

# Depth Data Error Modeling of the ZED 3D Vision Sensor from Stereolabs

Luis E. Ortiz\*, Elizabeth V. Cabrera\*, and Luiz M. Gonçalves\*

\* *Department of Computer Engineering and Automation, Federal University of Rio Grande do Norte, Natal, Brazil*  
*lortiz@dca.ufrn.br, vcabrera@dca.ufrn.br, lmarcos@dca.ufrn.br*

Received 6th May 2017; accepted 3rd Apr 2018

---

## Abstract

The ZED camera is binocular vision system that can be used to provide a 3D perception of the world. It can be applied in autonomous robot navigation, virtual reality, tracking, motion analysis and so on. This paper proposes a mathematical error model for depth data estimated by the ZED camera with its several resolutions of operation. For doing that, the ZED is attached to a Nvidia Jetson TK1 board providing an embedded system that is used for processing raw data acquired by ZED from a 3D checkerboard. Corners are extracted from the checkerboard using RGB data, and a 3D reconstruction is done for these points using disparity data calculated from the ZED camera, coming up with a partially ordered, and regularly distributed (in 3D space) point cloud of corners with given coordinates  $(x_e, y_e, z_e)$ , which are computed by the device software. These corners also have their ideal world (3D) positions  $(x_i, y_i, z_i)$  known with respect to the coordinate frame origin that is empirically set in the pattern. Both given (computed) coordinates from the camera's data and known (ideal) coordinates of a corner can, thus, be compared for estimating the error between the given and ideal point locations of the detected corner cloud. Subsequently, using a curve fitting technique, we obtain the equations that model the RMS (Root Mean Square) error. This procedure is repeated for several resolutions of the ZED sensor, and at several distances. Results showed its best effectiveness with a maximum distance of approximately sixteen meters, in real time, which allows its use in robotic or other online applications.

*Key Words:* Depth Data, RMS Error, Stereo Vision, Stereolabs ZED.

---

## 1 Introduction

The human being acquires information about the location and other properties of objects within an environment thanks to its powerful and sophisticated vision system. The perception of a third dimension (depth) occurs due to the difference between images formed in the retinas of the left and right eyes. In the process of image formation, the catches of each eye are not equal because they present a slight variation in the position of the observed objects, attributed to the separation between the eyes. Artificial stereo vision systems are generally inspired by the biological process to extract three-dimensional information from digital images, which can be used to perform 3D reconstructions, tracking, and detection of objects.

---

Correspondence to: lmarcos@dca.ufrn.br

Recommended for acceptance by Alicia Fornés

<https://doi.org/10.5565/rev/elcvia.1084>

ELCVIA ISSN: 1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

There are several devices that provide three-dimensional information, depending on the operating technology they can be classified into stereo vision sensors, structured light devices or sensors based on the principle of Time of Flight (ToF). These devices are used in several areas with multiple purposes, in Robotics they are employed as essential tools in navigation applications, three-dimensional parts review, among others.

One of the most recent developed stereo sensors is the ZED camera, it has two high-resolution cameras that capture images (left and right) at the same time and transmit them to an external computer device for processing. It is a passive device, that is, it needs additional equipment for the power supply. According to manufacturers information, it is designed and built to perceive the depth of objects in indoor and outdoor environments within a range of 1 to 20 meters and at 100 FPS. This device can be used on several platforms, including standard PC and also SBC (Single Board Computers) as the Nvidia development kit. By attaching the ZED camera to the recently launched Jetson board of the Nvidia constitutes a powerful tool for applications in the field of Robotic Vision [1].

However, depth data provided by stereo devices have errors attributed to several aspects related to cameras hardware and computational processes that are performed to obtain these values. It is possible to enumerate some sources of errors as hardware system error, camera calibration error, feature extraction and stereo matching errors [2]. These inherent errors that such data present should be taken into account in the applications where depth data generated by 3D vision sensors are used, such an example is the Robotic Vision. In real applications, such as autonomous robotics, it is important to consider and treat those visual errors in order to achieve correct decision-making process during a navigation task, for example.

In this direction, the main contribution of this work is the mathematical modeling of the depth error determined by the ZED camera considering a left RGB image and the depth map. This subject is discussed in detail and an algorithm is proposed coming up with a methodology that we apply here to find the mathematical models of the RMS error of the Stereolabs ZED camera for all of its resolutions of operation ( $2208 \times 1242$ ,  $1920 \times 1080$ ,  $1280 \times 720$  and  $672 \times 376$  px). Curves that model the estimated error are provided, and data on capture times and depth calculation are also presented. The obtained results can become initial information for the development of Robotic and Computational Vision applications that need as a useful parameter, the determined error. As we have noticed, this relevant and important information has not yet been disclosed in any of previous works.

Also as a contribution of this article, the experimental data that we found is made available to the Computer Vision and Graphics community besides being relevant to our further work. Unlike the works similar to ours explained in the related works Section, next, here we demonstrate mathematically and experimentally the behavior of the error of the ZED device.

This paper follows with Section 2, which presents the ZED camera and introduces the previous formalism in order to better understand the problem. Then the most related works to ours are presented in Section 3. The proposed mathematical model for the ZED camera error is proposed in Section 4. Section 5 shows some experiments performed seeking to analyze the accuracy and time processing performance. At the end, Section 6 describes some final considerations about the proposed model and its achieved results.

## 2 3D Reconstruction with the ZED Camera

The ZED device is composed of stereo 2K cameras with dual 4MP RGB sensors. It has a field of view of  $110^\circ$  and can stream uncompressed video at a rate up to 100 FPS in WVGA format. It is an UVC-compliant USB 3.0 camera backward compatible with USB 2.0. Left and right video frames are synchronized and streamed as a single uncompressed video frame in the side-by-side format. Several configurations parameters of on-board ISP (Image Signal Processor) as resolution, brightness, contrast, saturation can be adjusted through the SDK that is provided by ZED development team. This camera has a compact structure and reduced size (Figure 1a), compared to other stereo cameras such as the Bumblebee XB3 for example. These characteristics make it relatively simple to incorporate into robotic systems or drones. Details of the dimensions are presented in the Figure 1b. The most important characteristics of the ZED camera are summarized in Table 1.

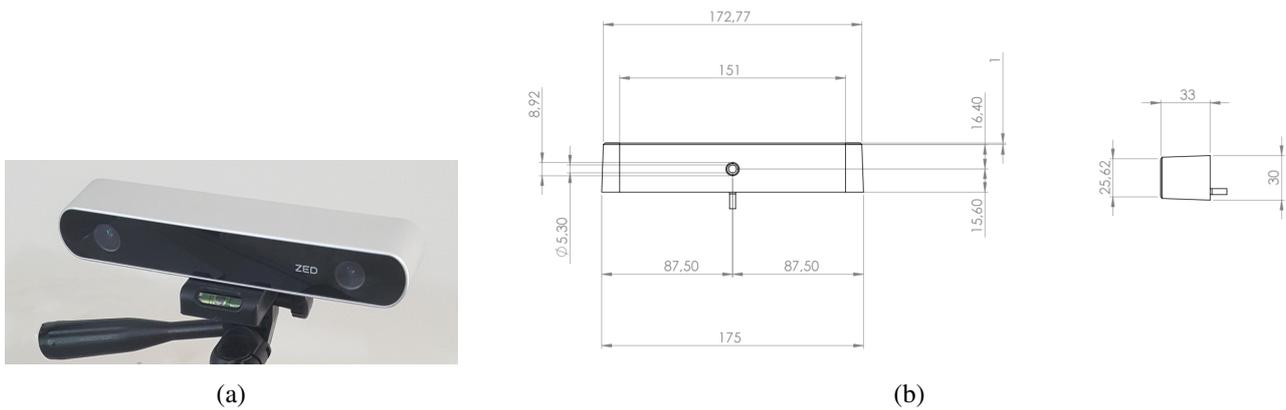


Figure 1: (a) ZED physical aspect and its (b) dimensions in millimeters.

<b>Size and weight</b>	Dimensions: 175×30×33 mm Weight: 159 g
<b>Individual image and depth resolution (pixels)</b>	HD2K: 2208 × 1242 (15 FPS) HD1080: 1920 × 1080 (30, 15 FPS) HD720: 1280 × 720 (60, 30, 15 FPS) WVGA: 672 × 376 (100, 60, 30, 15 FPS)
<b>Depth</b>	Range: 1-20 m Format: 32 bits Baseline: 120 mm
<b>Lens</b>	Field of View: 110 ° $f/2.0$ aperture
<b>Sensors</b>	Size: 1/3 Format: 16:9 Pixel Size: 2- $\mu$ pixels
<b>Connectivity</b>	USB 3.0 (5 V / 380 mA) 0°C to +45°C
<b>SDK System requirements</b>	Windows or Linux Dual-core 2.3 GHz 4 GB RAM Nvidia GPU

Table 1: Relevant features of the ZED.

## 2.1 Depth Calculation

As said, humans can have a three-dimensional perception of the world through the eyes due to the difference observed in the images formed in left and right retinas. In the imaging process, the images sent to the brain from each eye are not the same, with a slight difference in the position of the objects due to the separation between the eyes, which form a triangle with the scene points. Thanks to this difference, by triangulation the brain can determine the distance (depth) that the objects are in relation to the observer position. The implementation of stereo vision in computers uses this basic principle to recreate a 3D scene representation based on the two images of it taken from different viewing points. This is known as stereo reconstruction [3, 4]. In order to do stereo reconstruction, a series of steps are necessary, as calibration, rectification, and further depth determination, seen next.

### 2.1.1 Calibration

The calibration process estimates intrinsic and extrinsic parameters of the cameras. Intrinsic values include the focal length ( $f_x$ ,  $f_y$ ), principal point coordinates ( $C_x$ ,  $C_y$ ), radial ( $k_1$ ,  $k_2$ ,  $k_3$ ) and tangential ( $p_1$ ,  $p_2$ ) distortion factors. They are commonly used to obtain images without distortions, caused by the lenses and

camera construction process, and to obtain three-dimensional representations of a scene.

On the other hand, extrinsic parameters relate the real world reference systems and the camera, describing position and orientation of the device in the real world coordinate system (i.e. rotation matrix and translation vector). In addition to the aforementioned calibration (for each camera), may be developed a stereo calibration, this process allows obtaining information that relates the positions of the two cameras in space.

The ZED camera has factory calibration data, these parameters can also be optimized by self-calibration of ZED SDK to obtain non-distorted and aligned images after rectification process.

### 2.1.2 Rectification

Stereo rectification is the process in which a pair of stereo images are corrected, so that, it appears that they had been taken by two cameras with row-aligned image planes as shown in Figure 2. With such process the principal rays of the cameras are parallel, that is, they intersect at infinity. This step facilitates the stereo disparity estimation, a fundamental process prior to the estimation of the depth map. The ZED SDK provides rectified images.

### 2.1.3 Depth Estimation

The ZED camera computes depth information using triangulation (re-projection) from the geometric model of non-distorted rectified cameras, seen in Figure 2. Assuming that the two cameras are co-planar with parallel optical axes and same focal length  $f_l = f_r$ , the depth  $Z$  of each point  $L$  is calculated by Equation 1, here  $B$  is the baseline distance and  $x_i^l - x_i^r$  is the disparity value [4]. Notice that depth varies inversely proportional to the disparity.

$$Z = \frac{fB}{x_i^l - x_i^r} \quad (1)$$

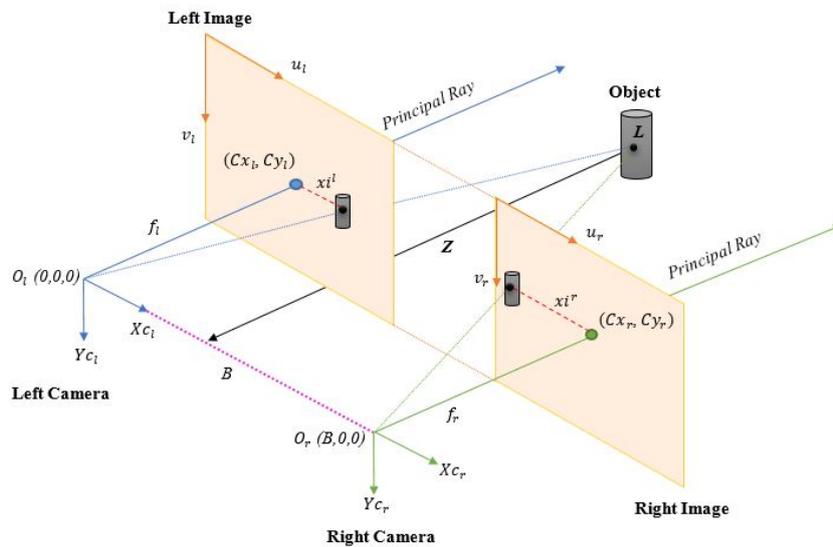


Figure 2: The operation principle of the ZED camera.

The Figure 3 shows how ZED cameras acquire 3D values considering the left camera as the origin frame. Given the coordinates of a pixel  $(u', v')$  in this image coordinate frame, these same coordinates are used to search their corresponding depth value  $Z$  in the depth map, Figure 3b.

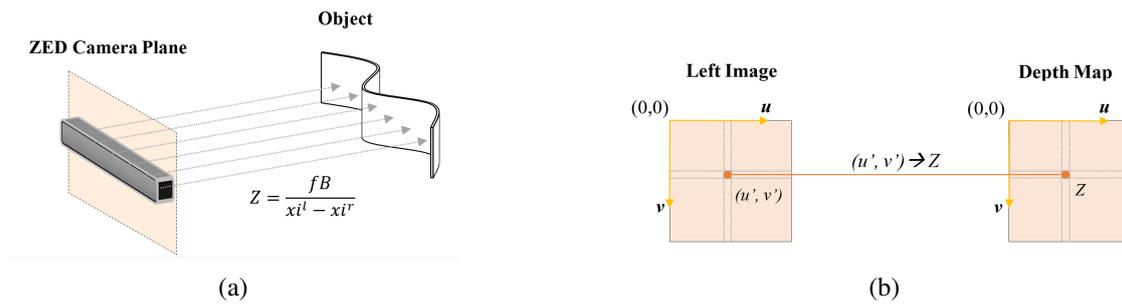


Figure 3: (a) Depth data estimation. (b) Left image and depth map relation.

### 3 Related Works

Several works in the literature make an analysis of the error determination in RGB-D and stereo sensor devices. It is reported in the literature the limitation of the MS Kinect [5], with the error in the depth measurements squared increasing according to the distance between the sensor and the target. For indoor mapping applications, it is recommended that data should be acquired from 1 to 3m of distance to the sensor. Several comparisons between versions of a device and between different devices are also made. A comparison of versions of MS Kinect shows that the accuracy of v2 is slightly better than v1 [6, 7]. Concerning static efficiency, its error is stable around 5mm up to a distance of 4.5m. Another comparison of MS Kinect v1 with Asus Xtion concludes that the first is more sensitive to bright illuminated objects [8]. A between different devices comparison, with the PointGrey Bumblebee XB3 and MS Kinect v2 [9], concludes that ToF cameras are more sensitive to ambient light while stereo vision devices have better resolution and are more robust mainly indicated for outdoor use.

In the same way, Sabattini [10] and Beltran [11] tested the PointGrey Bumblebee XB3 and MS Kinect v1 in a robotic application. The results show that, due to the resolution, the BumbleBee stereo camera has a better performance in the localization of a mobile robot. The performance of the stereo camera and RGB-D sensor is compared using the mobile robot and 2D landmarks demonstrating experimentally that the stereo cameras have a smaller error in the determination of the 3D position of known points in the image.

Procedures and metrics for error resolution are proposed with accuracy estimation methodologies for 3D vision sensors [12], and mathematical models to calculate the error at a given distance [5]. Other efforts towards obtaining the probability distribution of error measures in the three-dimensional location of corners in an image captured by a stereo sensor have been found in the literature. One of the first approaches in this direction tries to determine the probability that a 3D point has to reach a certain error value [13]. Latter, ways to provide the mathematical and statistical analyses of the error in stereo devices are proposed in the works of Chang [14, 15]. Equations are provided for the absolute, relative, and upper and lower depth data error limits. Song [16] suggests an approach generate a per-pixel confidence measure for each depth map captured by MS Kinect v1 in indoor scenes through supervised learning and the use of artificial intelligence.

In a work that we consider somewhat similar to our approach, the characteristics of MS Kinect v1 stochastic errors are presented for each direction of the axis, by Choo [17]. The depth error is measured using a 3D chessboard, and the results show that, for all three axes, the error should be considered independently. In the work of Kytö [18], it is presented a method to evaluate the depth accuracy of a stereo camera in applications focused on people detection. The depth resolution is determined using a 2D checkerboard and the variation of the focal length.

In the work that we consider most similar to ours found in the literature however applied to other device [11], it is performed the analysis of the error in-depth data provided by the Point Grey Bumblebee XB3. The error is calculated as the differences between actual and estimated depth measurements of sixteen Known points in a scene. Our methodology is different from the work since it uses another approach for the error determination

and also it is performed in much more points, in fact, some 112 points. So the main contribution of our work is the error determination for the ZED camera, and, more than that, to provide a way for extending this error measure to other devices that would use it in the same way.

## 4 Depth Error Estimation and Modeling

In this Section is described the sequence of steps for modeling the error behavior. Notice that is an important point in many tasks, mainly when there is a decision-making process depending on the noisy data from the camera.

### 4.1 Depth Error Estimation

The accuracy of the depth data given by the camera is first determined by the statistical analysis of 100 depth maps acquired consecutively, images are captured with the four resolutions at their highest image frame rate to assume that there are no changes in the environment as a variation of light. The images capture and the depth maps estimation is performed using the SDK of the stereo camera. The scene is in an external environment, and its characteristics do not vary in the course of the acquisition of the data. The camera is located approximately four meters from a flat surface. Depth values for the pixels in which the device does not estimate values are assumed to be null, in WVGA are eliminated the 2.09%, in HD720 the 2.06%, in HD1080 2.43% and in HD2K the 2.3%.

The color maps of Figure 4 show the computed standard deviations of each corresponding pixel for the depth maps. This graph allows knowing that although the sensitivity of the ZED varies for each resolution, in general, the majority of depth values present a standard deviation in a range from 0 to 0.1 m making it a poorly sensitive device. The depth data estimated in a resolution of HD2K present areas with greater sensitivity. These parts are those that are closer to the camera at distances less than 0.5 meters in which the device does not estimate correct depth values.

In the Figure 5 it is noticed that in the resolution of HD1080 the sensitivity of the depth estimation is smaller in comparison with the other resolutions. The standard deviations vary from 0 to 0.02 m and its median is 0.006. In WVGA and HD720 there is an increase in their deviations, but their maximum values do not exceed 0.04 m and its medians are 0.00816 and 0.008 m respectively. The HD2K resolution depth data are the ones that have the greatest variation (maximum of 0.9 m). However, 75% of the standard deviations are in the range of 0 to 0.04 m with a median of 0.01.

### 4.2 Modelling the Error

So, verified that stereo reconstruction has inherent errors in the estimation of depth data (and therefore in the determination of three-dimensional coordinate  $X, Y, Z$  of a point in space), it becomes necessary to determine this error in order to obtain more precise data to allow more robust applications. Towards achieving this goal, we have implemented a method to determinate the depth error reflected in the 3D coordinates provided by the device.

Considering two three-dimensional representations of a point (corner), an ideal (obtained from a chessboard pattern manually) with coordinates referenced to the camera system  $(X_{c_p}, Y_{c_p}, Z_{c_p})$ , and another given (estimated) by the ZED camera  $(X_{c_{p'}}, Y_{c_{p'}}, Z_{c_{p'}})$ , the error constitutes in the difference between the coordinates of the points  $p$  and  $p'$  aligned in a common coordinate system. Mathematically this value is defined by:

$$e = \left| \vec{p} - \vec{p}' \right| = \left| (X_{c_p} - X_{c_{p'}})i + (Y_{c_p} - Y_{c_{p'}})j + (Z_{c_p} - Z_{c_{p'}})k \right| \quad (2)$$

where,

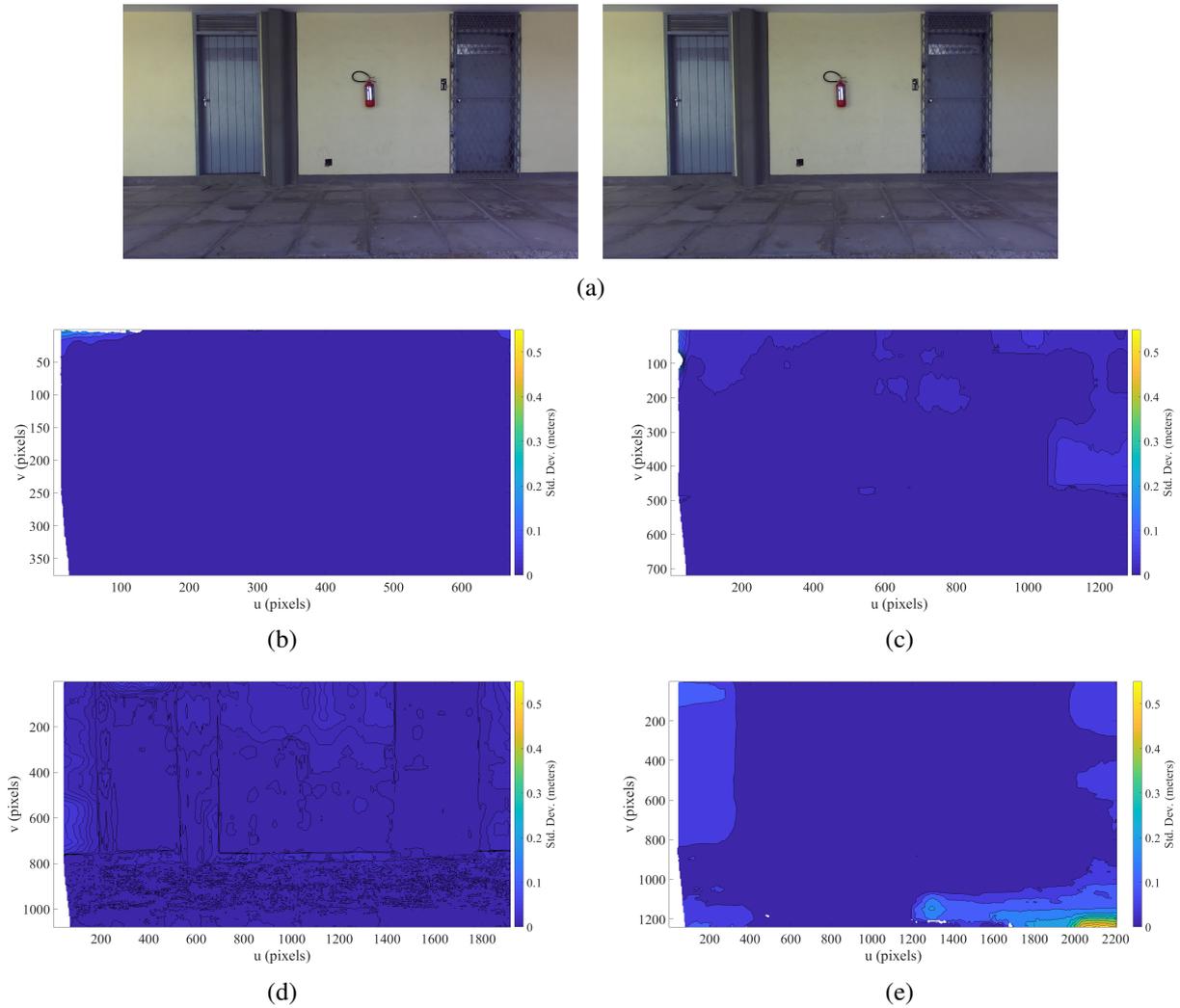


Figure 4: (a) Left and right RGB images acquired at a distance of 4.18 m. Standard deviation computed for each corresponding pixel for 100 subsequent depth maps generated by ZED, using all camera resolutions: (b) WVGA, (c) HD720, (d) HD1080 and (e) HD2K.

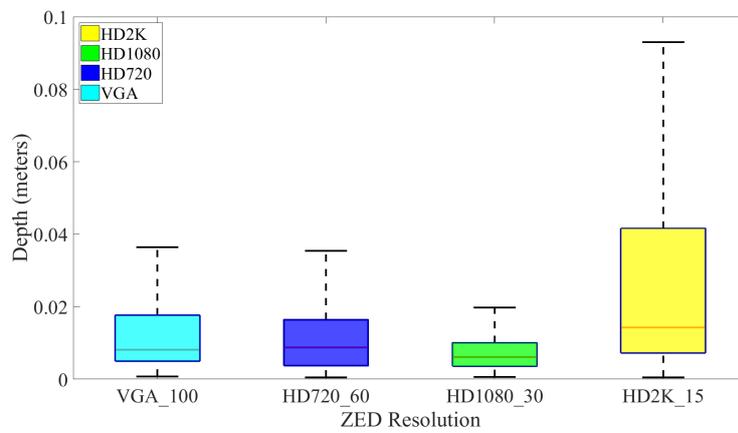


Figure 5: Box plot of standard deviations of depth ZED data.

$$X_{c_{p'}} = \frac{u_{p'} - Cx}{fx} Z_{c_{p'}} \quad (3)$$

$$Y_{c_{p'}} = \frac{v_{p'} - Cy}{fy} Z_{c_{p'}} \quad (4)$$

In the equations 3 and 4,  $u_{p'}$  and  $v_{p'}$  represent the position of a point in the image coordinate system, obtained from non-distorted rectified images.

#### 4.2.1 Method Implementation

- Capture: the implementation of the method requires data from stereo undistorted and rectified image of the left camera and the depth map of a pattern with three chessboards.
- Pattern Corners 2D Processing: detect the corners of the left image checkerboard. For well-known coordinates  $(u_{p'}, v_{p'})$  of each points of interest, search in the depth map the respective  $Z_{c_{p'}}$  values.
- Pattern Corners 3D Processing: creates two point clouds, one ideal and one real. The first cloud is constructed by measuring the position (in millimeters) of each corner of the checkerboard in reference to the point of intersection of three planes of the pattern. The real cloud is made using the reprojection equations 3 and 4. Lastly, make a register of the clouds in order to find a rigid transformation matrix that allows aligning their coordinated systems.
- RMS Error Estimation: after aligning the coordinate axes of the clouds, ideally the corresponding points should be placed in the same coordinate  $(X_c, Y_c, Z_c)$ ; however, it is observed that there is a variation. This is quantified by the calculation of the Euclidean distance between the two vectors of the same point with the origin of the coordinate system (Figure 6). To obtain a global representation of the error in the all corresponding points of clouds is computed the RMS error defined by the Equation 5, where  $\sum^2$  represents a sum of all individual Euclidean distances between two vectors and  $n$  are all corners in the pattern image.

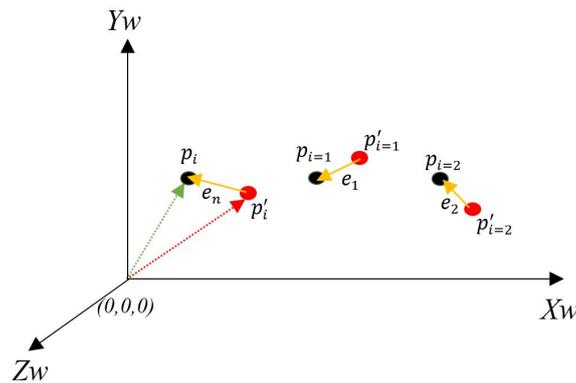


Figure 6: Representation of the calculation of the RMS depth error, determined from the difference between corresponding points of a ideal cloud (of points  $p_i$ ) and ZED estimated cloud (of points  $p'_i$ )

$$e_{RMS} = \sqrt{\frac{\sum^2}{n}} \quad (5)$$

- RMS error mathematical modeling: after obtaining the error data for several distances between the ZED camera and the pattern, it is possible to determine a mathematical model of the RMS error using curve fitting of least squares.

The Figure 7 indicates all development steps to determine the curves of the ZED's error.

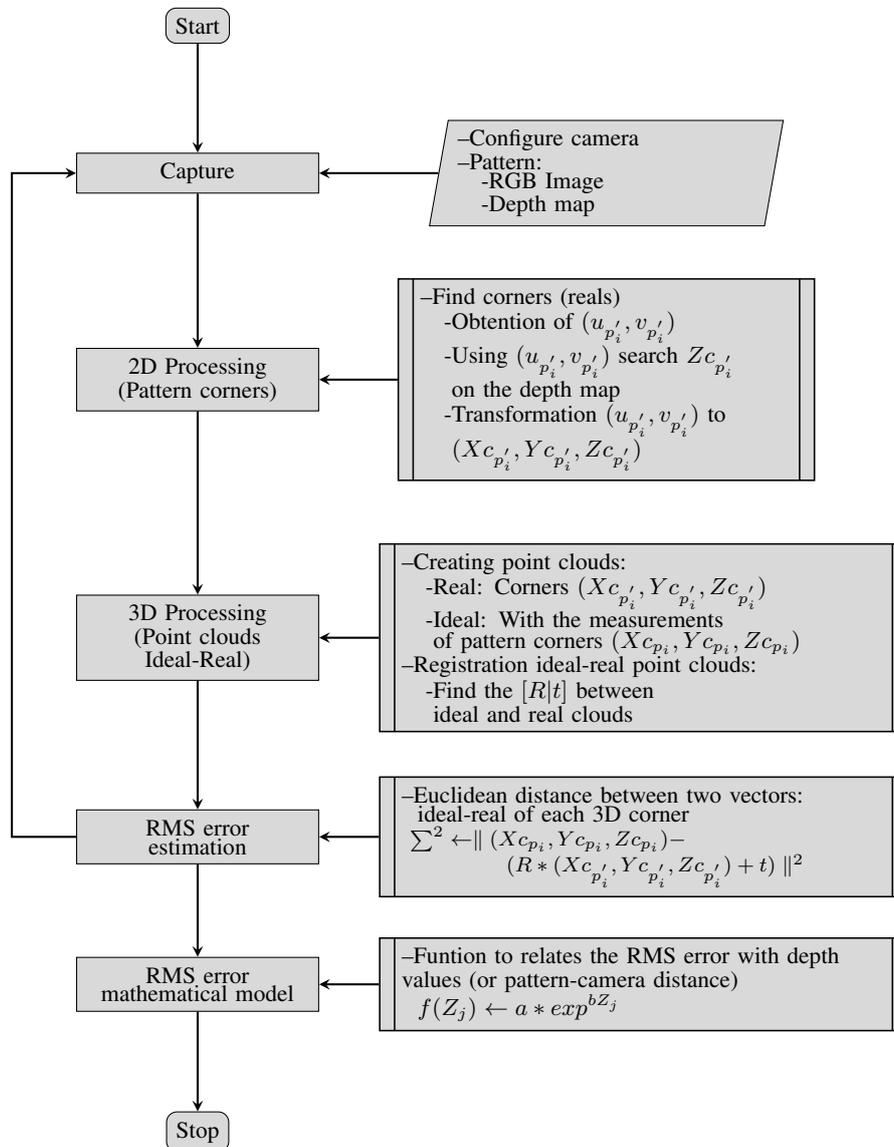


Figure 7: Implemented algorithm to compute the ZED's RMS depth error.

### 4.3 Curve Fitting for RMS Error Data Modeling

The curve fitting of RMS error data can be performed through several models such as linear, polynomial, exponential and power. Is necessary to evaluate if the set data fits the actual data, here is proposed to do by using the graphical and numerical analysis. In the graphical analysis an inspection of the residual values of the adjustment and the predicted limits is performed, meanwhile that the numerical analysis is done by computing statistical values as:

- **Sum of Squares due to Error (SSE):** this is a measure of the discrepancy between the data and an adjustment model. In Equation 6,  $y_i$  is the observed data value,  $\hat{y}_i$  is the predicted value from the fit, and  $w_i$  is the weighting applied to each data point in a set of  $n$  points.

$$SSE = \sum_{i=1}^n w_i (y_i - \hat{y}_i)^2 \quad (6)$$

- **$R_{square}$ :** measures how strong the fit is in explaining the variation of the data. In Equation 7,  $\bar{y}_i$  is the mean of the observed data.

$$R_{square} = 1 - \frac{\sum_{i=1}^n w_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^n w_i (y_i - \bar{y}_i)^2} \quad (7)$$

A curve fit using least squares regression is considered the best fit when SSE is close to zero, and the  $R_{square}$  has a value close to one [19]. With the above solution, the error determination and curve fitting for RMS error data modeling can be done, for the several resolutions and at several distances that the ZED cameras allow to capture data.

## 5 Experiments and Results

To determine the RMS error in the 3D data given by the ZED camera, a Nvidia Jetson TK1 board and a 3D pattern (checkerboard) were used in addition to the camera. The procedure is done for each of its capture resolutions (individual frames with  $2208 \times 1242$ ,  $1920 \times 1080$ ,  $1280 \times 720$ , and  $672 \times 376$  pixels). Also, the data collection experiment is repeated for several distances between the pattern and the camera, in the camera range (from 1 to 20 m), as shown in Figure 8. For each camera resolution, there were developed twenty experiments, what gives eighty experiments in total. These tests were done in an indoor environment with artificial light, achieving the maximum performance of the camera. Nonetheless, notice that it is well known that stereo vision works better for outside scenes where the texture is more present [3], so the results of this approach can be generalized to other scenes.

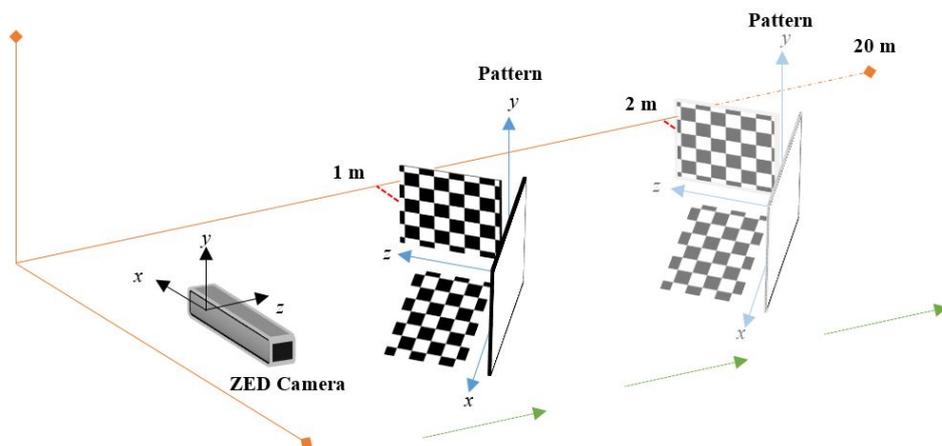


Figure 8: Captured data process for determination of the RMS depth error estimated by Stereolabs ZED.

## 5.1 Camera Settings

A helpful tool of the ZED camera is its Software Development Kit (SDK) that allows the developer to use optimized functions to obtain the maximum performance of the device. As seen above, by using this tool, it is possible to calibrate the device and to acquire rectified images and depth data [20].

In this proposal, the depth modes of the ZED camera are set to *STANDARD* and *QUALITY*. The *STANDARD* mode is a sensing mode that preserves distance metrics and shapes, but it contains holes due to visual occlusions and filtering. These holes are represented by black areas in the depth image. The *QUALITY* mode increase the level of accuracy and computational performance of the depth sensing module preserves edges and Z-accuracy while offering a higher reliability in areas with low texture. Also, the automatic calibration mode is activated, allowing the initial calibration parameters to be optimized according to the lighting and cameras position. All of these configurations are done to obtain depth data with the maximum possible precision in order to be able to determine the RMS error correctly.

## 5.2 The 3D test pattern

The test pattern used in this proposal is shown in the Figure 9. It is composed of three coupled chessboards with  $6 \times 7$ ,  $6 \times 4$  and  $7 \times 5$  inner corners, respectively. To achieve a robust detection of corners at long distances the size of the black and white squares is 50 mm. By employing this same pattern, it is obtained a set of 101 corners in each experiment.

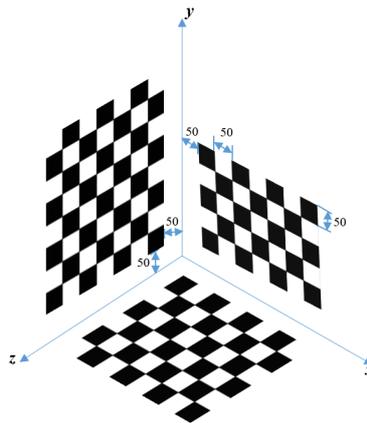


Figure 9: Chessboards size and corners distribution on 3D test pattern.

## 5.3 RMS Error Modeling

To define the best mathematical equations for representing the RMS error of each resolution of the ZED camera, a curve fitting is performed using linear, exponential, and polynomial models. After evaluating the graphical and numerical fittings through equations 6 and 7, it was empirically determined that the optimal fit for error data was the exponential model, which is shown in Figure 10. The fitting coefficients  $a$  and  $b$  are specified at Table 4. Here, it can be noticed that, for each resolution, the model has a component of a small random error (SSE close to zero), which the adjustment in each case explains between 96% and 97% of the total variation in the data on the mean ( $R_{square}$  values close to one). In summary, the RMS error may be calculated by using the equation 8 at any given depth  $Z$  for each ZED resolution.

$$f(Z) = a \exp^{(bZ)} \quad (8)$$

Resolution [px]	Coefficients	SSE	$R_{square}$
2208 × 1242	a = 0.01805; b = 0.1746	0.0026	0.968
1920 × 1080	a = 0.0106; b = 0.2215	0.0031	0.96
1280 × 720	a = 0.0184; b = 0.2106	0.0044	0.964
672 × 376	a = 0.0115; b = 0.2986	0.0004	0.97

Table 2: Numerical analysis of exponential fitting model.

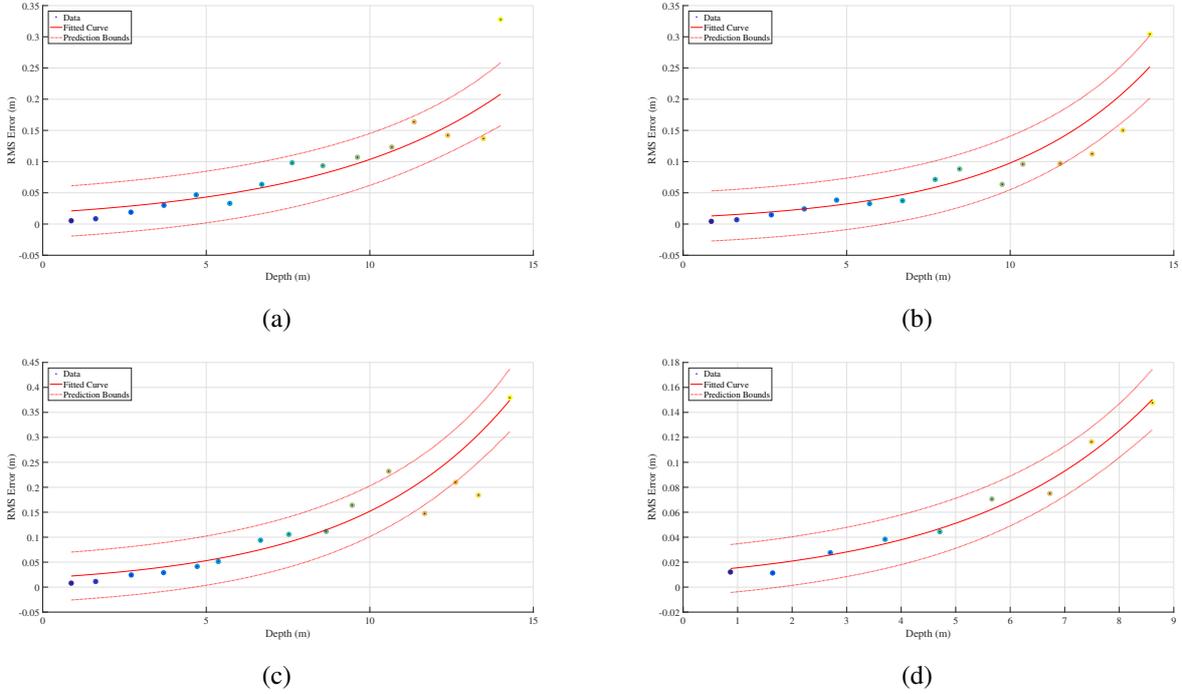


Figure 10: Exponential fitting curves for RMS error in depth data of the ZED camera, (a) 2208 × 1242 (b) 1920 × 1080 (c) 1280 × 720 and (d) 672 × 376 pixels. The dotted lines represent the data bounded prediction limits up to 95% of confidence.

By comparing the four curves presented at Figure 11, it can be seen that the ZED camera does not give 3D data with high variation in the amplitude of RMS error for distances up to seven meters. From eight meters and above there exists a rapid increase in the amplitude of the curve of 672 × 376 and 1280 × 720 pixels resolutions. This is due to the camera ceasing to deliver depth data for these resolutions at greater distances. On the other hand, the curves of 1920 × 1080 and 2208 × 1242 pixels remain with a minimum of amplitude variation between them, until thirteen meters.

Table 3 shows the mean and variance values of the RMS error for each resolution of the ZED camera, this overall error was calculated through the values estimated by the curves in Figure 10, the distances in which the sensor calculates the depth are considered.

Resolutions	$\mu[m]$	$\sigma^2[m]$
2208 × 1242	0.17	0.02
1920 × 1080	0.21	0.06
1280 × 720	0.30	0.12
672 × 376	0.83	1.54

Table 3: Mean and variance of the RMS error for all ZED resolutions.

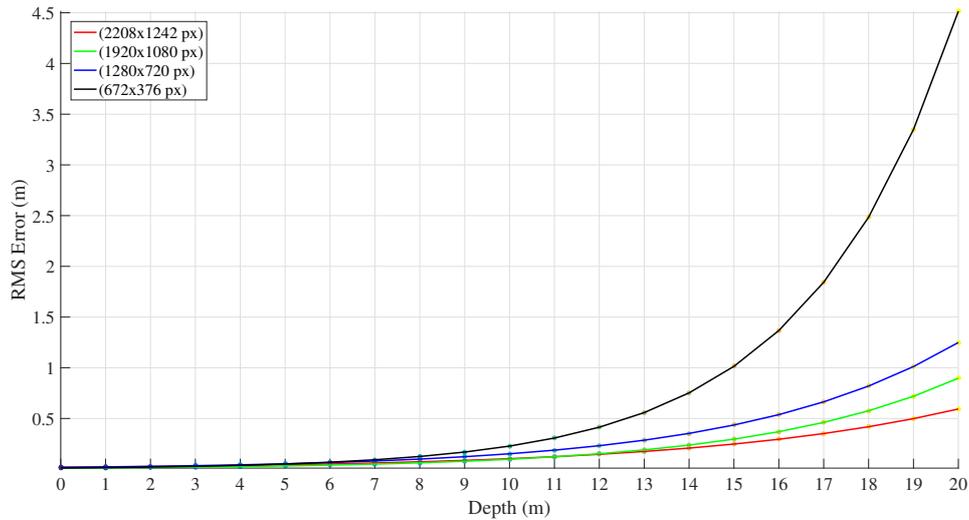


Figure 11: RMS error for all ZED resolutions (1-20 m).

## 5.4 Processing Time Analysis

The ZED camera provides three-dimensional data of a scene through depth maps and point clouds. For real-time applications, it is necessary to know how many of time this process takes. Times can vary depending on the resolution setting selected for the image capture. The average processing times for a test of one second of duration are shown in the Table 4. The four operating resolutions of the sensor are considered. Tests to capture such times were made using the Nvidia Jetson TK1 board with GPU Kepler (192 CUDA Cores) and 2 GB RAM.

Resolutions [px]	Frame-Rate [FPS]	Depth Map Time [ms]	Point Clouds Time [ms]
2208 × 1242	15	9.598	12.867
	30	7.407	9.640
1920 × 1080	15	7.301	9.551
	60	3.294	4.259
1280 × 720	30	3.277	4.252
	15	3.304	4.360
	100	1.130	1.327
672 × 376	60	1.063	1.254
	30	1.066	1.264
	15	1.070	1.279
	15	1.070	1.279

Table 4: Depth map and point cloud processing time for all ZED resolutions.

## 6 Conclusion

Depth data provided by stereo devices have errors attributed to several aspects related to cameras hardware and computer algorithms used to obtain these values. In real applications, such as an autonomous robot, it is important to consider and treat those errors in order to achieve correct decision-making process during a navigation task, for example.

In this paper, we proposed a mathematical modeling of the depth error determined for the ZED camera. Curves that model the estimated error were provided, and data on capture times and depth calculation were also analyzed. The obtained results can become initial information for the development of robotic and computational

vision applications that need as a useful parameter, the determined error. The introduced methodology can be followed by others stereo sensors, making it a general procedure.

In this particular case study, using the ZED camera, we conclude that the estimation of RMS error in-depth data of the ZED is directly related to the resolution of the camera. When this resolution is low, the depth capture distance is small, and RMS error is greater. It could also be thought that the detection of corners contributes to an additional error to the last RMS error. However, through the tests performed it is possible to verify and affirm that the detection algorithm is robust and detects corners at more than 20 meters in the maximum resolution of the ZED, resulting in a negligible detection error.

We noticed that in all the resolutions analyzed, from 15 meters, the camera begins to deliver some NaN depth data for some areas of the 3D testing pattern. This has a direct inference in the fact that at high distances between the pattern and the camera the RMS error goes quickly to infinity. In fact, in a short analysis from these experiments, mainly from Figure 11 and Table 3, it can be noticed that the ZED camera can be used with good precision up to some 16 meters. After that, the curves show errors that might not be acceptable depending on the application. For terrestrial robotics, for example, this would be considered an excellent working distance. However, for aerial robotics, this distance might have to be increased, mainly for fixed-wing aerial robots.

The resolution of operation to which the device has lower RMS error is  $2208 \times 1242$  pixels. However, the processing time in these images is greater. After analyzing RMS error and processing time in four resolutions, it can be concluded that for an application where it is sought to capture depth data up to thirteen meters, a capture resolution of  $1280 \times 720$  pixels is better, without a dramatic variation of the RMS error and with lower processing time with respect to resolution  $2208 \times 1242$  pixels. Similarly, in the case of applications where it is intended to use depth data up to seven meters, we could notice that the most appropriate resolution is  $672 \times 376$  pixels due to the little variation in the amplitude of the error and low processing time on the other resolutions.

## References

- [1] L. Ortiz, E. Cabrera, and L. Goncalves, "Sistema embarcado para reconstrução 3D em tempo real usando imagens estereo," in *Workshop of Works in Progress (WIP) in the 29th Conference on Graphics, Patterns and Images (SIBGRAP'16)*, J. P. Gois and M. Pamplona, Eds., São José dos Campos, SP, Brazil, Oct 2016.
- [2] X. Zou, H. Zou, and J. Lu, "Virtual manipulator-based binocular stereo vision positioning system and errors modelling," *Machine Vision and Applications*, vol. 23, no. 1, pp. 43–63, Jan 2012, doi: 10.1007/s00138-010-0291-y.
- [3] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. Prentice Hall, Mar 1998.
- [4] R. Szelisky, *Computer Vision: Algorithms and Applications*. Springer Link, 2012, doi: 10.1007/978-1-84882-935-0.
- [5] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 3812, pp. 133–138, Sep 2011, doi: 10.5194/isprsarchives-XXXVIII-5-W12-133-2011.
- [6] M. G. Diaz, F. Tombari, P. Rodriguez-Gonzalvez, and D. Gonzalez-Aguilera, "Analysis and Evaluation Between the First and the Second Generation of RGB-D Sensors," *IEEE Sensors journal*, vol. 15, no. 11, pp. 6507–6516, Nov 2015, doi: 10.1109/JSEN.2015.2459139.
- [7] M. Samir, E. Golkar, and A. A. A. Rahni, "Comparison between the Kinect V1 and Kinect V2 for respiratory motion tracking," in *2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, Oct 2015, pp. 150–155, doi: 10.1109/ICSIPA.2015.7412180.

- [8] H. Haggag, M. Hossny, D. Filippidis, D. Creighton, S. Nahavandi, and V. Puri, "Measuring depth accuracy in RGBD cameras," in *Signal Processing and Communication Systems (ICSPCS), 2013 7th International Conference on*, Dec 2013, pp. 1–7, doi: 10.1109/ICSPCS.2013.6723971.
- [9] W. Kazmi, S. Foix, G. Aleny, and H. J. Andersen, "Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison," *{ISPRS} journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 128 – 146, 2014, doi: 10.1016/j.isprsjprs.2013.11.012.
- [10] L. Sabattini, A. Levratti, F. Venturi, E. Amplo, C. Fantuzzi, and C. Secchi, "Experimental comparison of 3D vision sensors for mobile robot localization for industrial application: Stereo-camera and RGB-D sensor," in *Control Automation Robotics Vision (ICARCV), 2012 12th International Conference on*, Dec 2012, pp. 823–828, doi: 10.1109/ICARCV.2012.6485264.
- [11] D. Beltran and L. Basañez, *A Comparison between Active and Passive 3D Vision Sensors: Bumblebee XB3 and Microsoft Kinect*. Cham: Springer International Publishing, 2014, pp. 725–734, doi: 10.1007/978-3-319-03413-3\_54.
- [12] L. Fernandez, V. Avila, and L. Goncalves, "A Generic Approach for Error Estimation of Depth Data from (Stereo and RGB-D) 3D Sensors," 2017, doi: 10.20944/preprints201705.0170.v1.
- [13] S. D. Blostein and T. S. Huang, "Error Analysis in Stereo Determination of 3-D Point Positions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 6, pp. 752–765, Nov 1987, doi: 10.1109/TPAMI.1987.4767982.
- [14] C. Chang and S. Chatterjee, "Quantization error analysis in stereo vision," in *Conference Record of the Twenty-Sixth Asilomar Conference on Signals, Systems Computers*, Oct 1992, pp. 1037–1041 vol.2, doi: 10.1109/ACSSC.1992.269140.
- [15] C. Chang, S. Chatterjee, and P. R. Kube, "A quantization error analysis for convergent stereo," in *Proceedings of 1st International Conference on Image Processing*, vol. 2, Nov 1994, pp. 735–739 vol.2, doi: 10.1109/ICIP.1994.413668.
- [16] X. Song, F. Zhong, Y. Wang, and X. Qin, "Estimation of Kinect depth confidence through self-training," *The Visual Computer*, vol. 30, no. 6, pp. 855–865, 2014, doi: 10.1007/s00371-014-0965-y.
- [17] B. Choo, M. Landau, M. DeVore, and P. A. Beling, "Statistical Analysis-Based Error Models for the Microsoft Kinect Depth Sensor," *Sensors*, vol. 14, no. 9, p. 17430, 2014, doi: 10.3390/s140917430.
- [18] M. Kytö, M. Nuutinen, and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 78 640I–78 640I.
- [19] D' Agostino and Ralph B, *Goodness of fit techniques*. CRC press, 1986, vol. 68.
- [20] Stereolabs, "ZED Stereo Cameras Web Site," 2017, accessed: 2017-03-01. [Online]. Available: <https://www.stereolabs.com/>