FITTING A NORMAL PROBABILITY DISTRIBUTION TO DEPTH ESTIMATIONS OF THREE REALSENSE™ RGB-D CAMERAS TESTED IN SCENES WITH TRANSPARENCY

Eva Curto, Helder Araujo

Institute of Systems and Robotics, Dep. of Electrical and Computer Engineering, University of Coimbra, Portugal
(evacurto,helder)@isr.uc.pt

KEY WORDS: RGB-D cameras, Depth estimation, Low-cost 3D sensors, Active systems, Intel® Realsense™, Transparency

ABSTRACT:

In the last decade, various companies have released different versions of RGB-D sensors, improving their performance at various levels (resolution, frame rate, robustness). These devices can measure depth using one of the following optical technologies: Structured-Light, Active Stereoscopy or Time-of-Flight / Lidar. This paper aims to evaluate and compare the performance of three low-cost RGB-D cameras in the estimation of depth of a wall when transparent elements such as glass and water are added to the field of view. We propose an experimental setup for data acquisition involving an aquarium (empty and filled with water). The evaluation is based on the statistical distribution and dispersion of the Normal Probability Distribution estimated for each case.

1. INTRODUCTION

Low-cost RGB-D cameras have been widely used in robotics and computer vision due to their compactness and ability to perform real-time 3D reconstruction with a frame rate of 15-30 fps and resolutions up to 1280×720 (Carfagni et al., 2019), (Rosin et al., 2019). Intel® was one of the major tech companies that caught the wave of democratization of RGB-D cameras, launching the RealSense™ product line, specialized in the creation of consumer RGB-D cameras. RealSense™ offers several image sensing solutions, targeting different applications and contexts. RGB-D cameras are designed to obtain the shape of opaque surfaces through the analysis of the diffuse reflection components. However, not all surfaces exhibit an ideal light response, and so this type of camera is vulnerable to error sources such as scattering and absorption/attenuation of light. In order to evaluate and compare how these phenomena affect the depth estimation performance of RGB-D cameras, we designed an experimental assessment inspired by the assays of (Sarbolandi et al., 2015) and (Hansard et al., 2012). Transparent elements are added to the camera's field of view. Specifically, a large glass aquarium, first empty and then filled with water, is placed in front of the camera's target, which in this case is a simple wall. We compare the depth estimations of RealSense™ models D415, SR305, and L515, which operate with active depth sensing. Although all of them employ active technology, the operating principles and technology are different. The D415 relies on Active Stereoscopy (AS), SR305 uses Structured Light (SL) technology, and L515 works with Time-of-Flight (ToF) principle. Active Stereoscopy (the D415 model) comprises two ordinary cameras and an unstructured light pattern. Depth is estimated by triangulation using the disparity map between the two cameras. The pattern only serves to add artificial features to triangulate it. On the other hand, SR305 uses coded light technology to estimate depth. This means that there is a projector emitting one or more patterns sequentially onto the scene. These patterns are warped by the scene, reflected back, and captured by a standard camera. The depth of the scene is estimated by comparing the projected patterns and the distorted ones acquired by the camera. Finally, the LiDAR L515 is based on the Time-of-Flight principle (ToF). That is, the system estimates the depth by sending laser pulses on the scene and measuring the time between the laser pulse arriving at the target and returning to the receiver. The L515 camera uses an Edge-Emitter Laser (EEL) for the light source, a MEMS micro-mirror to scan the environment, a photodiode to measure the time of flight, and optical lenses to focus the beam. In order to evaluate and compare the depth estimates of the three cameras, a Normal Probability Distribution (NPD) was fitted to the depth estimates, enabling statistical characterization of the cameras response to each scenario.

2. MATERIALS AND METHODOLOGY

2.1 Experimental Setup

The experimental setup for this work consisted of a large glass aquarium with dimensions $0.84 \times 0.22 \times 0.58$ cm (and about 6mm thickness) placed on a table in front of a wall. On the opposite side of the wall and the aquarium, there was a tripod where the cameras were placed, facing the aquarium and the wall, with the optical axes approximately perpendicular to the wall. As mentioned in the Introduction, the cameras evaluated are the low-cost Intel® RealSense™ cameras D415(Intel® RealSenseTM, 2018), SR305 (Intel, 2016), and L515 (Intel® RealSenseTM, 2020), shown in Figure 1. All depth acquisition software was adapted from the C++ code samples provided by the Intel RealSense SDK, using their library *librealsense2*. The data acquisition routine was then run on an Ubuntu operating system. Due to the fact that the SR305 camera has only one unique resolution at $640\times$ 480, all cameras were set with the same resolution so that evaluation could be performed under the same conditions. Acquisitions were made in a controlled light environment, as the whole setup was in a dark room illuminated by an adjustable LED board.

In Figure 2, the three tests that comprise our experimental framework are illustrated. The *Wall* test, in Figure 2(a), is a starting test in which we acquire depth images directly from the wall, that is, without any transparency between the camera and the

(a) (b) (c)

Figure 1. Low-cost Intel® RealSense™ cameras: (a) D415 (b) SR305 (c) L515.

wall. These depth measurements will serve as a reference to the following tests since we do not have ground truth for the depth. As illustrated, the wall is approximately 63 cm away from the camera. Then Figure 2 (b) describes the *Empty* test, the aquarium is placed in front of the wall (9.7 cm away) to analyze the effect of the two transparent glass walls of the aquarium on depth estimation. The aquarium is filled with water (approximately 95 L) in the *Water full* test, as represented in Figure 2(c). This addition introduces, between the camera and the wall, one more transparent element in addition to the glass walls of the aquarium.

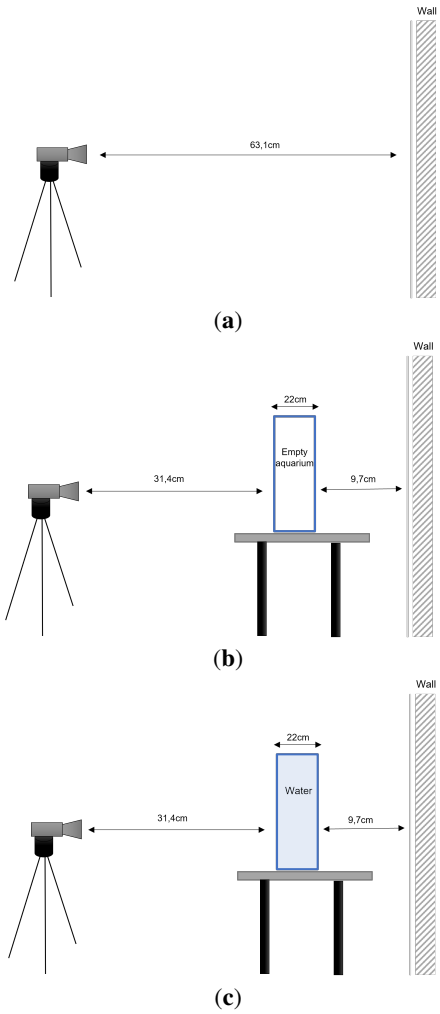

Figure 2. Experimental setup for the three tests: (a) *Wall* test; (b) *Empty* test; (c) *Water full* test.

2.2 Experimental Evaluation

For each of these three tests, 100 depth frames are captured and saved by each camera sequentially. Following the acquisition, the depth frames were converted into point clouds. To this end, we used the camera SDK to obtain the intrinsic parameters. Thus, having the pixel coordinates and the respective depth measurements d_m, it was possible to estimate the corresponding 3D points. The z coordinates correspond to the depth values measured at these points, i.e. $z = d_m$. The coordinates x and y were calculated as follows:

$$x = z \times \frac{u - pp_x}{f_x} \tag{1}$$

$$y = z \times \frac{v - pp_y}{f_y} \tag{2}$$

where (u, v) are the coordinates of the pixels, (pp_x, pp_y) are the coordinates of the principal point, and f_x and f_y are the focal lengths in pixel units.

All the following processing and analysis of the data were performed in MatLab (Mat, 2022). For each test, the information from the 100 point clouds was combined into a single one, where the depth of each 3D point (z coordinate) results from the average value of the 100 depth values corresponding to that point. To exclude non-stable depth values, we did not estimate the average depth for those points where less than 80 positive depths (out of 100 measurements) were obtained.

To analyze the distribution of the resulting average depth estimates throughout the tests, we modeled the probability distribution of our data. The Statistical and Machine Learning Toolbox (The Mathworks, 2022) allows us to fit a probability function to our data. In this case, observing the histograms of depth distribution, we empirically chose the Normal distribution to fit our data. A continuous random variable X follows a normal probability distribution if its probability function is expressed as

$$y = f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{3}$$

where μ is the population mean and σ^2 is the population variance (McClave and Sincich, 2012). The parameters μ and σ^2 of the normal distribution were estimated from the sample mean and sample variance, respectively, using maximum likelihood estimation (MLE).

3. RESULTS

The results of the fitting of an NPD (Normal Probability Distribution) to the average depth estimates are presented next. In Figures 3, 4 and 5 we can observe the histograms that represent the distribution of the average depth estimates (in relative frequency) of each camera for every test. Overlaid on each histogram is a red line that describes the adjusted NPD for each case. Associated to each estimated NPD, we have μ and σ parameters that can be seen in Table 1. In order to obtain a statistical overview of the NPD estimated to our data, additional statistical information is graphically described in the form of box charts. In Figures 7, 8 and 9, we can observe the change in terms of distribution and spreading of depth estimates for each camera, comparing the three tests. In addition, the percentage change of the mean (μ) relative to the first test (*Wall*) is also displayed. A comparative visualization between cameras for the same tests is also given by Figures 10, 11 and 12. In these graphs, it does not make sense to show the percentage change of the mean since there is no reference value.

In Figure 6, we give the legend for the information displayed in the box charts. The box charts provide a visual representa-

tion of the summary statistics of the fitted normal probability distributions, allowing us to understand and compare how our data are distributed and spread out. The statistical information provided is the following:

- Median - The median is represented by the horizontal line in the middle of the notch, dividing the two boxes.
- Interquartile range (IQR) - The IQR is the distance between the top and bottom edges of the box. The top edge represents the upper quartile (corresponds to the 0.75 quantile) and the bottom edge represents the lower quartile (corresponds to the 0.25 quantile).
- Outliers - Outliers are represented by a blue 'o' symbol and refer to values that are more than $1.5 \cdot IQR$ away from the top or bottom of the box.
- Whiskers - There are two whiskers in each box. One whisker connects the upper quartile to the nonoutlier maximum (the maximum value that is not an outlier), and the other connects the lower quartile to the nonoutlier minimum (the minimum value that is not an outlier).
- Notch - The notch is the tapered, shaded region around each median that displays the confidence interval. The top and bottom edges of the notch region correspond to $m + (1.57 \cdot IQR)/\sqrt{n}$ and $m - (1.57 \cdot IQR)/\sqrt{n}$, respectively, where m is the median, IQR is the interquartile range, and n is the number of data points.

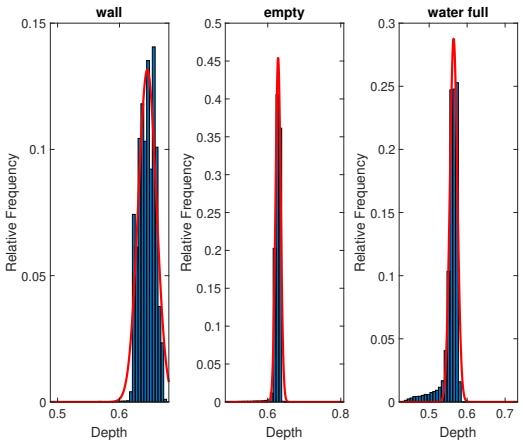

Figure 3. Normal distribution fitting of the depth estimates for D415 acquisitions.

4. DISCUSSION

Given the results obtained by the NPD fitting and statistical characterization, the effect of each test can be evaluated and compared for each camera. In general, for the three cameras, the shape and size of the fitting NPD change when transparent elements are added to the scene, which naturally translates into a different statistical characterization. When we compare the three cameras in the *Wall* test, it is noticeable that the central tendency concerning the depth estimates is quite identical among the three cameras. Thus, we can infer that the starting point of the three cameras will be approximately the same with respect to the accuracy of the estimates (taking into account the type of experiment we are going to perform). On the other hand, the reference test indicates that under the so-called ideal conditions, the L515 camera has less variability. In the *Empty* test, minor changes due to the transparency of the aquarium glass

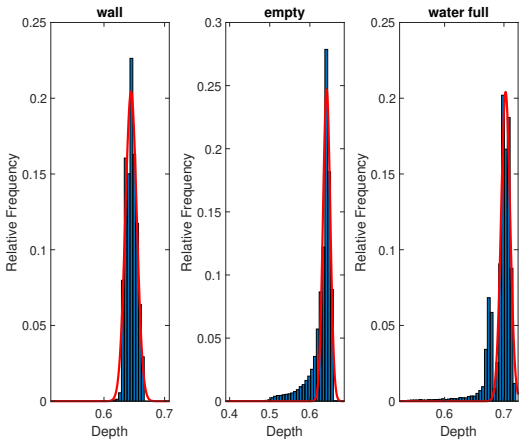

Figure 4. Normal distribution fitting of the depth estimates for L515 acquisitions.

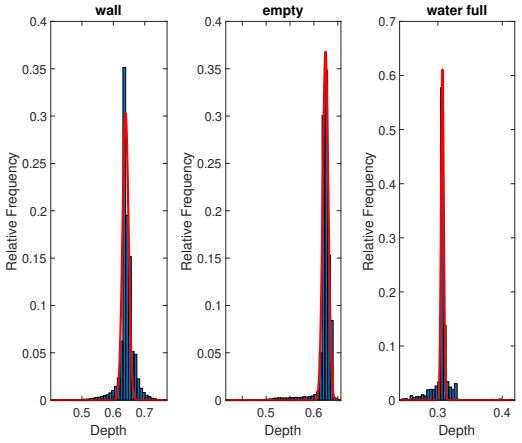

Figure 5. Normal distribution fitting of the depth estimates for SR305 acquisitions.

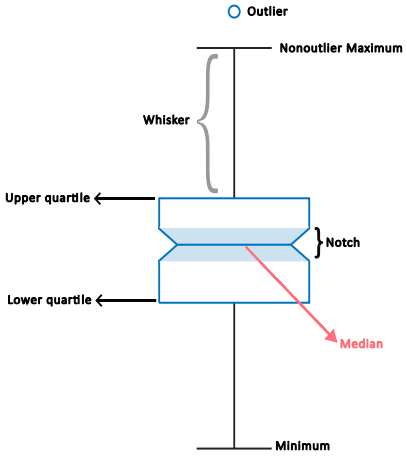

Figure 6. Legend of the information in the box charts.

walls are noticeable. The central tendency of the three cameras is slightly offset. More precisely, the percentage change of the

Table 1. Estimated parameters resulting from fitting to a normal distribution. μ - Mean and σ - Standard deviation.

D415			
	Wall	Empty	Water full
μ	0.6443m	0.6265m	0.5583m
σ	0.0130	0.0110	0.0214
L515			
	Wall	Empty	Water full
μ	0.6450m	0.6282m	0.6938m
σ	0.0089	0.0308	0.0220
SR305			
	Wall	Empty	Water full
μ	0.6415m	0.6220m	0.3063m
σ	0.0239	0.0168	0.0136

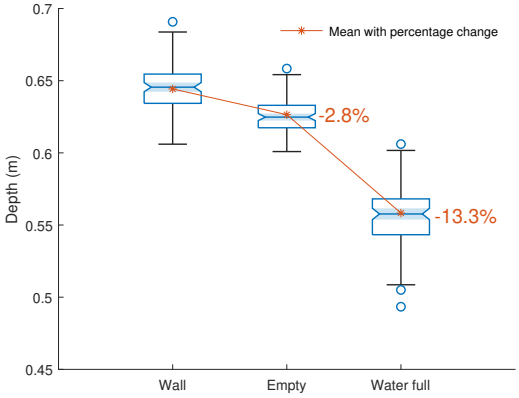

Figure 7. Normal Distributions - D415

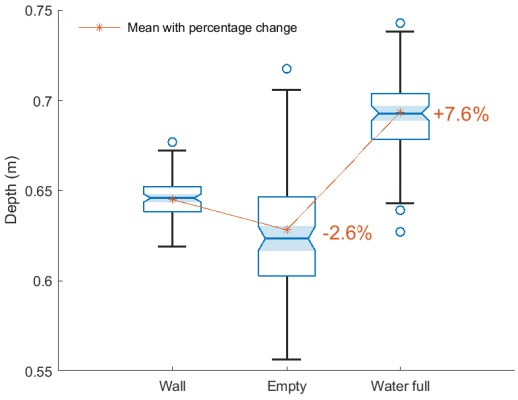

Figure 8. Normal Distributions - L515

mean is -2.8%, -2.6% and -3.0% for the D415, L515 and SR305, respectively. Therefore, the three cameras continue to display similar behavior between them. However, the variability of the L515 camera has increased substantially as it is visible from the σ value and the Interquartile range. A significant difference was found in the results with respect to the *Water full* test. The central tendency of the three cameras changes significantly. The mean value for D415 changes to 0.56 m, translating into a percentage change of -13% relative to the reference value (*Wall* test). On the other hand, the mean for L515 increases, having a percentage change of $+7\%$. This increase results in an unrealistic mean value of depth estimates of 0.69m (since the wall is at about 0.63m). The erroneous estimation of depth by L515 when the water is added to the aquarium is not particularly surprising, given that L515 works with the Time-

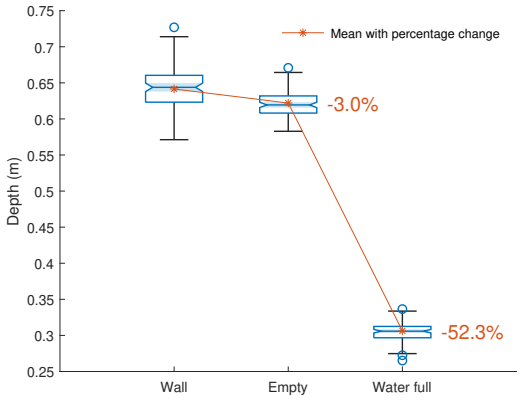

Figure 9. Normal Distributions - SR305

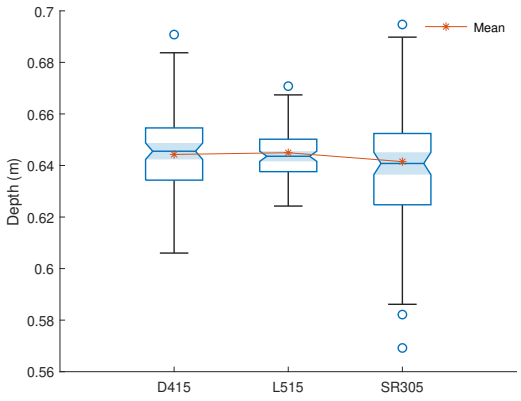

Figure 10. Normal Distributions - Wall test

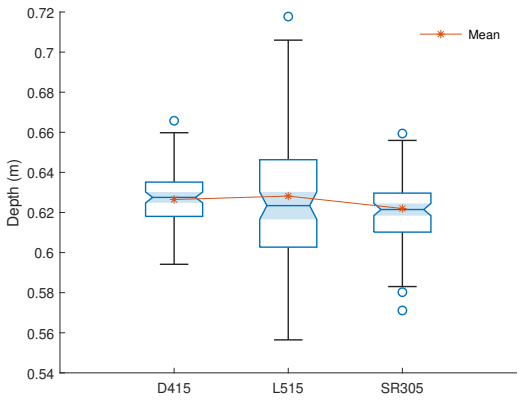

Figure 11. Normal Distributions - Empty test

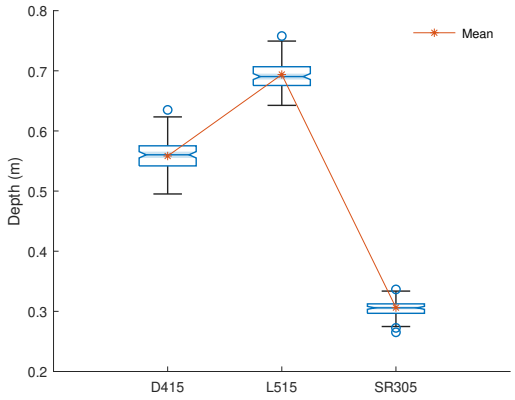

Figure 12. Normal Distributions - Water full test

of-Flight (ToF) principle. This principle relies on the speed of light in vacuum that is equal to 2.99×10^8 m/s. When we add transparent elements such as glass or water, we must take into account that the light traveling between the camera and the target is slowed down to 2.26×10^8 m/s in water and 2.00×10^8 m/s in glass. In addition attenuation of light also occurs. Another phenomenon that is present in media such as water is refraction. Since the speed of light is different in different materials refraction occurs, and therefore the path of the refracted light is different from that of the incident light. In summary, the flight time will be longer and the estimated distance will be greater than the actual distance. In addition to the L515 camera, the estimates from the SR305 camera have also been greatly affected, albeit in a opposite way. That is, the percentage change of the mean for SR305 in the *Water full* test is $-52,3\%$, which is the largest percentage change. The mean value of 0.31m suggests that the SR305 is estimating the depth of the first wall of the aquarium. Regarding the spreading of depth estimates, the results need to be interpreted with caution, because as stated in the method, non-stable depth points were excluded. In this way, when we look at the reduced variance of the SR305 camera, we have to consider that the number of depth estimates compared to the other cameras is much lower. It should also be taken into consideration that in the *Water full* test many more points were excluded.

5. CONCLUSION

This paper has investigated the effect of transparent elements such as glass and water on depth estimation performance for three different low-cost RGB-D cameras with different operating principles. We have presented an experimental framework to evaluate and compare RealSense™ models D415, SR305, and L515 in a scenario in which the target, a wall, is observed across the glass walls of an aquarium and then also by the water that fills it. One hundred depth frames were acquired for each test and converted into point clouds to facilitate the processing of depth data. Then, for each test, the depth frames were combined into a unique point cloud by averaging the depth values and removing non-stable depth measurements. The evaluation and comparison of the depth estimates have been performed in terms of the statistical analysis of the depth distribution. For that purpose, we modeled our data by a probability distribution. The Normal Probability Distribution (NPD) was chosen to fit the depth estimates. In the beginning of the experiment, in the *wall* test, the NPD of three cameras has a close mean, but in the following experiments the NPD of each camera changes in different ways. Our experiments confirm that the operating principles affect depth estimation differently. The most significant evidence is the behavior of the L515 camera in the *water full* test, in which the time-of-flight measurement is affected by the conditions of light propagation in the water (attenuation and refraction). Overall, the results obtained demonstrate that the D415 camera produces better depth estimates in the case of transparent objects.

6. ACKNOWLEDGMENTS

This research was partially supported by Project COMMANDIA SOE2/P1/F0638, from the Interreg Sudoe Programme, European Regional Development Fund (ERDF), by Fundação para a Ciência e a Tecnologia (FCT) through the Ph.D. Scholarship 2021.06127.BD. and also by Fundação para a Ciência e a Tecnologia (FCT) under the project UIDB/00048/2020.

REFERENCES

Carfagni, M., Furferi, R., Governi, L., Santarelli, C., Servi, M., Uccheddu, F., Volpe, Y., 2019. Metrological and critical characterization of the intel D415 stereo depth camera. *Sensors (Switzerland)*, 19(3).

Hansard, M., Lee, S., Choi, O., Horaud, R., 2012. *Time of Flight Cameras : Principles , Methods , and Applications*. Springer Science & Business Media.

Intel, 2016. Intel ® RealSense ™ Camera SR300 Embedded Coded Light 3D Imaging System with Full High Definition Color Camera. Technical Report June.

Intel® RealSenseTM, 2018. Product Family D400 Series Datasheet. Technical Report November.

Intel® RealSenseTM, 2020. LiDAR Camera L515 user guide. Technical Report July.

Mat, 2022. MATLAB version 9.12.0.2039608 (R2022a) Update 5.

McClave, J. T., Sincich, T. T., 2012. *Statistics*. Pearson, Boston.

Rosin, P. L., Lai, Y.-K., Shao, L., Liu, Y., 2019. *RGB-D Image Analysis and Processing*. Springer Nature.

Sarbolandi, H., Lefloch, D., Kolb, A., 2015. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Computer Vision and Image Understanding*, 139, 1–20. http://dx.doi.org/10.1016/j.cviu.2015.05.006.

The Mathworks, I., 2022. Statistics and machine learning toolbox.