IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 July 2024, accepted 6 August 2024, date of publication 9 August 2024, date of current version 20 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3441238

RESEARCH ARTICLE

Comparative Evaluation of Intel RealSense D415, D435i, D455, and Microsoft Azure Kinect DK Sensors for 3D Vision Applications

MICHAELA SERVI, ANDREA PROFILI, ROCCO FURFERI, AND YARY VOLPE
Department of Industrial Engineering, University of Florence, 50139 Florence, Italy
Corresponding author: Michaela Servi (michaela.servi@unifi.it)

ABSTRACT RGB-D cameras have been widely used in various research fields and applications in recent years. With the growing availability of commercial products, choosing the most suitable sensor for a specific application has become more complex. The lack of tools to evaluate and compare the metrological performance of these sensors was addressed by the recent ISO 10360-13:2021 standard, which defines the standards for the metrological characterization of 3D optical coordinate measuring systems. This paper applies the ISO 10360-13 methodology to characterize and compare four RGB-D cameras: three Intel® RealSense™ (D415, D435i, D455) and one Microsoft® Azure Kinect DK. The procedures described in the standard were integrated with three tests: the analysis of systematic errors measured acquiring a planar surface at increasing distances; a test to analyze the resolution performances with a specifically devised artifact; the evaluation of 3D reconstruction performances on two objects with different geometric characteristics. The results highlighted the strengths and weaknesses of each device. The D415 showed better performance in reconstructions from close-range acquisitions. The Azure Kinect DK had the best results for systematic error and resolution. The D455 generally achieved the best performance in the standard tests, followed by the D435i.

INDEX TERMS Intel RealSense, Microsoft Azure Kinect DK, ISO 10360-13:2021 standard, device characterization, performance evaluation.

I. INTRODUCTION

RGB-D cameras represent a significant technological innovation in the fields of computer vision and 3D reconstruction. These advanced optical sensors enable the capture of visual data in three dimensions by intelligently integrating color RGB information with depth data for every pixel in the image. In this way, RGB-D cameras are capable of creating a detailed depth map of the observed scene, paving the way for numerous revolutionary applications.

The key aspect that makes these cameras so advantageous is their cost-effective nature as 3D optical coordinate measurement systems, and despite their compact size and low cost, they offer large acquisition volumes, portability, and extreme ease of use. These characteristics have made them the preferred solution across various research and application sectors [1], [2], [3].

RGB-D cameras find use in a wide range of domains: in robotics, they allow systems to interact with their surroundings for indoor and outdoor environment mapping [1], [4], [5], [6], [7]; in autonomous driving, they provide essential 3D information [8]; in the biomedical field, they enable digitizing patient anatomy for human modeling, recognition, and tracking [9], [10], [11], [12]; in manufacturing sectors, they facilitate automation and optimization of production processes, such as in agriculture [13], [14], [15]. In general, wherever there is a need to replicate human vision capabilities in terms of field of view and working environmental conditions, RGB-D cameras offer an effective solution.

There are three main technologies behind these devices: stereoscopic systems that interpret the scene by comparing

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

two images from slightly different viewpoints, with the disparity allowing 3D mapping; time-of-flight (ToF) systems that measure the time a light signal takes to travel to the scene and back; and structured light systems that project a known pattern onto the scene and analyze its deformation to obtain depth. In addition to the operating principle, the surrounding environmental conditions and the type of observed object can also influence the final metrological performance of RGB-D cameras.

Knowledge of these performance characteristics has become a fundamental requirement for choosing which camera to use, especially when working at close ranges, since accuracy is generally proportional to the distance between the sensor and the object. Therefore, users typically need to work as close as possible to maximize the sensor's capabilities.

Numerous studies have presented metrological characterizations of depth cameras, comparing sensors from different manufacturers and evaluating their accuracy and precision [3], [16], [17], [18], [19], [20], [21], [22]. Moreover standards have been developed that define the tests to be performed for a comprehensive evaluation of such sensors. Recently, the ISO 10360-13:2021 standard was introduced and is currently the most advanced standard in the field of RGB-D camera evaluation.

The ISO 10360-13:2021 standard defines artifacts and procedures for assessing and verifying the performance of optical measurement systems in terms of probing error, distortion error, and flatness error.

In this work, this standard was used to compare the performance of four of the most compelling low-cost optical devices currently available on the market, i.e. the Intel® RealSense™ D415, D455, D435i and Microsoft® Azure Kinect DK. In addition to the tests provided by the standard, in order to provide an in-depth characterization that can support the choice of the most suitable device for each specific application, the work integrates three additional tests: the analysis of systematic depth errors, a dedicated test to analyze the resolution of this category of devices, and a test to evaluate the achievable performance in 3D reconstruction of objects with different geometric characteristics.

This work is therefore among the first to rigorously implement the recent ISO 10360-13:2021 standard for the characterization of low-cost optical 3D measurement systems and proposes a test methodology that integrates with additional tests using specifically designed artefacts. This combination offers a more comprehensive and multi-dimensional characterization of device performance, going beyond the limitations of standardized tests, providing a robust and replicable framework for the characterization of current and future sensors, regardless of manufacturer. A detailed comparative analysis of the four RGB-D devices is then provided to allow a more complete understanding of the relative capabilities of the devices in various usage scenarios.

The structure of this work is organized as follows: Section II provides a description of the analyzed optical sensors. Section III presents a brief overview of the ISO 10360-13:2021 standard and the additional tests conducted for a comprehensive evaluation. Section IV showcases the results obtained from the various tests performed. Section V discusses and analyzes the implications and significance of the obtained results.

II. EXAMINED SENSORS

Four 3D optical sensors were analyzed in this study, which are D415, D435i and D455 produced by Intel® RealSense™ [23] and Azure Kinect DK produced by Microsoft [24] (depicted in Fig. 1).

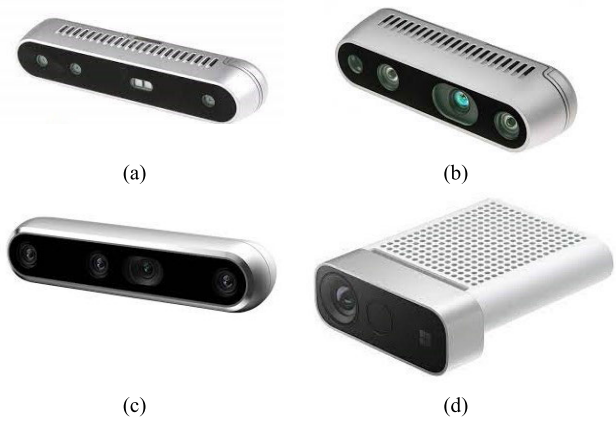


FIGURE 1. The four depth sensors analyzed in this study: (a) Intel® RealSense™ depth camera D415, (b) Intel® RealSense™ depth camera D435i, (c) Intel® RealSense™ depth camera D455, (d) Microsoft® Azure Kinect DK.

Numerous features are shared by the first three cameras in the Intel RealSense D400 series. They have similar design, the same frame rates (up to 90 fps), and they use a stereoscopic vision technology. The three Intel cameras add active lighting to improve depth reconstruction, which is performed by comparing the differences between two RGB images taken by two different sensors to enhance the texture of the scene under observation. The shutter type (global shutter for the D435i and D455, rolling shutter for the D415) and fields of view (FOV) of these three sensors differ from one another. While the D435i and D455 have bigger FOVs and optimum maximum ranges of 6 and 3 meters, respectively, the D415 has a lower FOV with a maximum range of only 3 meters.

On the other hand, continuous wave time-of-flight (CW ToF) technology is used for 3D reconstruction in the Microsoft Azure Kinect DK. With this technology, depth is determined by measuring the phase difference between the signal that is sent and received.

Depending on whether the camera uses pixel binning or not, the Azure Kinect DK sensor has many operating modes that give distinct FOVs (wide and narrow) and operating ranges. In comparison to the Intel RealSense cameras that were previously discussed, it possesses a lower frame rate (up to 30 fps).

Table 1 provide additional details regarding the four sensors, including their specifications. To ensure consistency and optimal performance during the experiments, the D415, D435i, and D455 cameras from Intel RealSense were configured using the default settings provided by the acquisition software. Additionally, these cameras were operated at their highest possible resolution of 1280 × 720 pixels, maximizing the available image quality and detail.

The Microsoft Azure Kinect DK camera was set to operate in its NFOV (Narrow Field of View) mode without binning. This configuration choice was made to take advantage of the camera's specialized operating modes, which can provide higher resolution and longer-range capabilities compared to other modes.

TABLE 1. Key characteristics of the D455, D415, D435i and Azure Kinect DK sensors.

General Characteristics

Device	*Technology*	*IMU*	*Ideal range*	*Dimensions (L×D×H)*
	-	-	*[mm]*	*[mm]*
D415	Stereoscopic	No	0.5 – 3	99×20×23
D435i	Stereoscopic	Yes	0.3 – 3	90×25×25
D455	Stereoscopic	Yes	0.6 – 6	124×26×29
Azure Kinect	Time of Flight	Yes	Up to 0.5 – 5.46[(1)] Up to 0.25 – 2.88[(2)]	103×39×126

[(1)] NFOV no binning 0.5 – 3.86 m; NFOV 2×2 with binning 0.5 – 5.46 m.
[(2)] WFOV no binning 0.25 – 2.88 m; NFOV 2×2 with binning 0.25 – 2.21 m.

Depth module

Device	*FOV (H×V)*	*Resolution*	*Frame rate*	*Shutter*
	[°]	*Pixel*	*fps*	-
D415	65×40	Up to 1280×720	Up to 90	Rolling
D435i	87×58	Up to 1280×720	Up to 90	Global
D455	87×58	Up to 1280×720	Up to 90	Global
Azure Kinect	75×65[(3)] 120×120[(4)]	Up to 640×576[(3)] Up to 1024×1024[(4)]	Up to 30	-

[(3)] NFOV no binning 640×576; NFOV 2×2 with binning 320×288.
[(4)] WFOV no binning 1024×1024; WFOV 2×2 with binning 512×512.

RGB module

Device	*FOV (H×V)*	*Resolution*	*Frame rate*	*Shutter*
	[°]	*Pixel*	*fps*	-
D415	69×42	1920×1080	Up to 30	Rolling
D435i	69×42	1920×1080	Up to 30	Rolling
D455	90×65	1280×800	Up to 30	Global
Azure Kinect	90×59[(5)] 90×74.3[(6)]	Up to 3840×2160[(5)] Up to 4096×3072[(6)]	Up to 30[(6)]	-

[(5)] Available resolutions: 1280×720, 1920×1080, 2560×1440, 3840×2160.
[(6)] Available resolutions: 2048×1536, 4096×3072 (up to 15 fps).

III. EXPERIMENTAL RESULTS

The metrological characterization procedures for the close-range performance of the four devices are presented below. The proposed methodology foresees four types of tests: application of the ISO 10360-13:2021 standard; evaluation of systematic depth errors, including depth offset errors and systematic nonplanarity errors; assessment of accuracy and resolution using a specifically designed artifact with known geometries; evaluation of 3D reconstruction capabilities by capturing objects with varying complexities.

To obtain the artifacts' reference dimensions and models the Romer Absolute Arm 7520 SI/SE [25] (accuracy of ±0.063 millimeters) was employed.

A. ISO 10360-13:2021 STANDARD

The ISO 10360-13:2021 standard describes procedures for confirming the stated performance of an optical coordinate measuring system (CMS). Three tests are required: probing error (size and shape error), distortion error, and flatness error. The tests aim to find the maximum error the optical sensor can make in the intended measurement volume, consisting of eight voxels, with the maximum enclosed length or distance between two points denoted by L_0. The documentation includes measurement details, materials used for artifacts, and required environmental factors.

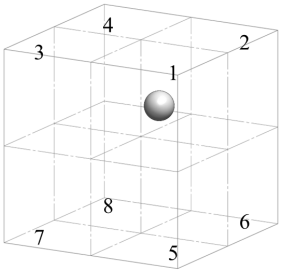
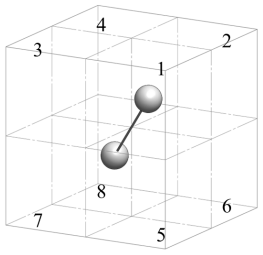


FIGURE 2. Example of a spherical artifact acquired in a voxel of the measurement volume.



Position	Shpere 1 in voxel	Shpere 2 in voxel
1	1	7
2	4	6
3	5	3
4	8	2
5	2	4
6	3	1
7	3	4
8	6	8
9	7	5
10	6	5
11	4	5
12	3	6

FIGURE 3. Ball-bar positions defined by the standard for measuring distortion error.

- Probing Error: A calibrated spherical artifact (Fig. 2) with diameter Φ between 0.02^*L_0 and 0.2^*L_0 is placed and acquired in each voxel, repeated three times. Size error (PS) and form error (PF) are calculated.
- Distortion Error: An artifact with two spheres connected center-to-center, with distance $LP > 0.3^*L_0$ and diameter Φ between 0.02^*L_0 and 0.2^*L_0, is acquired at 12 positions within the measurement volume (Fig. 3), repeated three times. The distortion error D is the difference between the measured and actual center-to-center distance.

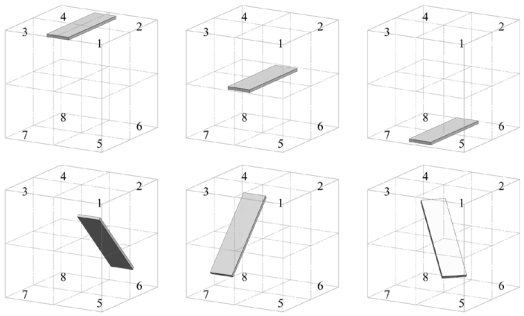


FIGURE 4. Flat surface positions defined by the standard for measuring flatness error.

- Flatness Error: A calibrated flat surface with long side $L_{max} > 0.5^{*}L_0$ and short side $L_{min} < 0.1^{*}L_0$ is acquired three times at six different positions (Fig. 4). The flatness error F is the minimum distance between two parallel planes encompassing the scanned artifact.

As said, the experiments were conducted considering the FOV ranging from 500 to 1500 millimeters, and differently from the standard, which likens the measurement volume to a parallelepiped, the measurement volumes of the devices under study more closely resemble truncated pyramids (Fig. 5). This difference in shape results in a non-negligible discrepancy between the area framed by the device at the closest distance to the camera and the area framed at the most distant distance. This complication introduces challenges in shaping the artifacts to be acquired, as they need to be fully enclosed within the four voxels closest to the sensor. Despite the different shape of the measurement volume, it was determined that within the 500-1500 mm acquisition range, it is possible to size the artifacts according to the standard and ensure their full acquisition at the required positions for all devices under consideration. To sizes for the artifacts the diagonals of the four sensors' measurement volumes were calculated.

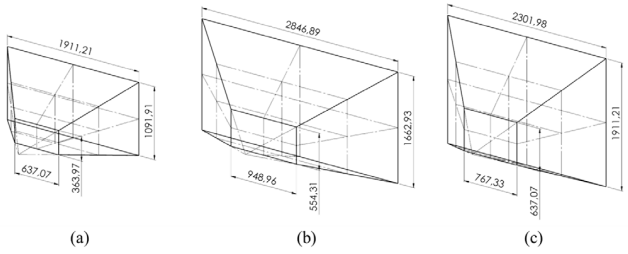


FIGURE 5. FOV of the four sensors in the range 500-1500 mm. From left to right: (a) D415, (b) D435i and D455, (c) Azure Kinect DK.

The measurements of the artifacts were decided as follows: the single sphere has a diameter (Ô) of 143.24 mm, the distance between the two centers of the two ball-bar spheres (LP) is 903.33 mm, and the plane has the dimensions of 1250 × 250 x 90 mm. These artifacts used for these experiments are depicted in Fig. 6.

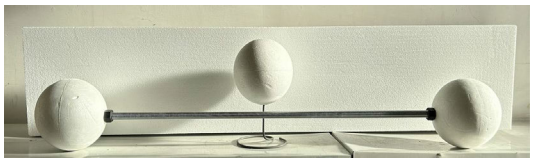


FIGURE 6. Artifacts selected for characterization according to ISO standard 10360-13:2021.

As is evident from the image, the surface of the artefacts has some localized irregularities. To quantify these imperfections, a measurement of the artefacts' surface was performed using the professional Romer Absolute Arm scanner. For the spheres, the best-fit sphere was extracted and the deviation between it and the acquired data was measured. A similar procedure was applied to the plane. The average error obtained on the single sphere is 0.004 mm with standard deviation 0.28 mm and maximum error 1.66 mm. The mean error obtained on the ball-bar is 0.007 mm with standard deviation 0.27 mm and maximum error 1.25 mm on the first ball, and 0.001 mm with standard deviation 0.19 mm and maximum error 1.76 mm on the second ball. The mean error obtained on the plane is 0.0007 mm with standard deviation 0.12 mm and maximum error 0.7 mm. It is important to mention that these surface imperfection values are orders of magnitude lower than the resolution of the RGB-D cameras under investigation in this study. Consequently, we can confidently state that these minimal artefact imperfections cannot significantly influence the measurements made with the sensors under investigation.

B. SYSTEMATIC DEPTH ERROR

To evaluate the errors arising from acquisitions at varying distances from the sensor, known as ''inhomogeneous distances'' errors, the following test was performed: the sensor was positioned orthogonally to a flat surface, and acquisitions were made at progressive distances ranging from 500 to 1500 millimeters, with 100-millimeter increments. Similarly to [19], a linear guide was set up perpendicular to the plane being observed, ensuring orthogonality with a reference cross marker placed on the flat surface. These markers act as alignment references, with the cross drawn in the center of the image as the sensor moves away from the surface (Fig. 7). Two types of errors were analyzed: systematic depth error and depth offset error. The systematic depth error was estimated by comparing the scan data to a plane of known flatness. The depth offset error was obtained by measuring the distance along the Z-axis between the scan and the reference planes.

C. ACCURACY AND RESOLUTION ANALYSIS

A test was performed to evaluate the sensors' resolution capabilities similar to the test described in [22]. For this experiment, an artifact made of four cuboids of varying thicknesses, all perpendicular to a plane, was appositively devised. The steps between the cuboids (illustrated in Fig. 8) were measured with the professional laser scanner.

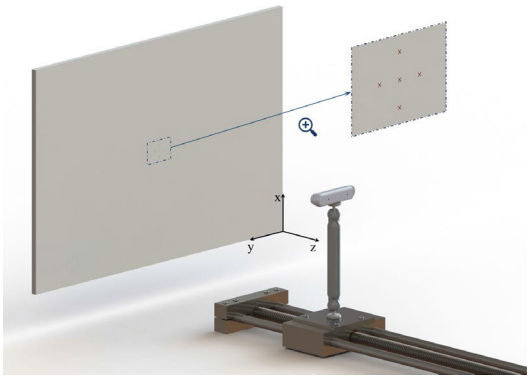


FIGURE 7. Acquisition setup for the systematic depth error.

The objective of this test was twofold: to evaluate the sensors' ability to accurately delineate the distinct planes and their relative distances, as well as to observe the impact of height differences on the sensors' resolution performance.

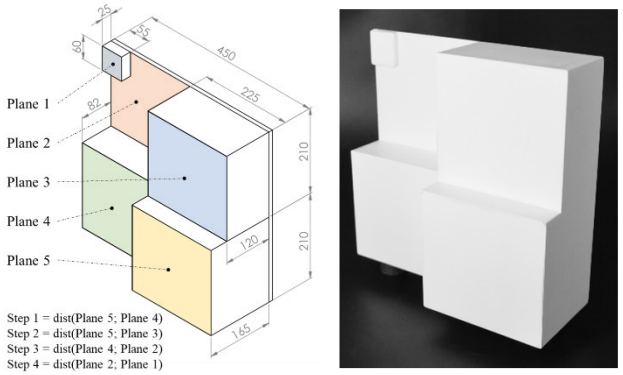


FIGURE 8. Artifact made for analysis on accuracy and resolution.

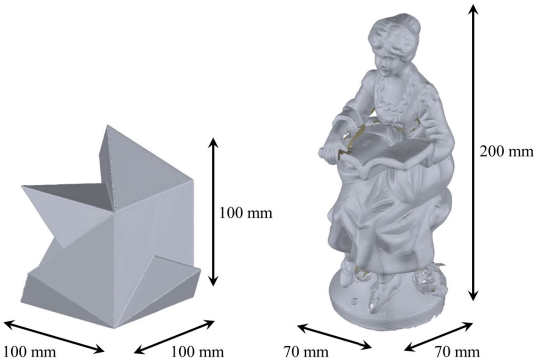


FIGURE 9. 3D Tangram and statue used for 3D object reconstruction.

D. OBJECT RECONSTRUCTION

The final test focused on evaluating the sensors' 3D reconstruction capabilities by assessing their performance in accurately capturing object geometries.

Following similar methodologies from three previous works [19], [20], [21], two different objects were selected: a free-form object (approximately 200 mm high and 70 mm wide) and an object with known 3D geometries (Fig. 9).

The acquisition process involved rotating the objects while the sensor remained fixed in a stable position. This approach enabled capturing the objects from multiple angles, facilitating comprehensive 3D reconstruction. To optimize performance, the distance between the sensor and the object was set to the minimum of 500 mm as the sensors' accuracy and precision tend to be higher at closer distances.

IV. RESULTS

The acquired point clouds were processed independently for each acquisition leveraging the Geomagic Design X® [26] software and best-fit operations were used to generate planar and spherical primitives from the acquired data as well as deviation analysis operation, to compare the extracted information with the nominal values. The results obtained for the four tests are shown in the following.

A. ISO 10360-13:2021 STANDARD

1) PROBING ERROR

Fig. 10 shows the spherical artifact acquisitions obtained by the four sensors at the 8 positions indicated by the standard within the measurement volume.

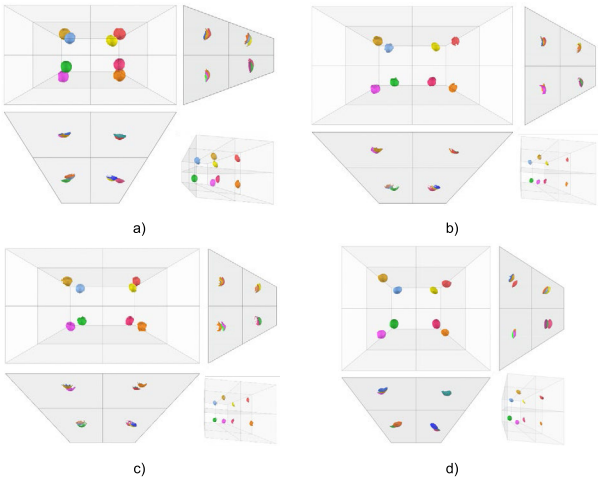


FIGURE 10. Sphere acquisitions performed by the four sensors according to the ISO 10360-13:2021 standard: a) D415 b) D435i c) D455 d) Azure Kinect DK.

Table 2 shows value of the probing size error (PS), calculated for the various positions and for all cameras as the difference between the actual value of the sphere diameter and the measured value. The values in the table are obtained as the average over the 3 acquisitions for each position indicated by the standard.

Table 3 shows the probing form (PF) error for each sensor, calculated as the average value of the difference between the maximum and minimum distance of the points on the surface of the i-th sphere from the relative best-fit sphere. The values

TABLE 2. Mean error and standard deviation of size error PS.

			D415	D435i	D455	Azure Kinect DK
Average PS	Mean	[mm]	0.928	3.167	3.528	14.947
	Std	[mm]	1.401	3.774	4.329	18.627

in the table are obtained as the average over the 3 acquisitions for each position indicated by the standard.

TABLE 3. Mean error and standard deviation of shape error PF.

			D415	D435i	D455	Azure Kinect DK
Average PF	Mean	[mm]	8.723	15.369	19.354	9.757
	Std	[mm]	1.232	2.279	2.592	1.454

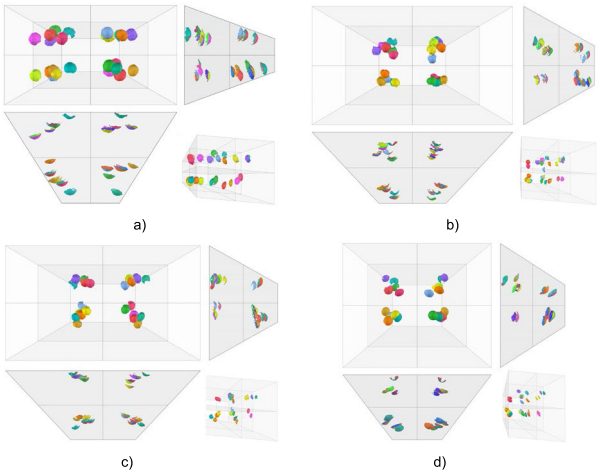


FIGURE 11. Ball-bar acquisitions performed by the four sensors according to ISO standard 10360-13:2021: a) D415 b) D435i c) D455 d) Azure Kinect DK.

2) DISTORTION ERROR

Fig. 11 shows the ball-bar acquisitions obtained by the four sensors at the 12 positions indicated by the standard within the measurement volume.

Table 4 shows the average value of the distortion error calculated as the difference between the actual and measured distance between the centers of the two spheres of the ball-bar. The values in the table are obtained as the average over the 3 acquisitions for each position indicated by the standard.

TABLE 4. Mean error and standard deviation of the distortion error D.

			D415	D435i	D455	Azure Kinect DK
Average D	Mean	[mm]	15.994	2.925	3.538	34.040
	Std	[mm]	17.015	1.981	3.967	38.253

3) FLATNESS ERROR

Fig. 12 shows the flat surface acquisitions obtained by the four sensors at the 6 positions indicated by the standard within the measurement volume.

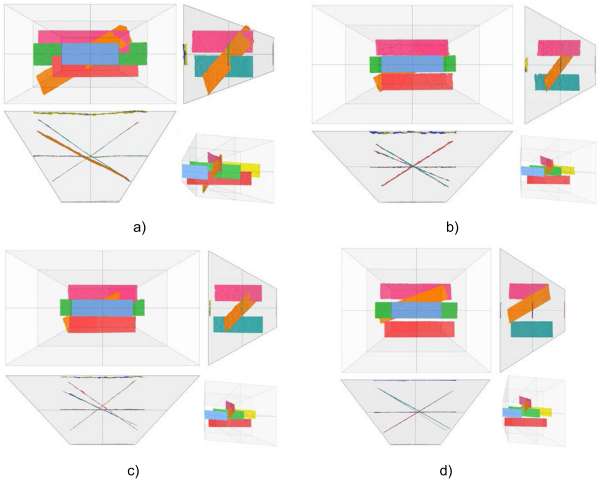


FIGURE 12. Flat surface acquisitions performed by the four sensors according to ISO standard 10360-13:2021: a) D415 b) D435i c) D455 d) Azure Kinect DK.

Table 5 shows the average value of the flatness error calculated as the difference between the maximum and minimum distance of the surface points from the relative best-fit plane. The values in the table are obtained as the average over the 3 acquisitions for each position indicated by the standard.

TABLE 5. Mean error and standard deviation of the flatness error F.

			D415	D435i	D455	Azure Kinect DK
Average F	Mean	[mm]	24.666	29.136	20.896	15.031
	Std	[mm]	4.709	3.199	2.016	1.479

B. SYSTEMATIC DEPTH ERROR

This section shows the results obtained on systematic depth errors, which are (1) depth offset errors and (2) systematic nonplanarity errors.

The first type of error was analyzed both qualitatively and quantitatively by considering the differences in Z distances between the scanned planes and the reference planes (where Z is the optical axis of the camera). Fig. 13 qualitatively shows the result of this test.

A quantitative analysis was carried out by calculating for each step the distance between the known reference and the respective best-fit plane, imposing parallelism between the planes as a constraint. Fig. 14 shows the trend of offset error as the distance between plane and sensor increases.

The second type of error was analyzed by referring to the best-fit planes and calculating their deviation from the acquired images (Fig. 15).

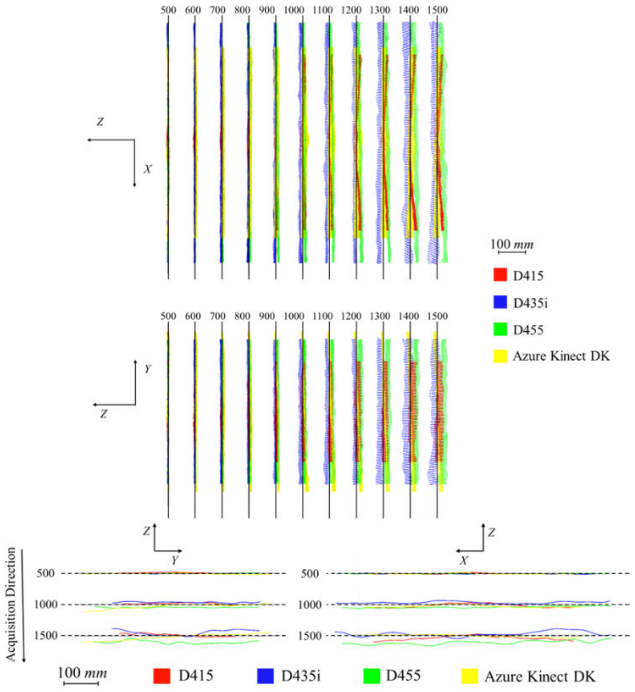


FIGURE 13. Qualitative offset error in the range 500-1500 mm.

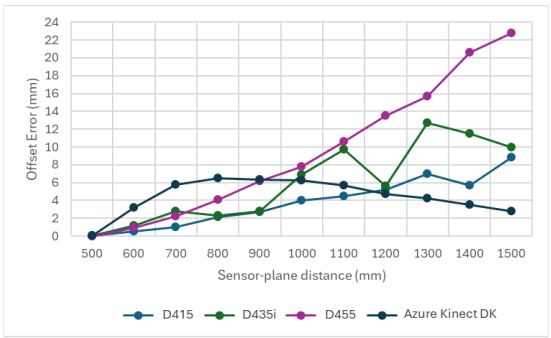


FIGURE 14. Systematic depth error in the range of 500-1500 mm: offset error.

Table 6 shows the errors, averaged over all acquisitions, in terms of mean distance, standard deviation and range defined as the difference between the maximum and minimum distance of the surface points from the best-fit plane.

TABLE 6. Systematic depth error in the range 500-1500 mm.

			D415	*D435i*	*D455*	*Azure Kinect DK*
Systematic depth error	*Mean*	*[mm]*	0.509	-0.075	0.144	0.307
	Std	*[mm]*	3.900	3.026	2.202	2.050
	Range	*[mm]*	21.866	22.842	18.878	16.752

Fig. 16 shows the trend, for each position, of the error range of the acquired points with respect to the best-fit surface.

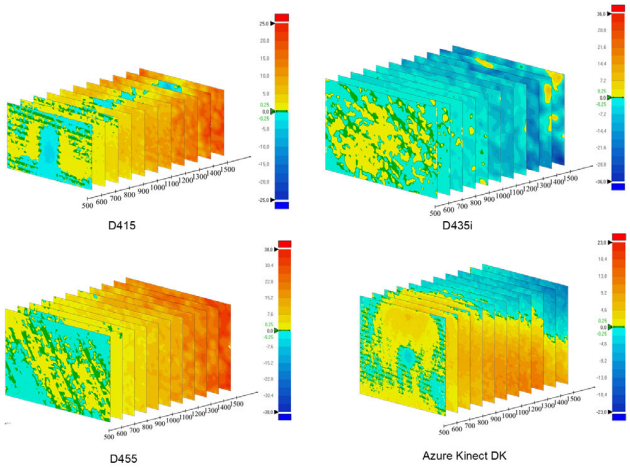


FIGURE 15. Deviation of the acquired planes from the relative best-fit planes.

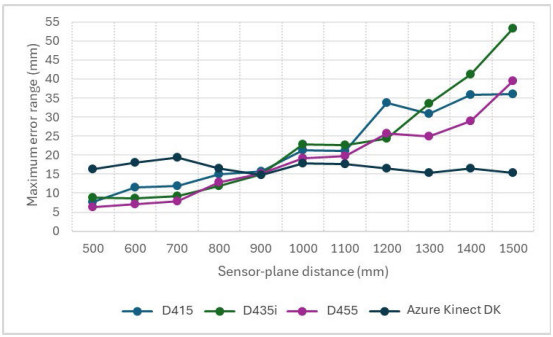


FIGURE 16. Systematic depth error in the range of 500-1500 mm: maximum error range.

C. ACCURACY AND RESOLUTION ANALYSIS

To evaluate the accuracy of the RGB-D cameras in measuring distances between planes, a multi-step analysis was conducted.

After an artifact point isolation process, where the point clouds representing the surfaces of interest were extracted from the overall acquired data, best-fit planes were constructed using the point clouds for each planar surface of the artefact in Fig. 8. To minimize the influence of edge effects and point scattering, the best-fit planes were generated considering areas as far from the object edges as possible. This approach aimed to ensure that the generated planes accurately represented the central regions of the acquired surfaces, where the data quality was expected to be highest.

With the established best-fit planes, the distances between these planes were compared to the corresponding reference distances obtained from the highly accurate laser scanning data.

Fig. 17 presents a visual representation of the analysis results to provides a clear overview of the cameras' performance in measuring inter-planar distances (indicated as ''Step'' in Fig. 17 referring to the steps shown in Fig. 8),

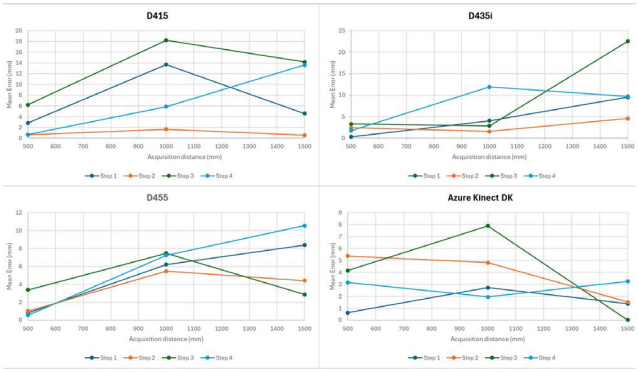


FIGURE 17. Error in evaluating distances between planes.

enabling easy identification of potential biases or deviations from the expected values.

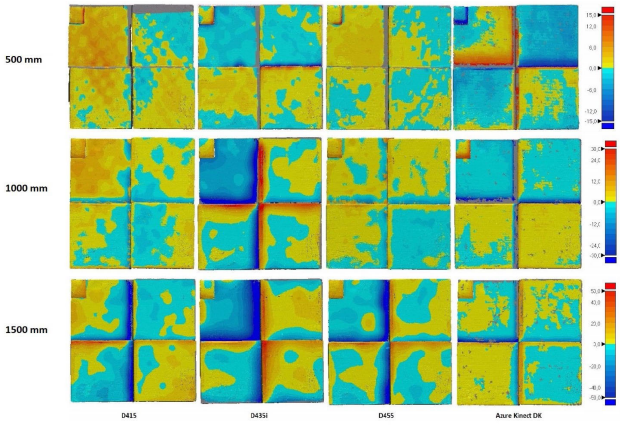


FIGURE 18. Investigation of camera resolution at the various distances considered.

For the resolution assessment the acquired point clouds were manually aligned with the reference model, and a depth map was generated by comparing the two datasets. Fig. 18 illustrates this depth map, where a color scale is used to represent the deviations between the acquired data and the reference model in millimeters. Positive values, indicating regions where the acquired data protrudes beyond the reference model, are marked by shades ranging towards red. Negative values, representing areas where the acquired data falls short of the reference model, are marked by shades ranging towards blue. Areas that were not recorded in the point clouds obtained from the cameras are depicted in gray.

D. OBJECT RECONSTRUCTION

As for the 3D reconstruction test, in order to recreate the three-dimensional model of each observed object, after removing the edges from each acquisition, it was necessary to perform an alignment process consisting of two steps, an initial coarse alignment based on manual selection of the corresponding points and a fine alignment using the Iterative Closest Point (ICP) algorithm.

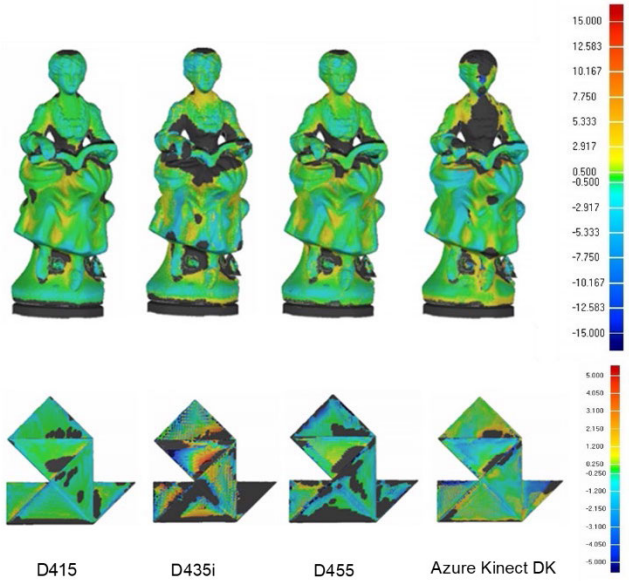


FIGURE 19. Deviation of the two 3D reconstructions (free-form statue object and 3D tangram).

Fig. 19 shows the deviation between the acquired artifacts and the reference models. The color scale depicts this difference between the two in mm: positive values are marked with colors from green to red, negative values are marked with colors from green to blue, while gray areas were not considered in the deviation analysis (in these areas the data may be missing or too far from reality).

Tables 7 and 8 show the reconstruction errors in terms of mean distance, standard deviation and range of error between the reconstructed data and the base model.

TABLE 7. Mean distance, standard deviation and error range in statue reconstruction.

			D415	*D435i*	*D455*	*Azure Kinect DK*
Statue reconstruction	*Mean*	*[mm]*	-0.236	-0.101	0.002	-0.660
	Std	*[mm]*	1.098	1.981	1.437	6.532
	Range	*[mm]*	13.995	21.334	32.271	39.995

TABLE 8. Mean distance, standard deviation and error range in 3D tangram reconstruction.

			D415	*D435i*	*D455*	*Azure Kinect DK*
3D Tangram reconstruction	*Mean*	*[mm]*	-0.499	-0.522	-0.825	-0.694
	Std	*[mm]*	1.031	3.186	1.499	2.080
	Range	*[mm]*	15.219	19.898	16.909	19.992

V. DISCUSSION

Building upon the results previously reported, this section aims to discuss and compare the performance of the four devices under consideration, with the goal of outlining their merits and limitations when employed for metrology

applications. It is important to note that this analysis is approached from the perspective of utilizing these low-cost sensors in their ''out-of-the-box'' configuration, without any modifications or deviations from their factory-provided features. This includes refraining from recalibration procedures, the use of external programs, filters, or libraries, and relying solely on the default acquisition and filtering settings.

Analyzing the results obtained from the tests indicated by the standard, several key observations can be made regarding the performance of the four devices. The Intel RealSense D415 exhibited the lowest probing size error (PS) and probing shape error (PF) among the tested cameras. Specifically, it achieved a mean PS of 0.928 mm and a mean PF of 8.723 mm, demonstrating its capability in accurately measuring dimensions and capturing geometric forms. Regarding the distortion error (D), the Intel RealSense D435i emerged as the top performer, with a mean error of 2.925 mm and a standard deviation of 1.981 mm. This result highlights the D435i's ability to minimize distortions and maintain geometric integrity throughout its field of view. For the flatness error (FD), the Microsoft Azure Kinect DK delivered the best result, with a mean value of 15.031 mm and a standard deviation of 1.479 mm. This performance suggests that the Azure Kinect DK excels in capturing planar surfaces with minimal deviations from flatness.

Interestingly, during the analysis of the images acquired by the Azure Kinect DK, it was observed that these images exhibited lower background noise and higher resolution of the observed artifacts compared to the other cameras. This improved image quality is likely attributable to an internal smoothing process implemented in the camera's software, which enhances the definition of the acquired data at the expense of greater error in measuring effective sizes.

As an example of this trade-off, in the test with the single spherical artifact, the Azure Kinect DK recorded a PS error with a mean value of 14.947 mm and a standard deviation of 18.627 mm, along with a PF error with a mean value of 9.757 mm and a standard deviation of 1.454 mm.

Fig. 13 illustrates the acquisitions of the plane surface captured by the cameras at progressively greater distances. A qualitative analysis of these images reveals a gradual increase in the distance between the plane approximating the acquired data and the relative known reference surface across all four sensors. However, when considering both the offset and flatness analyses, the Microsoft Azure Kinect DK exhibited superior overall performance compared to the other cameras. In terms of offset, the Azure Kinect DK presented a maximum error of less than 7 mm, indicating its ability to maintain accurate depth measurements across the varying distances.

Regarding the flatness analysis, the Azure Kinect DK demonstrated remarkable results, with an average maximum range of 16.752 mm and an average standard deviation of 2.050 mm. Notably, Fig. 14 shows that the trend of the maximum point error range with respect to the best-fit plane remained stable for all distances for the Azure Kinect DK, in contrast to the other three cameras. This exceptional flatness performance can be attributed to the greater smoothing that the Azure Kinect DK applies to the acquired data, resulting in flatter and more uniform surfaces in the captured images. While this smoothing process may introduce some trade-offs in terms of capturing fine details or sharp edges, it enhances the overall flatness and consistency of the acquired data, which can be advantageous for specific applications that prioritize planar surface accuracy.

Analyzing Fig. 18, which presents the resolution assessment, it is evident that at close range, all the cameras exhibit similar performance. However, as the distance increases, a progressive deterioration in the resolution error of the acquired images becomes noticeable for the Intel® produced sensors. Consequently, these cameras struggle to sharply delineate the step features, unlike the Azure Kinect DK, which demonstrates uniform results across all three acquisition distances.

Regarding the 3D reconstruction test, the results reveal varying performances among the cameras. The Intel RealSense D415 and D455 provided very similar and accurate reconstructions of the statue object. In contrast, the D435i yielded a slightly worse reconstruction, characterized by the presence of numerous gray areas, indicating missing or unreliable data. The Microsoft Azure Kinect DK, on the other hand, exhibited a comparatively poorer performance in reconstructing the statue due to the strong flattening effect it applies to the acquired images. This smoothing process, while beneficial for maintaining flatness, can potentially compromise the capture of intricate details and surface variations. For the 3D reconstruction of the tangram object, the D415 once again recorded the best data, closely followed by the Azure Kinect DK. The Kinect's ability to accurately recognize and capture flat surfaces proved advantageous for remodeling the tangram object, surpassing its performance on the statue reconstruction. The D455 still provided reasonably good data for the tangram reconstruction, despite the presence of several gray areas. Finally, the D435i recorded the worst data for this object, potentially due to the tangram's relatively small size, which may have challenged the camera's capabilities.

When looking at the analysis results, it is important to mention that the different technologies necessarily lead to different measurement outcomes. For example, it is worth noting that the systematic depth error values measured with the Azure Kinect DK, despite being higher in the first part of the range, are much more stable as the distance increases - and indeed, it is the only camera for which the error decreases over time. This feature of the error trend is in fact imputable to the ToF technology, which tends to be more effective for longer distances for several reasons including temporal resolution (for short distances, the differences in flight times are very small and require an extremely high temporal resolution, at longer distances, the temporal differences are more significant and therefore easier to detect accurately). In general, at close distances active stereo cameras offer

greater resolution and surface detail, although they require more processing time.

VI. CONCLUSION

In conclusion, this work involved a comprehensive metrological characterization and performance comparison of four RGB-D devices – three from the Intel® RealSense™ family (D415, D435i, and D455) and the Azure Kinect DK produced by Microsoft®. The study aimed to evaluate and compare the capabilities of these devices across various aspects through a wide range of qualitative and quantitative tests.

The testing methodology incorporated elements from previous similar works, as well as adherence to the new ISO 10360-13:2021 standard, which specifically addresses the metrological characterization of optical devices. Despite the different technologies employed by these devices to obtain 3D coordinates of the observed scene, the tests were conducted consistently across all three devices, ensuring a fair and robust evaluation. The extensive tests performed revealed the unique strengths and weaknesses of each device, shedding light on their suitability for different applications and operating conditions. In terms of 3D reconstruction at close distances, the Intel RealSense D415 demonstrated superior quality, excelling in capturing intricate details and accurately representing geometric features from proximal acquisitions.

Regarding the tests on systematic depth errors and resolution quality, the Microsoft Azure Kinect DK emerged as a standout performer, exhibiting a remarkable ability to accurately represent planar surfaces with minimal deviations from flatness.

Finally, when it came to the tests related to the ISO 10360-13:2021 standard, which assess metrological performance in terms of measurement error, distortion error, and flatness error, the Intel RealSense D455 exhibited the highest overall performance, closely followed by the D435i.

ACKNOWLEDGMENT

The authors would like to express their utmost gratitude to Lorenzo Iacopini for his invaluable help with the whole process.

REFERENCES

[1] V. Tadic, A. Toth, Z. Vizvari, M. Klincsik, Z. Sari, P. Sarcevic, J. Sarosi, and I. Biro, "Perspectives of RealSense and ZED depth sensors for robotic vision applications," *Machines*, vol. 10, no. 3, p. 183, Mar. 2022, doi: 10.3390/machines10030183.

[2] J. D. Mejia-Trujillo, Y. J. Castano-Pino, A. Navarro, J. D. Arango-Paredes, D. Rincon, J. Valderrama, B. Munoz, and J. L. Orozco, "Kinect and Intel RealSense D435 comparison: A preliminary study for motion analysis," in *Proc. IEEE Int. Conf. E-Health Netw., Appl. Services (HealthCom)*, Oct. 2019, pp. 1–4, doi: 10.1109/healthcom46333.2019.9009433.

[3] F. Lourenço and H. Araujo, "Intel RealSense SR305, D415 and l515: Experimental evaluation and comparison of depth estimation," in *Proc. 16th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2021, pp. 362–369, doi: 10.5220/0010254203620369.

[4] A. K. Ramasubramanian, M. Kazasidis, B. Fay, and N. Papakostas, "On the evaluation of diverse vision systems towards detecting human pose in collaborative robot applications," *Sensors*, vol. 24, no. 2, p. 578, Jan. 2024, doi: 10.3390/s24020578.

[5] A. M. A. Zaki, M. Carnevale, C. Schlette, and H. Giberti, "On the use of low-cost 3D stereo depth camera to drive robot trajectories in contact-based applications," *Int. J. Adv. Manuf. Technol.*, vol. 128, nos. 9–10, pp. 3745–3759, Oct. 2023, doi: 10.1007/s00170-023-11851-x.

[6] S. Alexovič, M. Lacko, and J. Bačík, "3D mapping with a drone equipped with a depth camera in indoor environment," *Acta Electrotechnica et Inf.*, vol. 23, no. 1, pp. 18–24, Mar. 2023, doi: 10.2478/aei-2023-0003.

[7] R. Haenel, Q. Semler, E. Semin, P. Grussenmeyer, and S. Tabbone, "Evaluation of low-cost depth sensors for outdoor applications," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-2, pp. 101–108, Dec. 2022, doi: 10.5194/isprs-archives-xlviii-2-w1-2022-101-2022.

[8] Z. Qiu, J. Martínez-Sánchez, V. M. Brea, P. López, and P. Arias, "Low-cost mobile mapping system solution for traffic sign segmentation using Azure Kinect," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 112, Aug. 2022, Art. no. 102895, doi: 10.1016/j.jag.2022.102895.

[9] F. Buonamici, R. Furferi, L. Governi, S. Lazzeri, K. S. McGreevy, M. Servi, E. Talanti, F. Uccheddu, and Y. Volpe, "A practical methodology for computer-aided design of custom 3D printable casts for wrist fractures," *Vis. Comput.*, vol. 36, no. 2, pp. 375–390, Feb. 2020, doi: 10.1007/s00371-018-01624-z.

[10] L. Lv, J. Yang, F. Gu, J. Fan, Q. Zhu, and X. Liu, "Validity and reliability of a depth camera–based quantitative measurement for joint motion of the hand," *J. Hand Surg. Global Online*, vol. 5, no. 1, pp. 39–47, Jan. 2023, doi: 10.1016/j.jhsg.2022.08.011.

[11] G. Maculotti, L. Ulrich, E. C. Olivetti, G. Genta, F. Marcolin, E. Vezzetti, and M. Galetto, "A methodology for task-specific metrological characterization of low-cost 3D camera for face analysis," *Measurement*, vol. 200, Aug. 2022, Art. no. 111643, doi: 10.1016/j.measurement.2022.111643.

[12] U. Özsoy, Y. Yıldırım, S. Karaşin, R. Şekerci, and L. B. Sşzen, "Reliability and agreement of Azure Kinect and Kinect v2 depth sensors in the shoulder joint range of motion estimation," *J. Shoulder Elbow Surg.*, vol. 31, no. 10, pp. 2049–2056, Oct. 2022, doi: 10.1016/j.jse.2022.04.007.

[13] Y. Bai, B. Zhang, N. Xu, J. Zhou, J. Shi, and Z. Diao, "Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review," *Comput. Electron. Agricult.*, vol. 205, Feb. 2023, Art. no. 107584, doi: 10.1016/j.compag.2022.107584.

[14] S.-T. Liong, Y.-L. Wu, G.-B. Liong, and Y. S. Gan, "Moving towards agriculture 4.0: An AI-AOI carrot inspection system with accurate geometric properties," *J. Food Eng.*, vol. 357, Nov. 2023, Art. no. 111632, doi: 10.1016/j.jfoodeng.2023.111632.

[15] G. Dhanush, N. Khatri, S. Kumar, and P. K. Shukla, "A comprehensive review of machine vision systems and artificial intelligence algorithms for the detection and harvesting of agricultural produce," *Sci. Afr.*, vol. 21, Sep. 2023, Art. no. e01798, doi: 10.1016/j.sciaf.2023.e01798.

[16] G. Kurillo, E. Hemingway, M.-L. Cheng, and L. Cheng, "Evaluating the accuracy of the Azure Kinect and Kinect v2," *Sensors*, vol. 22, no. 7, p. 2469, Mar. 2022, doi: 10.3390/s22072469.

[17] M. Heinemann, J. Herzfeld, M. Sliwinski, J. Hinckeldeyn, and J. Kreutzfeldt, "A metrological and application-related comparison of six consumer grade stereo depth cameras for the use in robotics," in *Proc. IEEE Int. Symp. Robotic Sensors Environ. (ROSE)*, Nov. 2022, pp. 1–7, doi: 10.1109/ROSE56499.2022.9977421.

[18] S. Rijal, S. Pokhrel, M. Om, and V. P. Ojha, "Comparing depth estimation of Azure Kinect and realsense D435i cameras," *Available SSRN*, Oct. 2023, doi: 10.2139/SSRN.4597442.

[19] M. Servi, E. Mussi, A. Profili, R. Furferi, Y. Volpe, L. Governi, and F. Buonamici, "Metrological characterization and comparison of d415, d455, L515 RealSense devices in the close range," *Sensors*, vol. 21, no. 22, p. 7770, Nov. 2021, doi: 10.3390/s21227770.

[20] M. Carfagni, R. Furferi, L. Governi, C. Santarelli, M. Servi, F. Uccheddu, and Y. Volpe, "Metrological and critical characterization of the Intel D415 stereo depth camera," *Sensors*, vol. 19, no. 3, p. 489, Jan. 2019, doi: 10.3390/s19030489.

[21] M. Carfagni, R. Furferi, L. Governi, M. Servi, F. Uccheddu, and Y. Volpe, "On the performance of the Intel SR300 depth camera: Metrological and critical characterization," *IEEE Sensors J.*, vol. 17, no. 14, pp. 4508–4519, Jul. 2017, doi: 10.1109/JSEN.2017.2703829.

[22] A. Lopez Paredes, Q. Song, and M. H. Conde, "Performance evaluation of state-of-the-art high-resolution time-of-flight cameras," *IEEE Sensors J.*, vol. 23, no. 12, pp. 13711–13727, Jun. 2023, doi: 10.1109/JSEN.2023.3273165.

[23] *Intel RealSense D400 Series Product Family Datasheet*. Accessed: Apr. 15, 2024. [Online]. Available: https://dev.intelrealsense.com/docs/intel-realsense-d400-series-product-family-datasheet
[24] *Azure Kinect DK Documentation | Microsoft Learn*. Accessed: Apr. 15, 2024. [Online]. Available: https://learn.microsoft.com/en-gb/azure/Kinect-dk/?source=recommendations
[25] *ROMER Absolute Arm | Hexagon Manufacturing Intelligence*. Accessed: Oct. 26, 2019. [Online]. Available: https://www.hexagonmi.com/products/portable-measuring-arms/romer-absolute-arm
[26] *Geomagic Design X 3D Reverse Engineering Software*. Accessed: Oct. 29, 2021. [Online]. Available: https://www.3dsystems.com/software/geomagic-design-x

MICHAELA SERVI received the master's degree program in computer engineering from the School of Engineering, University of Florence, in 2015, and the Ph.D. degree in industrial engineering from the Department of Industrial Engineering in Florence, Italy, in 2020.

She is currently an RTDA Researcher with the Department of Industrial Engineering, University of Florence, within the SSD ING-IND/15. Her master's thesis was titled "Disparity coherent stereo video watermarking." She is developing a research project focused on the development of procedures for the realization of patient-specific medical devices. She has authored more than 50 publications in these scientific areas. Her research interests include reverse engineering, computer vision, software engineering, CAD, 3D printing, tools, and methods to support personalized medicine.

ANDREA PROFILI received the master's degree program in industrial engineering from the School of Engineering, University of Florence, in 2021. He passed the qualifying examination for the profession of engineer, in 2021. He is currently pursuing the Ph.D. degree with the Department of Industrial Engineering, University of Florence.

His research interests include reverse engineering, computer vision, additive manufactoring, CAD, 3D printing, tools, and methods to support personalized medicine.

ROCCO FURFERI received the Ph.D. degree. He is currently an Associate Professor with the Department of Industrial Engineering of Florence (DIEF), University of Florence, Italy, where he teaches the courses mechanical drafting and reverse engineering for optimization. He is a Referee for the mechanical engineering master's degree course. He is responsible for numerous national and international projects for DIEF, including the PNRR Cultural Heritage Active Innovation for Next-Gen Sustainable Society (CHANGES) and Tuscany Health Ecosystem - Spoke 3 Advanced technologies, methods, and materials for human health and well-being. He is the author of more than 150 articles in the fields of *Reverse Engineering*, *Computational Geometry*, CAD, *Textile*, *Mechanical Engineering*, and *Artificial Intelligence*.

YARY VOLPE received the M.Sc. degree in mechanical engineering and the Ph.D. degree in machine design and construction from the University of Florence, Italy, in 2006. After working as a Postdoctoral Researcher with the Department of Mechanics and Industrial Technologies, University of Florence (DIEF), in 2020, he assumed the faculty position as an Associate Professor in "Design and Methods of Industrial Engineering." He has coordinated the T3Ddy Laboratory, (joint laboratory between Meyer Children's Hospital and DIEF), since 2016. He is the author and co-author of more than 150 publications printed in international journals and participated in a number of international conferences. His main scientific research interests include computer aided design, image processing, virtual prototyping, FE simulation, reverse engineering, and rapid prototyping.

Open Access funding provided by 'Università degli Studi di Firenze' within the CRUI CARE Agreement