Morphological parameters assessment with a depth camera based measurement system

Olivera Tomašević, Luka Mejić, Darko Stanišić, Nikolina Maravić

***Abstract*—Lately, depth cameras are being widely used for computer vision applications such as human pose estimation, activity recognition, object and people tracking, 3D mapping and localization. Possibilities for depth sensing integration in economy sectors like agriculture and healthcare services are growing as researchers are stating numerous advantages in contactless measurements done by robust low-cost depth camera systems. In this paper, we discuss an application of two different depth-sensing technologies for morphological parameters assessment.**

***Index Terms*— Morphological parameters, active depth-sensing technology, Intel RealSense.**

I. Introduction

Since computer vision has largely been concerned with obtaining 3D information from 2D images, the invention of low-cost commercial depth sensing cameras made a significant contribution to solving computer vision problems. Derived 3D depth map allowed segmentation of the scene into foreground and background, which facilitated identification and tracking of simple objects. This opened new opportunities for many practical computer vision applications such as human pose estimation [1], activity recognition [2], object and people tracking [3], 3D mapping and localization [4].

Depth maps and resulting 3D reconstructions of objects acquired in real time also enabled non-contact measurement of objects' morphological properties. Researchers have been assessing objects' static properties such as size [5] and volume [6], and dynamic such as breathing patterns [7] in conditions of clinical interest. With this application, depth sensing is likely to gain significant role in economic sectors such as agriculture and healthcare services, but also in industrial and clothing design, and ergonomics.

Namely, in [8] it was concluded that when it came to body measurement, depth camera acquired results close to real data, and as a more affordable method, it could be an appropriate alternative to high cost laser scanner. In similar fashion, in [9] the depth camera based body measurement was shown to be an efficient approach for anthropometric data collection, and in [10] it was demonstrated that low-cost depth camera based system could be used for rapid and robust tracking of body shapes and anthropometric changes in children.

Contactless measurement of body size and volume has also been described as an innovative tool for making management decisions for livestock. In [11], the authors demonstrated an accurate body measurement system based on Microsoft Kinect v2 camera, which was applied to livestock. They successfully validated the quality of generated model against manually measured body references such as height, depth, length, and girth of certain body parts. Similarly, in [12], the authors presented the measurement of body parameters such as linear and integral characteristics along directional lines and local areas, geodesic distances and perimeters of cross sections. While in [13], the measurement system was also based on Microsoft Kinect sensor, the aim was the livestock's growth assessment through the repetitive analysis of 3D models of their certain body portions.

In a similar manner, and in order to optimize the harvesting by the growth state and crop yield assessment, researchers have been measuring structural parameters of fruits and vegetables. For instance, from the segmented lettuce point clouds in [14], they extracted the volume, surface area, leaf cover area, height predictors, and correlated them to the fresh weight. Analysis showed that the calculated surface areas correlated strongly with measured fresh weight. With the similar aim, in [15] the authors used a Microsoft Kinect sensor and reported that accuracy of the 3D models of cauliflower crops deviated from the ground truth measures by less than 2 cm in diameter/height, while the fruit volume estimation showed an error below 0.6% overestimation.

When it comes to expected measurement accuracy, it immensely depends on the choice of depth-sensing technology and application context, particularly lighting conditions and working distance. Namely, depth-sensing technology has been developing for many years in several different forms. Stereo [16] may be the most basic approach for acquiring 3D depth maps. It is a "passive" depth sensing method that uses two or more standard RGB cameras, and calculates depth by finding correspondences between image points. In contrary, "active" [17] approaches use their own light sources, and can potentially overcome the limitations of passive stereo, such as large textureless surfaces and low light levels.

As revised in [18], on the market there are several representatives of state-of-the-art active depth-sensing solutions. Intel RealSense product line is one of them. It supports active depth-sensing technologies that marked the last decade: structured (coded) light (SL), active IR stereo (AIRS), and time-of-flight (TOF). In this paper, we discuss the measurement of morphological parameters using two

Olivera Tomašević is with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: olivera.tomasevic@uns.ac.rs).

Luka Mejić is with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: mejic@uns.ac.rs).

Darko Stanišić is with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: darkos@uns.ac.rs).

Nikolina Maravić is with the Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000 Novi Sad, Serbia (e-mail: nikolina.maravic@uns.ac.rs).

 ISBN 978-86-7466-930-3

different Intel RealSense solutions, which we find suitable for close range applications.

The paper is organized as follows. First, we describe the acquisition system and software, then we look at the expected accuracy depending on the chosen device, and in the end, we propose methods for morphological parameters assessment based on a point cloud data.

II. THE METHOD

A. Acquisition system

Namely, Intel RealSense has produced depth cameras through three different series: SR300, D400, and L500. Each series represents a distinct depth sensing technology. What separates them is their ideal operating parameters, along with differences in how accurate the depth information is in different situations [19]. As SR300 and D400 series allow measurements of similar distance ranges, we opted to incorporate a representative of each in our acquisition system, independently.

SR305 camera is Intel RealSense's only representative of SR300 series [20], and thus it's only representative of structured light technology. The camera uses patterned light to determine depth within the scene, thus it is ideal for use in indoor, controlled lighting situations. It is also intended for very short-range applications under one meter, but its operating range goes even up to 2m. It utilizes rolling shutter technology, has a field of view (FOV) of 71.5° x 55° and will provide up to 640×480 resolution depth images at 60 fps.

D435i camera is one of Intel RealSense's representatives that belong to D400 series [21]. It is based on active IR stereo, which uses stereo vision to calculate depth, but also projects an invisible infrared pattern to improve depth accuracy in scenes with low texture. Taken that it is based on stereovision, it is expected to perform equally well in both indoor and outdoor environments. It has an ideal operating range of 0.3m to 3m. It utilizes global shutter technology, has a wide field of view of 85.2° x 58°, and provides up to 1280 x 720 resolution depth images at 90 fps. Since depth camera D435 is another representative in the series that only differs in IMU presence, in the following text we will use the term D435, as equally.

For the purposes of visualization, Intel has developed specialized tools such as Intel RealSense Viewer and Intel RealSense Depth Quality Tool, but has also developed an open source SDK [22] that supports various platforms and programming languages, thus enabling us to build applications for data acquisition and processing adapted to our needs.

B. Acquisition software

Software for data acquisition on PC is written in Matlab (ver. R2021a, MathWorks, USA). It allows acquisition parameters setting, acquisition and processing, and visualization.

Acquisition parameters refer to stream configuration and depth property settings. These parameters are supported by the SDK library, which allows the configuration of the camera with a number of internal settings.

Stream configuration implies settings like stream type, stream format, stream resolution, and sampling frequency. Stream types refer to different types of data provided by RealSense devices. In order to get point cloud that contains RGB information for every data point, our software allowed acquisition of both, depth frame, i.e. data from depth sensor, and color frame, i.e. image from embedded RGB camera. These were thereafter aligned in order to get final point cloud representation of the scene. Available stream format, which identified binary data encoding within a frame, was set to a predefined value, which depended on a type of a stream.

Sampling frequency can take several different predefined values. In cases of dynamic parameter assessment, the sampling frequency value also determined the time resolution of dynamic morphological parameters. Thus, in cases when official documentation did not recommend any value that would ensure good quality depth, we set the highest possible value that still allowed depth and RGB frame alignment. The same applied to stream resolution parameters.

In order to get good quality depth data, we set specific values to a combination of depth property parameters such as depth projector power, accuracy, filter option and depth preset setting. The values were set specifically to achieve as better depth accuracy as possible.

Depth preset setting is a working mode that implies predefined settings for all the depth property parameters that official documentation offers as optimal for certain conditions. Considering that we were after high accuracy mode, if it was supported for given device, we opted for high accuracy preset.

In case of SR305 device, depth accuracy was set through the depth property parameter of the same name, which defined the number of patterns projected per frame. In case of D435 device, high depth accuracy was achieved through the high accuracy preset, which set high confidence threshold value of depth.

Intel SDK library includes post-processing filters to enhance the quality of depth data and reduce noise levels. When done in real time, it negatively affects the temporal resolution of the acquired frames, which should be taken in consideration when assessing dynamic parameters of an object.

As recommended in [23], images acquired with SR305 camera were filtered with default depth filter, which had moderate smoothing effects.

C. Expected accuracy

As said before, expected measurement accuracy immensely depends on application context, particularly lighting conditions and working distance. However, even when these conditions are met in accordance to camera specifications, 3D reconstructed model is still very likely to contain inaccurate, i.e. dimensions that differ from ground truth.

Depth error of SR305 camera is described in [23]. The systematic distance inhomogeneity in case of this camera brings two main systematic error components increasing with distance: non-planarity and depth offset. Non planarity can be explained through error map between the target plane and multiple point clouds acquired with increasing distance from the plane itself, while depth offset can be approximated by a parabolic function with a high coefficient of determination. For instance, for distance of 65 cm, depth offset is somewhat larger than 12 mm.

The error inherent in SR305 camera acquisition is perhaps best portrayed through evaluation of the error introduced in

 ISBN 978-86-7466-930-3

measuring a flat surface that compares SR305 camera model with D415 camera model from D400 series. The comparison can be found in [24]. Namely, D415 has a considerably lower error both in terms of planarity and of distance accuracy.

D435 camera is, on the other hand, a lot more similar to D415 camera - they use the same vision processor to provide RGB-D data. The main difference lays in depth quality, which is usually portrayed through RMS error - the depth noise for a localized plane fit to the depth value. Comparison between the two models can be found in [25]. Namely, the depth RMS error scales as the square of the distance away from the camera, in both cases. However, due to the wider FOV, the smaller baseline, and the smaller sensor resolution, the D435 has more depth noise at any given range. At the same time, this model benefits from having smaller minimum operation distance, which allows acquisition at shorter distances, and global, as opposed to rolling, shutter sensor, so is expected to be more accurate when dealing with dynamic scenes.

D. Data acquisition and processing

The acquisition was performed indoors under artificial lighting conditions. Chosen camera was fixed at a distance of about 50 cm from the horizontal table top on which objects of interest were placed. It was taken into account that placed object should be in the camera's field of view and that all its interest points should also be within the recommended range of distances for a given camera type.

In the case of SR305 camera, RGB-D images were acquired in resolution of 640x480, and with sampling frequency of 60 Hz, and in the case of D435, they were acquired in resolution of 848x480, and with sampling frequency of 30 Hz. Upon the start of acquisition, the first couple of frames were discarded to allow the camera time to settle.

The steps we found useful in the process of parameters extraction were the following:

- points selection within a region of interest in the point cloud so as to lower the computational recources needed for further processing steps
- potential selection of markers that would be used for parameter assessment
- potential rigid transformations so as to align the direction of markers propagation with a certain axes of the camera local coordinate system
- potential plane fitting through the table surface so as to assess camera tilt relative to the table, and perform appropriate tilt correction through rigid transformations
- calculation of the distance between markers based on Euclidean distances between the each two nearest adjacent points
- calculation of surface area in a region of interest based on Delaunay triangulation [26] of point cloud data
- extraction of cross sections defined by the markers trajectory
- calculation of the cross-section area
- in the cases of dynamic parameters assesment, tracking of marker koordinates over succesive frames

III. RESULTS

The point cloud acquisition and calculation of morphological parameters is demonstrated on the cropped RGB-D image of a human body torso. Fig. 1 contains RGB-D image acquired with SR305 (on the left), and with D435 camera (on the right), respectively.

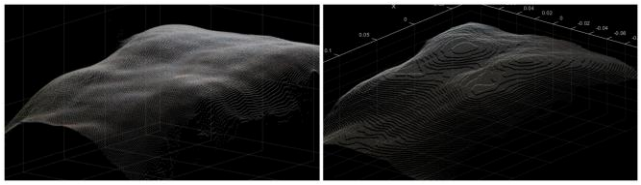

Fig. 1. Cropped RGB-D image of a human body torso captured with SR305 (on the left) and D435 (on the right).

In the case of D435, different levels of depth values are obvious. More specifically, we acquired depth maps with resolution of 1mm, which is a default depth unit of underlying vision processor. In the case of SR305, depth unit is a lot smaller (0.000125 m), thus its depth resolution is higher, as can also be noted in the figure. Other than on visual representation, these differences did not have any significant affect on parameters assessment methods. They are processed in the same manner, and results are stated in the following figures.

Fig. 2 contains example of static morphological parameters derivation based on trajectories in certain directions. Resulting cross-sections are visible in the point cloud (on the left), and isolated so that their length could be assessed (on the right). Top row refers to measurements based on SR305 camera, while bottom row refers to D435 camera. As previously stated, the cross-section lengths were estimated by the calculation of Euclidian distances between the each two nearest adjacent points.

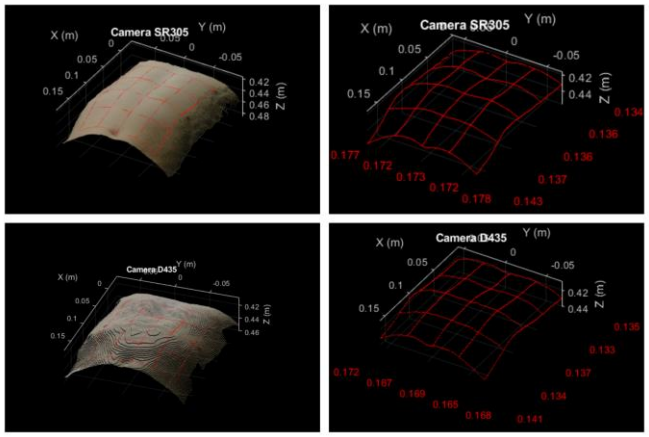

Fig. 2. Static morphological parameters derivation: cross-sections marked in a point cloud (on the left), and extracted cross-sections (on the right) along with their calculated dimensions.

Due to the anatomical planes in which they are located, these cross-sections are referred to as transverse and sagittal cross-sections, respectively. They depended on the marker positions, which were set on nipples and navel, by the user selection. Transverse cross-sections were defined based on direction on which the nipples were lying, and they were equidistantly distributed between the level of the nipples and level of the navel. Sagittal cross-sections were perpendicular with respect to them. The outermost positions of the sagittal cross-sections were left and right nipple location, respectively.

 ISBN 978-86-7466-930-3

These measurements were rounded to a precision of 1 mm. As can be seen on Fig. 2, there are certain differences in the cross-section length measurements that depend on the used camera. They are most prominent in sagittal sections and amount the most in the case of difference of 10 mm in favor of SR305 camera for the outermost left sagittal cross-section (from the perspective of the respondent). We assume that significant part of this difference lays in depth-offset error of camera SR305, which reportedly [23] takes out around 4 mm for the object distance of 400 mm, and 5.433 mm for the object distance of 450 mm.

Extracted cross-sections and their dimensions may indicate local torso characteristics. On the other hand, the object, the torso in this case could also be described globally by parameters such as surface area and volume. An example of surface area calculation is given in Fig. 3. Figure represents marked surface in the region of interest defined by the same markers that defined cross-sections: locations of nipples and navel. The area was estimated on tridimensional polygon mesh generated with Delaunay triangulation applied to the point cloud data illustrated in Fig 4. The resulting surface area is calculated upon all the areas of triangles at the base of the resulting mesh.

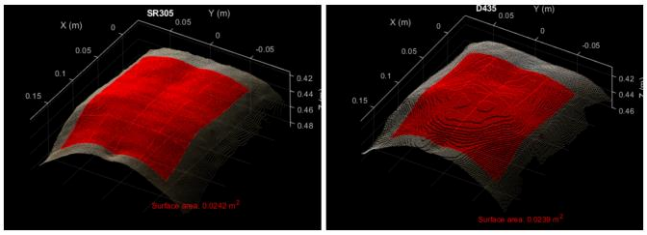

Fig. 3. Static morphological parameters derivation: marked surface in the region of interest – SR305 image (on the left), and D435 image (on the right) along with their respective results.

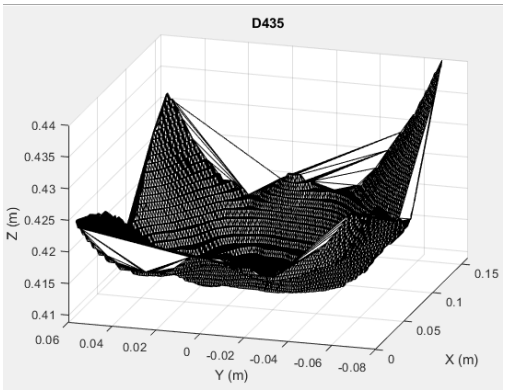

Fig. 4. Tridimensional polygon mesh calculated on the point cloud region of interest. This specific example refers to RGB-D image acquired with camera D435.

As Fig. 3 indicates, the image from SR305 camera resulted in slightly higher value of surface area parameter – 0.0242 m^2 as opposed to 0.0239 m^2 that resulted from D435 acquisition. This is in accordance with the conclusion regarding the length of the cross-sections acquired with these cameras.

Considering that demonstrated parameters can directly allow estimating of the dimensions of either torso or its parts, non-contact measurements of this type could have a purpose in clinical practice.

IV. Discussion

Results above show that visual representation of the scene did not vary significantly depending on the used camera. However, there are certain differences in their performance. Namely, results confirm the higher resolution of camera SR305 compared to D435 model (see Fig. 1). Considering that SR305 is capable of detecting depth with a higher resolution than D435, if the morphological characterization of fine details within small samples was at task, it could be expected that SR305 would be able to better reconstruct the object model, and thus would allow drawing contours that might not be visible in the acquisition with D435 camera.

On the other hand, SR305 has non-planarity and depth offset issues that contraindicate its use in precise measurement systems. As previously stated, in terms of planarity and distance accuracy, SR305 is outperformed by D400 series representatives.

When it comes to D400 series, D415 has more precision than D435. This regularity is accounted to the larger field of view of D435 camera. However, considering that D435 is better suited for dynamic scenes than D415, it could be expected that both these representatives of D400 be equally used in applications concerning morphological parameters assessment.

V. Conclusion

The aim of this study was to demonstrate derivation of morphological parameters with different depth sensing technologies. We discussed acquisition parameters that would allow good depth in static, as well as in dynamic scenes, and gave an example of static parameters calculation that could have a purpose in clinical practice.

As we discussed two different acquisition systems, we can conclude that for this type of application, one should use D400, rather than SR300 series, as it is more likely to provide sufficiently reliable measurements for the task at hand.

References

[1] Marin-Jimenez, M.J., Romero-Ramirez, F.J., Munoz-Salinas, R. and Medina-Carnicer, R., 2018. 3D human pose estimation from depth maps using a deep combination of poses. *Journal of Visual Communication and Image Representation*, *55*, pp.627-639.

[2] Park, S.U., Park, J.H., Al-Masni, M.A., Al-Antari, M.A., Uddin, M.Z. and Kim, T.S., 2016. A depth camera-based human activity recognition via deep learning recurrent neural network for health and social care services. *Procedia Computer Science*, *100*, pp.78-84.

[3] Zhou, Q.Y. and Koltun, V., 2015. Depth camera tracking with contour cues. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 632-638).

[4] Schubert, S., Neubert, P. and Protzel, P., 2017, July. Towards camera based navigation in 3d maps by synthesizing depth images. In *Annual Conference Towards Autonomous Robotic Systems* (pp. 601-616). Springer, Cham.

[5] Vo-Le, C., Van Muoi, P., Son, N.H., Van San, N., Duong, V.K. and Huyen, N.T., 2021, January. Automatic Method for Measuring Object Size Using 3D Camera. In *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)* (pp. 365-369). IEEE.

[6] Dellen, B. and Rojas Jofre, I.A., 2013. Volume measurement with a consumer depth camera based on structured infrared light. In *Proceedings of the 16th Catalan Conference on Artificial Intelligence, poster session* (pp. 1-10).

[7] Kotoku, J.I., Kumagai, S., Uemura, R., Nakabayashi, S. and Kobayashi, T., 2016. Automatic Anomaly Detection of Respiratory Motion Based on Singular Spectrum Analysis. *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*, *5*(01), p.88.

 ISBN 978-86-7466-930-3

[8] Yüksel, H. and Oktav, M.B., 2020. Analyses of body measurement with depth image data using motion capture sensor. *Industria Textila*, *71*(6), pp.530-537.

[9] Lin, Y. L., Wang, M. J. and Wang, B., 2015. Body dimension measurements using a depth camera. *New ergonomics perspective* (pp. 367-371), CRC Press/Balkema.

[10] Park, B.K., Lumeng, J.C., Lumeng, C.N., Ebert, S.M. and Reed, M.P., 2015. Child body shape measurement using depth cameras and a statistical body shape model. *Ergonomics*, *58*(2), pp.301-309.

[11] Ruchay, A., Kober, V., Dorofeev, K., Kolpakov, V. and Miroshnikov, S., 2020. Accurate body measurement of live cattle using three depth cameras and non-rigid 3-d shape recovery. *Computers and Electronics in Agriculture*, *179*, p.105821.

[12] Ruchay, A.N., Dorofeev, K.A., Kalschikov, V.V., Kolpakov, V.I. and Dzhulamanov, K.M., 2019, October. A depth camera-based system for automatic measurement of live cattle body parameters. In *IOP Conference Series: Earth and Environmental Science* (Vol. 341, No. 1, p. 012148). IOP Publishing.

[13] Pezzuolo, A., Guarino, M., Sartori, L. and Marinello, F., 2018. A feasibility study on the use of a structured light depth-camera for three-dimensional body measurements of dairy cows in free-stall barns. *Sensors*, *18*(2), p.673.

[14] Mortensen, A.K., Bender, A., Whelan, B., Barbour, M.M., Sukkarieh, S., Karstoft, H. and Gislum, R., 2018. Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Computers and Electronics in Agriculture*, *154*, pp.373-381.

[15] Andujar, D., Ribeiro, A., Fernandez-Quintanilla, C. and Dorado, J., 2016. Using depth cameras to extract structural parameters to assess the growth state and yield of cauliflower crops. *Computers and Electronics in Agriculture*, *122*, pp.67-73.

[16] Flusser, B.Z.J.: Image registration methods: a survey. Image Vis. Comput. 21(11), 977–1000 (2003)

[17] Krig, S., 2014. *Computer vision metrics: Survey, taxonomy, and analysis* (p. 508). Springer nature.

[18] Giancola, S., Valenti, M. and Sala, R., 2018. *A survey on 3D cameras: Metrological comparison of time-of-flight, structured-light and active stereoscopy technologies*. Springer Nature.

[19] Intel® RealSense™ Depth and Tracking Cameras, 2019. *Which Intel RealSense device is right for you? (Updated June 2020)*. [online] Available at: https://www.intelrealsense.com/which-device-is-right-for-you/ [Accessed 15 May 2022].

[20] Intel ® RealSense TM Depth Camera SR300 Series Product Family Datasheet Intel ® RealSense™ Depth Camera SR305, Intel ® RealSense™ Depth Module SR300 Revision 002, 2019. [online] Available at: https://www.intelrealsense.com/wp-content/uploads/2019/07/RealSense_SR30x_Product_Datasheet_Rev_002.pdf [Accessed 15 May 2022].

[21] Intel ® RealSense TM Product Family D400 Series Datasheet Intel ® RealSense™ Vision Processor D4, Intel, n.d. [online] Available at: https://www.intelrealsense.com/wp-content/uploads/2022/04/Intel-RealSense-D400-Series-Datasheet-April-2022-v2.pdf?_ga=2.226817501.1950346234.1651486981-1661064436.1599063643 [Accessed 15 May 2022].

[22] Intel® RealSense™ Depth and Tracking Cameras, n.d. *Intel RealSense SDK 2.0 – Intel RealSense Depth and Tracking cameras*. [online] Available at: https://www.intelrealsense.com/sdk-2/.

[23] Carfagni, M., Furferi, R., Governi, L., Servi, M., Uccheddu, F. and Volpe, Y., 2017. On the performance of the Intel SR300 depth camera: metrological and critical characterization. *IEEE Sensors Journal*, *17*(14), pp.4508-4519.

[24] Carfagni, M., Furferi, R., Governi, L., Santarelli, C., Servi, M., Uccheddu, F. and Volpe, Y., 2019. Metrological and critical characterization of the Intel D415 stereo depth camera. *Sensors*, *19*(3), p.489.

[25] Intel® RealSense™ Developer Documentation, n.d. *Tuning depth cameras for best performance*. [online] Available at: https://dev.intelrealsense.com/docs/tuning-depth-cameras-for-best-performance.

[26] www.mathworks.com. n.d. *Delaunay triangulation in 2-D and 3-D - MATLAB*. [online] Available at: https://www.mathworks.com/help/matlab/ref/delaunaytriangulation.html [Accessed 15 May 2022].

 ISBN 978-86-7466-930-3