-flight kinect. Comput. Vis. Image Underst. 139, 1–20. https://doi.org/10.1016/j.cviu.2015.05.006.

Stavrakakis, S., Li, W., Guy, J.H., Morgan, G., Ushaw, G., Johnson, G.R., Edwards, S.A., 2015. Validity of the Microsoft Kinect sensor for assessment of normal walking patterns in pigs. Comput. Electron. Agric. 117, 1–7. https://doi.org/10.1016/j.compag.2015.07.003.

Wang, K., Guo, H., Ma, Q., Su, W., Chen, L., Zhu, D., 2018. A portable and automatic Xtion-based measurement system for pig body size. Comput. Electron. Agric. 148, 291–298. https://doi.org/10.1016/j.compag.2018.03.018.

Zaragoza, L.E.O., 2009. Evaluation of the accuracy of simple body measurements for live weight prediction in growing-finishing pigs. MSc. Diss.. Univ. of Illinois, Urbana, Illinois.

Zhang, S., 2018. High-speed 3D shape measurement with structured light methods: a review. Opt. Lasers Eng. 106, 119–131. https://doi.org/10.1016/j.optlaseng.2018.02.017.

Zhu, Q., Ren, J., Barclay, D., McCormack, S., Thomson, W., 2015. Automatic animal detection from kinect sensed images for livestock monitoring and assessment. In: 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE, pp. 1154–1157. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.172.