

Article

Experimental Procedure for the Metrological Characterization of Time-of-Flight Cameras for Human Body 3D Measurements

Simone Pasinetti ¹, Cristina Nuzzi ^{1,*}, Alessandro Luchetti ², Matteo Zanetti ², Matteo Lancini ^{3,*}
and Mariolino De Cecco ²

¹ Department of Industrial and Mechanical Engineering, University of Brescia, Via Branze 38, 25125 Brescia, Italy

² Department of Industrial Engineering, University of Trento, Via Sommarive, 9, 38123 Trento, Italy

³ Department of Medicine and Surgery, Radiology and Public Health, University of Brescia, Viale Europa 11, 25123 Brescia, Italy

* Correspondence: cristina.nuzzi@unibs.it (C.N.); matteo.lancini@unibs.it (M.L.)

Abstract: Time-of-flight cameras are widely adopted in a variety of indoor applications ranging from industrial object measurement to human activity recognition. However, the available products may differ in terms of the quality of the acquired point cloud, and the datasheet provided by the constructors may not be enough to guide researchers in the choice of the perfect device for their application. Hence, this work details the experimental procedure to assess time-of-flight cameras' error sources that should be considered when designing an application involving time-of-flight technology, such as the bias correction and the temperature influence on the point cloud stability. This is the first step towards a standardization of the metrological characterization procedure that could ensure the robustness and comparability of the results among tests and different devices. The procedure was conducted on Kinect Azure, Basler Blaze 101, and Basler ToF 640 cameras. Moreover, we compared the devices in the task of 3D reconstruction following a procedure involving the measure of both an object and a human upper-body-shaped mannequin. The experiment highlighted that, despite the results of the previously conducted metrological characterization, some devices showed evident difficulties in reconstructing the target objects. Thus, we proved that performing a rigorous evaluation procedure similar to the one proposed in this paper is always necessary when choosing the right device.

Keywords: machine vision; human-body reconstruction; time-of-flight; metrological characterization; Azure Kinect; Basler Blaze 101



Citation: Pasinetti, S.; Nuzzi, C.; Luchetti, A.; Zanetti, M.; Lancini, M.; De Cecco, M. Experimental Procedure for the Metrological Characterization of Time-of-Flight Cameras for Human Body 3D Measurements. *Sensors* **2023**, *23*, 538. <https://doi.org/10.3390/s23010538>

Academic Editor: Gregorij Kurillo

Received: 7 October 2022

Revised: 20 December 2022

Accepted: 28 December 2022

Published: 3 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the fourth industrial revolution started, industrial manufacturing technologies have evolved rapidly. Robots and machines are now equipped with intelligence and sensing devices such as cameras to see their environment. This is especially important when the machine operates alongside human workers. Therefore, the human body reconstruction capability of the camera is necessary for safety reasons and to guarantee the correct execution of detection and monitoring software. Typically, RGB cameras are the common choice for most industrial applications due to their competitive prices and low computational complexity. However, they lack information about the actual dimension of the environment, the relative distance of objects and their volume. Therefore, for some applications, 3D cameras are best suited for the task.

Three-dimensional cameras have been intensively adopted for years as measurement systems in a wide variety of applications. The exploited technologies are structured light, stereoscopy, and time-of-flight [1,2]. Time-of-flight (ToF) cameras are optical devices that measure the distance of objects from the sensor based on the calculation of the elapsed time between the emission/reflection of a light source. Compared to stereoscopy-based

devices, they are best suited for indoor environments due to the possible interference from direct natural light. However, the depth image obtained from these devices is affected by artifacts due to quantization in the range of low amplitudes. This issue affects ToF technology due to large operating distances and low reflectivity. Moreover, the presence of bias and other systematic errors is caused by anharmonic signals and overexposure [3,4]. Illumination conditions, such as the target's color properties and material reflectivity, are external error sources that cause depth errors, which increase linearly with distance [5]. Other common artifacts are flying pixels (erroneous depth estimates that appear close to the edge discontinuities in depth data) and multipath errors [6,7].

Two types of ToF cameras are available on the market: consumer-end and industrial. The main difference lies in the communication protocol: a Gigabit Ethernet connection is adopted by industrial cameras in contrast to USB 3.0 available in consumer-end devices. Industrial cameras are specifically designed for harsh environments; hence, the sensor and its lenses are encapsulated in a high-protection cover adhering to the IP protection index standard. They are typically adopted to reconstruct objects with high accuracy in controlled illumination and temperature conditions to guarantee the best performance. As such, their performance is often unsatisfactory at high distances and for moving subjects. In contrast, consumer-end devices are best suited for human activities, also thanks to the gaming industry, which intensively adopted this technology to enhance their user experience. As a result, consumer-end devices are compact and easy to use. Their 3D reconstruction capabilities are worse in terms of the accuracy of small objects and unusual surface materials; however, they work well with moving subjects. The most famous consumer-end device is the one developed by Microsoft, which has been improved over the years. Since the release of the Microsoft Kinect v2 sensor [8] in 2013, this commercial device has been intensively used for research purposes due to its affordability and performance [9]. For example, it has been used in 3D reconstruction for object modeling [10–13] and indoor scenes [14,15] and mobile robots' navigation and mapping [16–19]. The industrial applications include palletizing tasks [20,21], safety [22–24], teleoperation [25,26], human body detection and tracking [27–29], and gesture recognition tasks [30–32]. Healthcare applications involve gait analysis and elderly monitoring [33–36], the reconstruction of human body kinematics thanks to augmented and virtual reality software based on Kinect v2 [37,38]. However, Microsoft interrupted the production of Kinect v2 devices in favor of its new product Kinect Azure released in 2020. Currently, only a few works performed a metrological characterization of the new sensor. The work presented in [39] describes a set of experiments focused on gait analysis aimed at comparing the performance of the new Kinect Azure with the old Kinect v2 and a Vicon system. The authors of [40] performed a thorough characterization of the Kinect Azure sensor compared with its predecessors, Kinect v1 [41] and Kinect v2. They evaluated the three sensors testing their depth repeatability, noise-to-reflectivity, warm-up time, depth precision, reflectivity sensitivity, lens aberration, indoor versus outdoor performance, and flying pixel error. Finally, in [42], the authors explored the depth errors of Kinect Azure in comparison with Kinect v2.

To our knowledge, there is no standard experimental procedure to determine the typical error sources of this technology (namely the temperature influence and depth-related errors). As a result, researchers often come up with their own procedure resulting in non-comparable results and inconsistencies, spreading confusion among the scientific community. For example, without a careful characterization of ToF devices (or a document clearly stating their metrological properties), researchers may end up using the camera of choice without considering the intrinsic errors this technology implies, errors that can be corrected in post-processing if known and estimated beforehand. Working with depth devices without this knowledge may lead researchers to reach wrong conclusions in their own works, for example, when bias correction is not correctly estimated and corrected or when temperature influence is not considered when designing experiments. Another example is when ToF devices are adopted as monitoring systems in industrial workspaces, leading to incorrect depth estimation that may result in harmful behavior of

robotic systems and machines or in healthcare applications when patients' body volume is incorrectly estimated. The datasheet information is typically not enough in the case of particular tasks; thus, a metrological characterization is always needed to ensure the success of the application. Therefore, our first contribution is the proposed experimental set-up and characterization procedure described in Sections 2–4.

Second, there is little to nonscientific literature that compares consumer-end ToF cameras with industrial ones because the two have very specific application fields that usually do not overlap. However, because of the Industry 4.0 paradigm, there are a plethora of new applications and requirements that challenge both worlds, thus making this comparison meaningful for both researchers and practitioners helping them choose the suitable device for their task. Therefore, the investigation conducted compares the performance and metrological characteristics of two industrial Basler ToF cameras with the new Kinect Azure in a set of experiments based on our previous work conducted on different ToF cameras [43].

Finally, considering the challenging task of human recognition and tracking in industrial environments [44], the devices have also been tested in two 3D reconstruction set-ups to evaluate the quality of their point clouds when used to reconstruct both geometrical objects and human body segments. This is, in fact, one of the tricky yet more interesting applications in which ToF cameras are typically involved as the metrological device to measure the objects' dimension and volume. To this aim, the experimental procedure conducted to evaluate the correctness of both objects and human shapes dimensions is detailed in Section 5. The results are compared with the reconstruction obtained with a high-performing digitizer (gold standard).

2. Materials and Methods

ToF measurement is performed using a continuous-wave modulation based on the phase-shifting principle [45]. A periodic wave is emitted from the device and, after hitting an object, is sent back to the system. The resulting distance is calculated by analyzing the time between the emission of the wave and the corresponding received signal.

2.1. Specifications of the Evaluated Sensors

The cameras analyzed in this work are (i) Microsoft Kinect Azure (consumer-end), (ii) Basler Blaze 101 (industrial), and (iii) Basler ToF 640 (industrial). Table 1 details the main technical characteristics of the three devices.

Table 1. Summary of cameras' depth mode technical characteristics.

	Kinect Azure NFOV Unb.	Basler ToF 640	Basler Blaze 101
Resolution	640 × 576 px	640 × 480 px	640 × 480 px
Frame rate	30 fps	20 fps	30 fps
FoV	75 × 65 deg	57 × 43 deg	67 × 51 deg
Working range	0.5–3.86 m	0.5–5.8 m	0.5–5.5 m
Dimension	103 × 39 × 126 mm	141.9 × 76.4 × 61.5 mm	100 × 81 × 64 mm
Power	5.9 W	15 W	22 W
Weight	0.440 kg	0.400 kg	0.690 kg

It is worth noting that, compared to Kinect v2, Kinect Azure's depth camera may be used in two modalities, which may be binned or unbinned: NFOV (narrow field-of-view) and WFOV (wide field-of-view). However, in the context of this work, only the NFOV unbinned modality was used to perform the tests since the binning operation is applied by the SDK. The depth sensor adopted by the camera is based on the one detailed in [46].

The Basler Blaze 101 ToF camera mounts a Sony DepthSense IMX556 sensor, which makes the camera more robust to natural light. According to the datasheet, the optimal operating range is 0.5–5.5 m, where a depth precision of ±5 mm is guaranteed.

The Basler ToF 640 industrial camera is based on pulsed ToF technology adopting a Panasonic CCD depth sensor. It is optimized to work in indoor environments due to the technology's sensitivity to natural light. The optimal operating range is 0.5–5.8 m, where a depth precision of ± 10 mm is guaranteed.

2.2. Evaluation of Error Sources and Performance

The error sources are subdivided into (i) systematic and non-systematic errors according to their nature [47], (ii) camera-dependent errors, (iii) and scene-dependent errors [48]. However, in this study, depth errors related to fixed pattern noise and internal light scattering have not been evaluated.

The study detailed in [8] shows a metrological characterization based on the Guide to the expression of Uncertainty Measurements [49] for the uncertainty analysis of 3D scene reconstruction. Moreover, in [43], a metrological characterization comparing the performance of Kinect v2 and Picoflexx was performed. As in [43], the error sources considered are (i) temperature-related errors and (ii) depth-related errors.

Temperature-related errors are systematic and camera-related errors that are relevant for ToF cameras because their technology is strongly affected by heat [3–5,9]. The internal temperature of the camera is due to the heating of the illumination unit and image sensor, which produces drifts in the depth measure. After reaching a stable temperature, the components' characteristics do not change anymore. Consequently, it is suggested by the literature [48,50–52] to use ToF cameras after a warm-up time to obtain stable depth readings. However, the warm-up time is different according to the device, and some may compensate internally for this effect while others may not.

Except for the temperature-related errors, the other three error sources analyzed in this study are all depth-related: (i) depth amplitude error, (ii) depth distortion, and (iii) temporal error.

Depth amplitude is a systematic and camera-related error because the precision of depth measurements depends on the amount of light observed on each pixel [47]. Both underexposed and overexposed amplitudes may result in depth discrepancies since the illumination intensity is the highest at the center of the image and grows weaker around the borders. This effect leads to the overestimation of the depth values around the edges. However, if the object is very close to the emitter, the observed intensity may be higher, leading to pixel saturation.

Depth distortion is a systematic and camera-related error that occurs when the emitted light (typically a sinusoidal signal) is not generated correctly due to irregularities in the modulation process. Therefore, an offset is produced that only depends on the measured depth observed at each pixel. This error is assessed by comparing the depth measurements from devices with a reference ground truth distance [47,48].

Temporal errors are non-systematic camera-related errors that represent the depth variation of a pixel over time caused by measurement noise, which is more evident when the scene illumination is non-uniform, or the observed surface has low reflectivity [48]. This error also depends on the depth uniformity of the scene and on the integration time.

The overall measurement uncertainty was evaluated following the methodology presented in [8]. In addition, two other tests were performed to evaluate the performance of the sensors when used for (i) 3D reconstruction and (ii) body kinematic measurements. In the first experiment, the capability of the devices to reliably reconstruct and measure body segments is determined, while the second is aimed at evaluating the performance in human body segmentation.

2.3. Measuring Set-Up

Figure 1 shows a scheme of the setup. The three cameras were tested indoors at an environment temperature of 24 °C. An opaque white sheet of paper was used as the target for the acquisitions, mounted on a planar panel with verified planarity of 0.1 mm around the center area of 400 × 400 mm. Each camera was mounted at a fixed height of 1.5 m

from the ground, and they were oriented perpendicularly with respect to the target using a tripod with integrated bubble levels. A laser rangefinder EXTECH DT40M with an accuracy of ± 2 mm was used to verify the nominal position D_n of the camera with respect to the target according to the experiment. The overall illumination of the scene was kept constant without the influence of natural light.

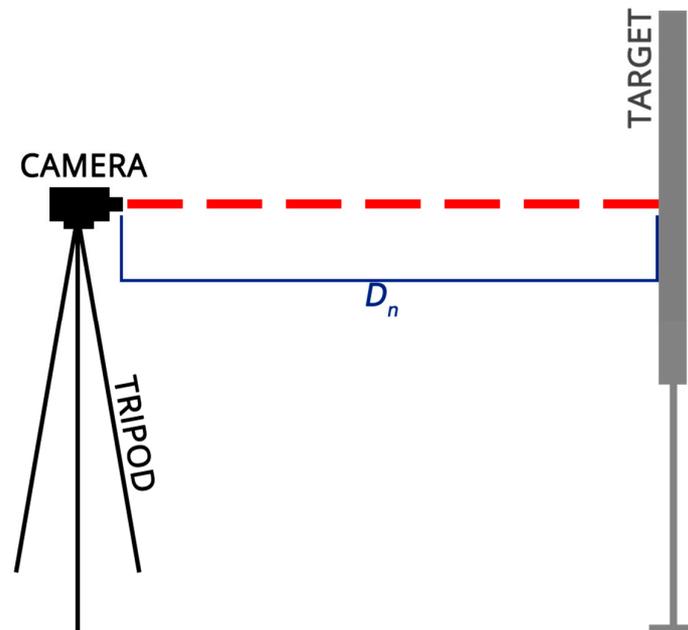


Figure 1. Scheme of the measuring set-up.

3. Evaluation of Temperature-Related Errors

Referring to the set-up in Figure 1, in this experiment, the cameras were positioned one at a time at a constant nominal distance of $D_n = 2$ m, corresponding to half of their operating range. They were kept in the test area while turned off for at least 4 h before starting the experiments.

The experiment lasted 2 h for each camera. The images were acquired every 10 s at 30 fps. Frames belonging to a time window of 5 min were grouped together, resulting in groups of 30 frames. For each, a region of interest of 15×15 px centered around the central pixel of the image was extracted. Finally, for each group the mean depth value μ_t and the standard deviation σ_t was computed, resulting in 24 data points d_t .

Depth measurements performed by ToF cameras are typically affected by a systematic error (bias). Thus, the measured depth must be corrected by this quantity to obtain correct readings. Bias estimation, in this case, was conducted using the following formula:

$$\begin{aligned} \mu_t &= \frac{1}{5} \sum_{i=19}^{24} d_t \\ b &= \mu_t - D_n \\ d_t^* &= d_t - b \end{aligned} \quad (1)$$

We chose to calculate μ_t as the mean of the last five data points d_t because they are the most stable. Otherwise, the bias correction would have taken in consideration the great difference between the actual distance D_n and the measured distance D_m that is observed in the first data points before the warm-up kicks in (see Figure 2). Hence, the last five d_t correspond to the time intervals after the warm-up time of the cameras. The mean depth values after bias correction d_t^* are represented in Figure 2.

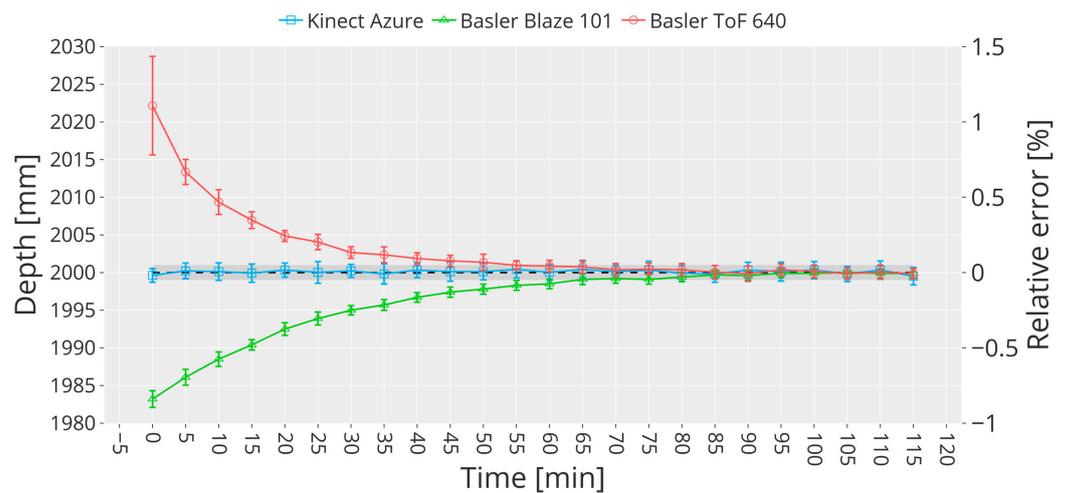


Figure 2. Temperature-related error. The plot shows the depth measurements obtained from the three ToF cameras considering a time interval of 5 min. Bias correction applied.

Surprisingly, Kinect Azure does not need to warm up to obtain a stable output, meaning that the depth readings are not correlated with the device temperature. This result contrasts with the work presented in [40], where the authors stated that the device needed at least 60 min of warm-up. However, in our experiment, Kinect Azure achieves stable values from the beginning without the need for a warm-up time. In fact, the bias correction for Kinect Azure could have been performed considering all the 24 data points d_t in Equation (1) instead of the last five only since they were not significantly different. The experiment was performed twice to confirm this finding, highlighting the robustness of our measuring method in contrast with reference [40]. A possible reason to explain why we obtained a significantly different result is that the bias was not corrected in [40]. Furthermore, in our experiment, the relative distance camera-target was $D_n = 2$ m, while in [40], they used $D_n = 0.8$ m. Environmental characteristics could have impacted the measurement, for example, if the ambient illuminance interfered with Azure's IR rays. In our case, even if the camera is used continuously for almost 120 min, the output is stable in terms of deviation from the actual distance, revealing a small wiggling behavior that leads to outputs that differ from the nominal distance of about -15 mm (around 1 mm after bias correction). Moreover, according to the experimental comparison in [7], the same behavior could be observed for Kinect v1, while the Kinect v2 mean depth data show a strong correlation with the device temperature (stable readings after 25 min). On the other hand, both the Basler Blaze 101 and the Basler ToF 640 cameras need a warm-up time of at least 50 min. It is worth noting that the deviation (shown as error bars) decreases with time for both Basler Blaze 101 and Basler ToF 640, but this could not be said for Kinect Azure, which shows stable deviation values uncorrelated with the device temperature. The reason for this different behavior could be that industrial cameras' design considers the hazardous environment in which they could be deployed; hence, manufacturers may not account for temperature-related errors internally due to lack of space or design limitations.

4. Evaluation of Depth-Related Errors

Referring to the setup in Figure 1, in this set of experiments, the three cameras were placed at 15 nominal positions D_n in the range of 1.7–4.5 m from the planar target, with a step of 0.2 m. For each nominal position, a total of 30 frames were recorded at 30 fps.

4.1. Depth Amplitude Errors Evaluation

To evaluate the depth amplitude of each sensor, it is necessary to analyze both the quality of the captured IR image and the corresponding depth map. In fact, depth accuracy is related to the amount of light received by each pixel because both under and over-exposed pixels result in depth errors.

On the one hand, IR images contain the light intensity of the reflected ray emitted from the ToF camera. Observing the images on the left in Figure 3, it is evident that the intensity of the captured light decreases around the corners of the images while it is at its maximum in the center. In the case of Kinect Azure, the dark circles appearing in the image are due to the visualization of the image but are not an error source, while the black patches on the corners of the image are due to the reduced field of view that happens in the NFOV modality and are absent in WFOV. In the context of this work, only the NFOV was evaluated since its characteristics are similar to those of the other cameras. In contrast, for the Basler Blaze 101 and the Basler ToF 640 cameras, the resulting intensity image is lighter and more uniform. Considering that the images in Figure 3 refer to a $D_n = 1.7$ m, the differences in the pictures are due to (i) the optics mounted on the cameras, (ii) the different field of view, and (iii) the positioning of the tripod along the x -axis with respect to the target; hence, the scene is captured differently. This is also the reason why only a portion of the frame around the central pixel was considered for computing our analysis of the depth frames.

On the other hand, depth images contain the measured depth value corresponding to each pixel. In this work, they were obtained by analyzing the point cloud because resolution and data accuracy are higher, especially for the two industrial cameras. The resulting depth amplitude error is estimated from these depth images. Considering the experimental set-up described in Section 4, to each nominal camera position D_n corresponds to a total of $m = 30$ measured depth maps D_m . Hence, to each D_n corresponds an average depth map μ_m computed by:

$$\mu_m = \frac{1}{m} \sum_{i=1}^m D_i \quad (2)$$

The resulting data must be centered around zero, so for each D_m we first compute:

$$\varepsilon_m = \mu_m - D_n \quad (3)$$

Then, we calculate the average ε_m as:

$$\overline{\varepsilon_m} = \frac{1}{m} \sum_{i=1}^m \varepsilon_m \quad (4)$$

Finally, for each nominal distance D_n we obtain the error map (centered around zero) by removing the mean:

$$\varepsilon_n = \varepsilon_m - \overline{\varepsilon_m} \quad (5)$$

The images on the right in Figure 3 represent the depth amplitude error calculated following the abovementioned procedure. In the case of Kinect Azure, the depth amplitude error at $D_n = 1.7$ m is mostly concentrated around ± 5 mm. On the upper corners, we can observe peaks of -17 to $+19$ mm due to reflections occurring on the surface. Basler Blaze depth is denser thanks to an increased point cloud resolution; however, more than half of the image shows an overestimation of $+5$ to $+13$ mm, probably due to the higher amount of reflected light in this area. It is also possible to observe concentric waves due to the non-ideal wave generated by the camera's emitter, which may be another error source causing this overestimation. The peak of -31 mm refers to a non-planar edge of the target. Finally, the depth amplitude error map of Basler ToF 640 shows that most pixels are underestimated by -5 to -19 mm, especially in the center area, while the angles of the target are overestimated by $+20$ to $+32$ mm. It is worth noting that even if the camera-target alignment was ensured before the acquisition, displacements might have occurred without us noticing. This is the reason why for the analysis described in the following Sections, we only consider a sub-portion of the depth map. Furthermore, we checked the depth amplitude error in a region of interest of 40×40 mm centered around the central pixel of the depth map. In this area, Kinect Azure's error ranges from 2 to 5 mm, Basler Blaze 101's

ranges from 3 to 7 mm, and Basler ToF 640's ranges from -10 to -6 mm. This region has been chosen to obtain comparable results with respect to the cameras' datasheets.

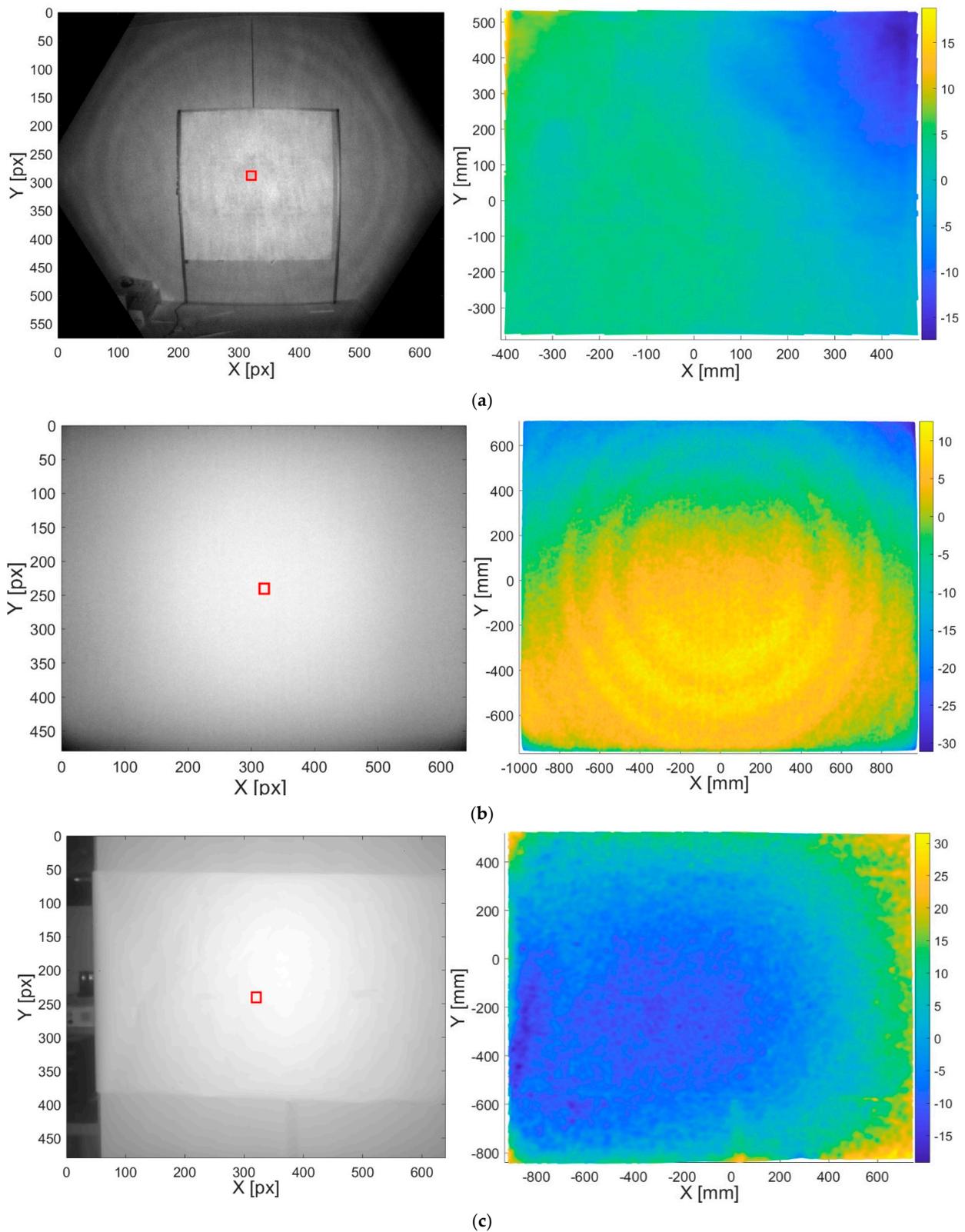


Figure 3. IR and Depth amplitude images of the target at $D_n = 1.7$ m acquired with (a) Kinect Azure, (b) Basler Blaze 101, (c) Basler ToF 640. The central region of interest is shown in red.

4.2. Depth Distortion Evaluation

Depth distortion errors typically increase with distance. Therefore, it is important to evaluate the relative trend of depth values according to the nominal positions D_n instead of the absolute values. Considering the experimental set-up described in Section 4 and the mathematical procedure detailed in Section 4.1 to compute ε_n , in this case each nominal camera position D_n corresponds to a total of $m = 30$ measured depth values D_m calculated considering only the central pixel of each frame [43].

Figure 4 shows the resulting depth distortion error ε_n of the three cameras for each nominal position. For Kinect Azure and Basler Blaze 101, the error is very low, showing a wiggling trend that is more prominent in the case of Azure Kinect. The depth distortion of Kinect Azure spans from -18 mm to 10 mm, which corresponds to the nominal distances of 2.7 m and 3.3 m, respectively. In the case of Basler Blaze 101, it has an almost constant behavior and spans from -11 mm to 9 mm, corresponding to the nominal distances of 1.7 m and 4.5 m, respectively. However, for Basler ToF 640, the distortion error is evident since the trend increases with distance. It ranges from -48 mm to 76 mm, corresponding to nominal distances of 1.7 m and 4.5 m.

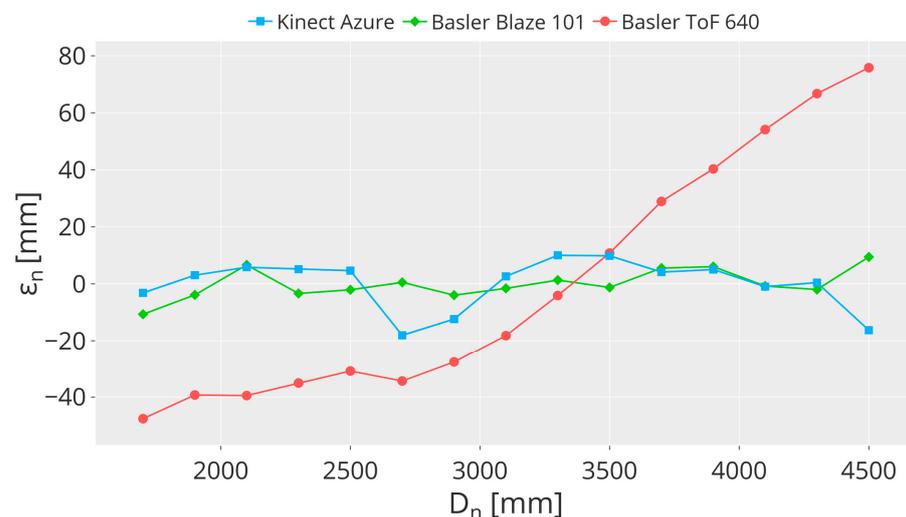


Figure 4. Plot showing the depth distortion error ε_n computed for each camera for each D_n .

4.3. Temporal Errors Evaluation

Temporal errors refer to fluctuations of the depth values due to measurement noise. Hence, considering the experiment described in Section 4.1, the temporal error is the standard deviation σ_m over 30 frames for each measured position D_m :

$$\sigma_m = \sqrt{\frac{1}{m} \sum_{i=1}^m (D_i - \mu_m)^2} \quad (6)$$

The mean depth value μ_m appearing in this formula is obtained from Equation (2).

The deviation values are shown in Figure 5 according to the μ_m values. For all the cameras, the standard deviation σ_m has an increasing trend. The results for both the Kinect Azure and Basler Blaze 101 cameras are similar. On the other hand, the standard deviation of the Basler ToF 640 camera is sometimes very high ($\sigma_m = 3.75$ mm for $\mu_m = 3429$ mm) or very low ($\sigma_m = 0.61$ mm for $\mu_m = 1786$ mm). The values of σ_m are slightly higher for the Kinect Azure, which has values in a range of 0.81 mm and 3.19 mm versus the σ_m computed for the Basler Blaze 101, which is in a range of 0.52 mm and 2.99 mm. Moreover, the trend lines (represented by the black dashed lines) are obtained as a linear regression over values σ_m . The regression coefficients R^2 are equal to 90% (slope 2.67, intercept 1.01 mm), 91% (slope 2.59, intercept 1.62 mm), and 48% (slope 2.64, intercept 0.52 mm) for the Kinect Azure, Basler Blaze 101, and Basler ToF 640, respectively. This shows that the performance

of the Kinect Azure and the Basler Blaze 101 are equivalent, while highlighting that the Basler ToF 640 has higher variability compared to the other two cameras.

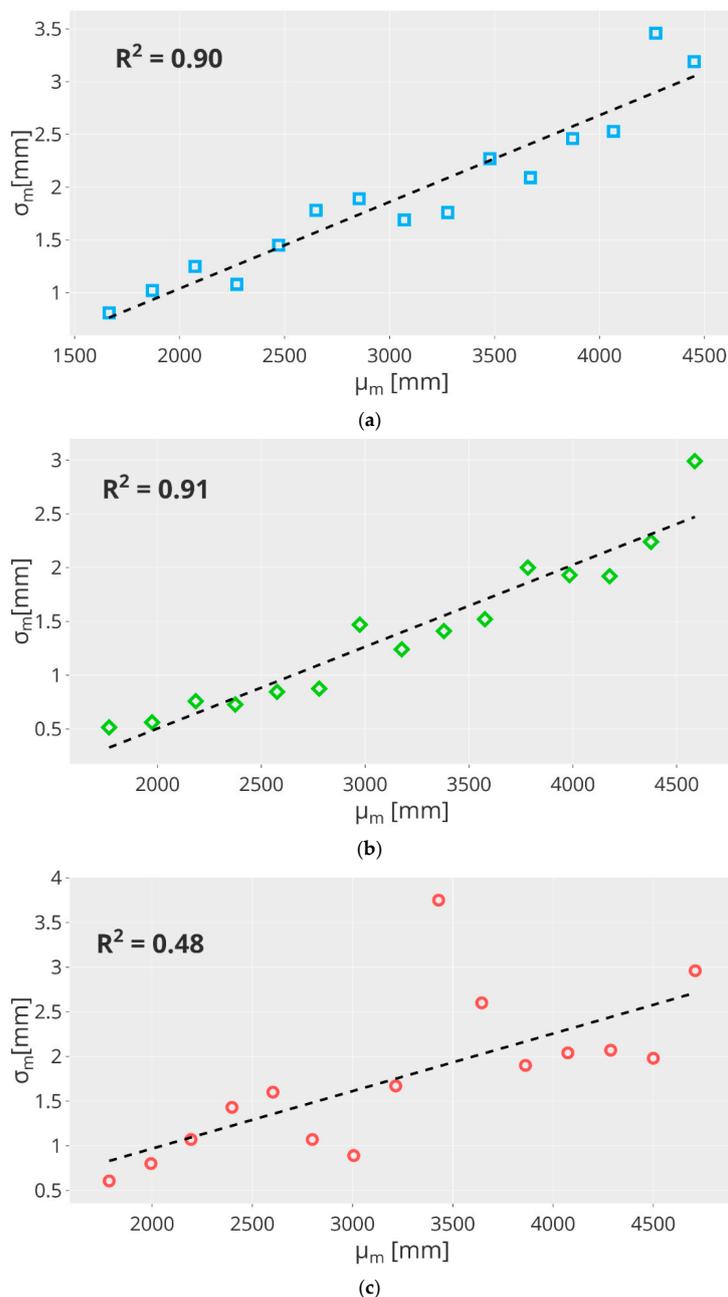


Figure 5. Plot showing the temporal error as a function of depth for (a) Kinect Azure, (b) Basler Blaze 101, and (c) Basler ToF 640.

It is worth noting that Kinect Azure's results adhere to those obtained in [42], which show equivalent standard deviation values that increase with distance.

4.4. Overall Depth Measurement Uncertainty Evaluation

To obtain the overall measurement uncertainty, a neighborhood of 20×20 px centered around the central pixel of the frame was extracted, and each depth value belonging to this squared region was considered, resulting in 400 data points per frame. The neighborhood is shown in red in Figure 3.

For each D_n , 30 frames were taken in the range of 1.7–4.5 m with a step of 0.2 m corresponding to 15 positions. The collected measurements are shown in Figure 6 for the three cameras and correspond to a total of around 168,000 data points (except the outliers, which were removed from the calculations). The results show that the Kinect Azure has very little dispersion since the measured depth values corresponding to the squared blue markers (Figure 6a) have small variation. However, there are few occurrences of incorrect readings observed only for nominal distances greater than 4.1 m due to some reflections occurring in the scene. This resulted in a mean deviation σ_0 equal to 13 mm, which is slightly lower than the value obtained for the Kinect v2 sensor resulting from the same experiment carried out in [43], which reported a σ_0 value equal to 18 mm. The dispersion observed in the case of the Basler Blaze 101 camera is even better, as shown by the reduced variability of the green diamond markers (Figure 6b). The sensor seems less prone to dispersion effects, and its σ_0 is equal to 6 mm. Nonetheless, a high number of incorrect readings appear for nominal distances greater than 3.1 m. Compared to Kinect Azure, this effect seems more generally distributed since the incorrect measures span a wider range (D_n in a range of 3.1–4.5 m corresponding to a D_m range of 0–0.8 m) than in the case of Kinect Azure (D_n in a range of 4.1–4.5 m corresponding to a D_m range of 0.8–1.6 m). In contrast, the Basler ToF 640 camera has higher dispersion (bottom image in Figure 6c) but it does not achieve incorrect depth values, resulting in a σ_0 of 13 mm, which is the lowest of the three. Moreover, the linear regression computed for the three devices and represented by the black dashed line in each plot highlights their measurement linearity. The corresponding R^2 coefficients are 99.98%, 100%, and 99.98% for the Kinect Azure, Basler Blaze 101, and Basler ToF 640 cameras, respectively. It is worth noting that the outliers observed were removed before performing the analysis.

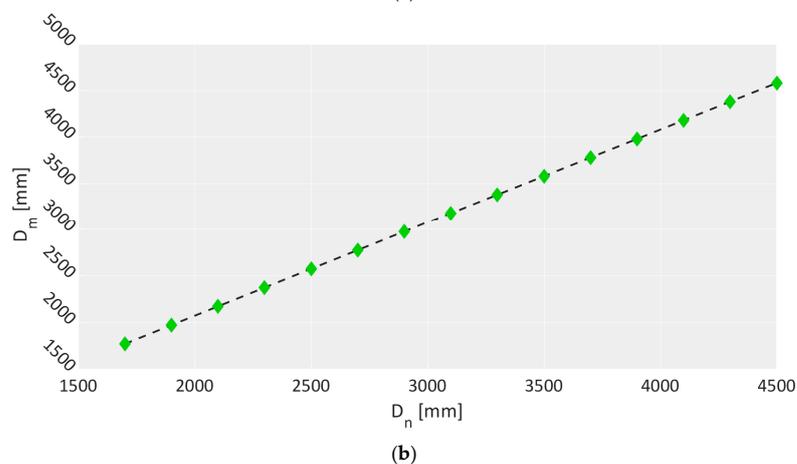
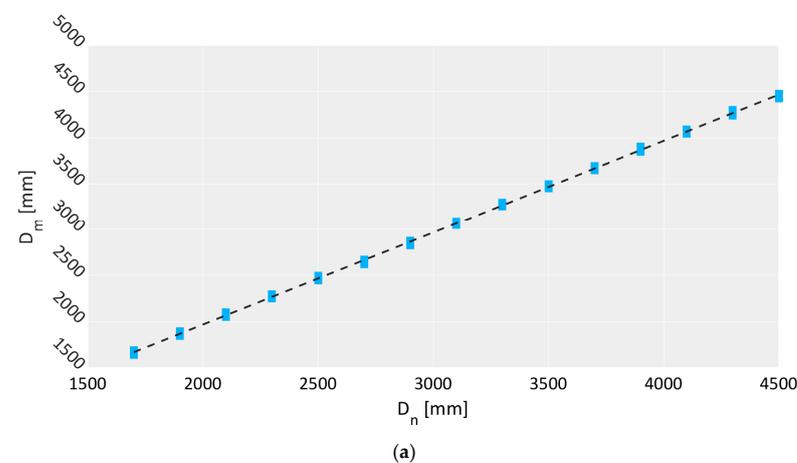


Figure 6. Cont.

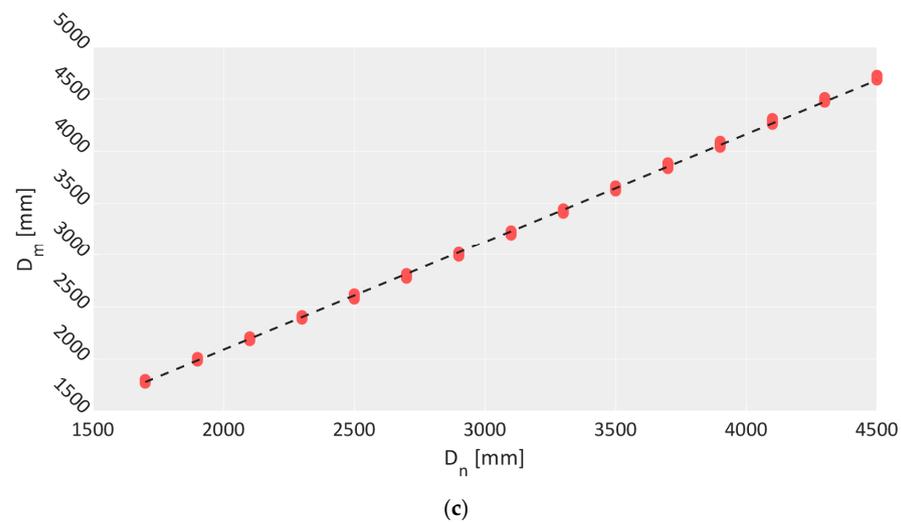


Figure 6. Measured depths D_m as a function of nominal depths D_n for (a) Kinect Azure, (b) Basler Blaze 101, (c) Basler ToF 640.

5. Application Example: 3D Reconstruction

ToF devices are typically used indoors to perform a plethora of measurement tasks, for example, the relative distance monitoring between human subjects and moving machines [23] and object volume and size estimation [53]. Depth cameras are also extremely important for the healthcare sector since human body reconstruction and volume estimation are crucial for the gait analysis of patients with reduced mobility [35] and for the evaluation of particular diseases involving malformations of the body [54].

A careful metrological characterization of the ToF camera of choice following the procedure described in Sections 2–4 is needed to determine the error sources and to design the application set-up in order to obtain correct data. However, in the case of 3D reconstruction, the assessment of error sources may not be enough to determine which device is best suited for the task. Therefore, in this Section, we propose two experimental set-ups in which the three cameras of our choice (Kinect Azure, Basler Blaze 101, and Basler ToF 640) are compared in terms of 3D reconstruction capabilities.

5.1. Object Reconstruction

This experiment aims to evaluate the sensors' capabilities to accurately reconstruct objects from the acquired point cloud. The cylindrical object used in [55] with an external radius of 122 mm was used as the measurement target. The cylinder was industrially produced, and its radius was measured with a caliber with a 0.01 mm resolution. Considering the field-of-view (FoV) and range of the cameras, the cylinder was placed at 15 positions spanning the FoV symmetrically. However, since the reconstruction performance may vary according to the positioning of the target with respect to the camera in the vertical direction, two set-ups were considered where the bottom end of the cylinder is placed (i) at 0.7 m from the floor (odd stations numbers), and (ii) at 1.5 m (even stations numbers) with the aid of an adjustable carrier. This results in the set-up shown in Figure 7, where the red dots represent odd stations, and the green dots represent even stations. Stations 19, 23, 25, and 29 were moved to allow the cylinder to fit inside the camera FoV (black dashed lines). It is worth noting that in this experiment, the aim was not to evaluate the multipath effect; hence, the cylinder was not positioned at floor height.

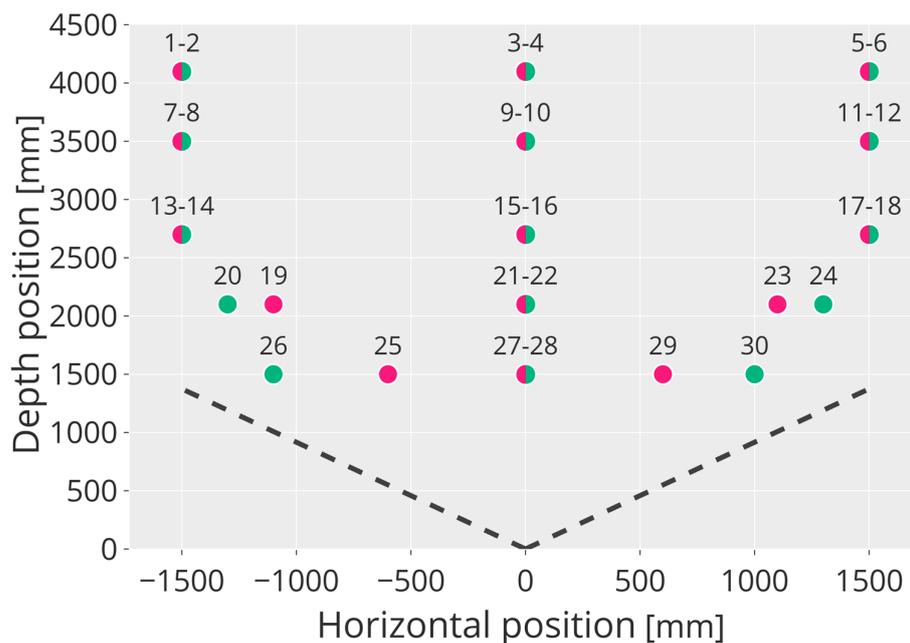


Figure 7. Cylinder positions adopted in the experiment. Red dots refer to odd station numbers, green dots to even station numbers.

A total of 30 frames at 30 fps for each station were acquired for each camera. By inspecting the data, it resulted that only the point clouds acquired by Kinect Azure were of enough quality for further analysis because the point clouds of the other two cameras were too noisy to allow a proper reconstruction of the object. For the Basler ToF 640 camera, the point cloud is affected by mixed pixel errors and multipath making it impossible to properly detect the target. In the case of the Basler Blaze 101 camera, the object is sometimes reconstructed at different depth values than the carrier, leading to incorrect readings. This is probably due to the target shape since industrial cameras are usually optimized to work with planar surfaces. Therefore, the following analysis was conducted only on Kinect Azure data:

1. Each point cloud was manually inspected to remove the elements of the scene not belonging to the cylinder. This was performed by applying a depth filter to cut off data outside the area of interest, thus obtaining only the point cloud of the cylinder.
2. The camera performance was evaluated by comparing the external radius of the measured cylinder with respect to the nominal one of 122 mm. The measured external radius was estimated individually for each acquisition by analyzing the point cloud with MATLAB using a cylindrical fit provided by the software.
3. For each station, the mean value over 30 frames of the external radius and the corresponding standard deviation was computed.

Figure 8 shows the measured external radius for the odd and even positions, respectively. The diameter of the colored circle represents the mean value μ_d of the measured diameter plus the standard deviation value σ_d (upper bound, *UB*). The diameter of the white hole represents the mean value μ_d minus the standard deviation value σ_d (lower bound, *LB*). If this subtraction results in values less than zero, no hole is drawn.

$$\begin{aligned} UB &= \mu_d + \sigma_d \\ LB &= \mu_d - \sigma_d \end{aligned} \quad (7)$$

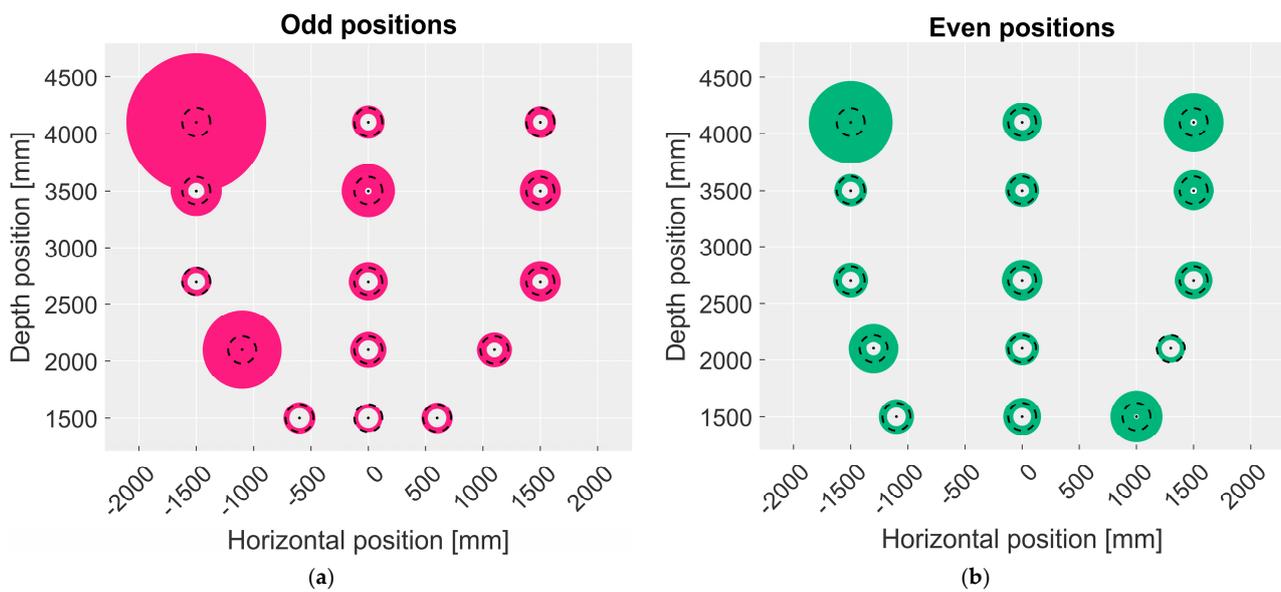


Figure 8. Image showing the upper bound of the cylinder’s diameter as the dimension of the colored circle, and the lower bound as the dimension of the white circle that results for each odd station. The dashed black circle represents the reference diameter. (a) Odd positions, (b) even positions.

The resulting mean values of the external radius span from 140 mm to 100 mm; however, the standard deviation is very high for stations 1, 2, 6, and 19, and it is above average for stations 7, 9, 11, 11, 20, and 30. Considering the camera operative range in NFOV unbinned mode, it is evident that the best performance is achieved when the cylinder is positioned at 1.5 m from the floor (even stations). Moreover, for stations positioned at depth values higher than 3.5 m, the standard deviation of the measure is higher, especially for stations closer to the edges of the FoV (1, 2, and 6, which have values of 432, 218, and 115 mm, respectively). It is unclear why, at positions 19, 20, and 30, the standard deviation is higher with respect to the other values achieved for stations inside the range of 1.5–3.0 m (values above 100 mm versus an average standard deviation of 50 mm). The reason may be that these are near the FoV edges. In conclusion, this experiment shows that Kinect Azure performance in 3D reconstruction is better at the center of its FoV, corresponding to the central area of the set-up.

5.2. Human Body Reconstruction

This experiment has been designed to evaluate the human body reconstruction capabilities according to the procedure described in [42]. A 70 × 30 cm mannequin representing the human upper body with movable arms was adopted for the test. Angles between its body segments were measured and compared with a reference measure taken by a commercial 3D digitizer Konica Minolta VIVID-920 with a resolution of 640 × 480 px and depth accuracy of ±0.40 mm. This device was chosen as the gold standard because its accuracy is one degree higher than the average measured depth values obtained from Kinect Azure.

Seven body segments can be extracted from the mannequin: (i) head and trunk, (ii) left forearm, (iii) right forearm, (iv) left arm, (v) right arm, (vi) left hand, (vii) and right hand. According to [43] and by considering the mannequin adopted to evaluate the human body kinematics, six absolute angles of interest should be considered:

- α_L : angle between the vertical axis and the left forearm.
- α_R : angle between the vertical axis and the right forearm.
- β_L : angle between the left forearm and the left arm.
- β_R : angle between the right forearm and the right arm.

- γ_L : angle between the left arm and the left hand.
- γ_R : angle between the right arm and the right hand.

The cameras were placed about 2 m from the mannequin, according to their FoV. The mannequin was moved in seven different poses, as shown in Figure 9, simulating a variety of human postures. For each configuration, a single point cloud was acquired. Points belonging to the mannequin were extracted using PolyWorks from each point cloud. Then, for each body segment, the Principal Component Analysis (PCA) was performed to compute the direction of their principal components considering the body axial symmetry [53,56]. Finally, the angles of interest were obtained by performing the scalar product between the principal components accordingly.

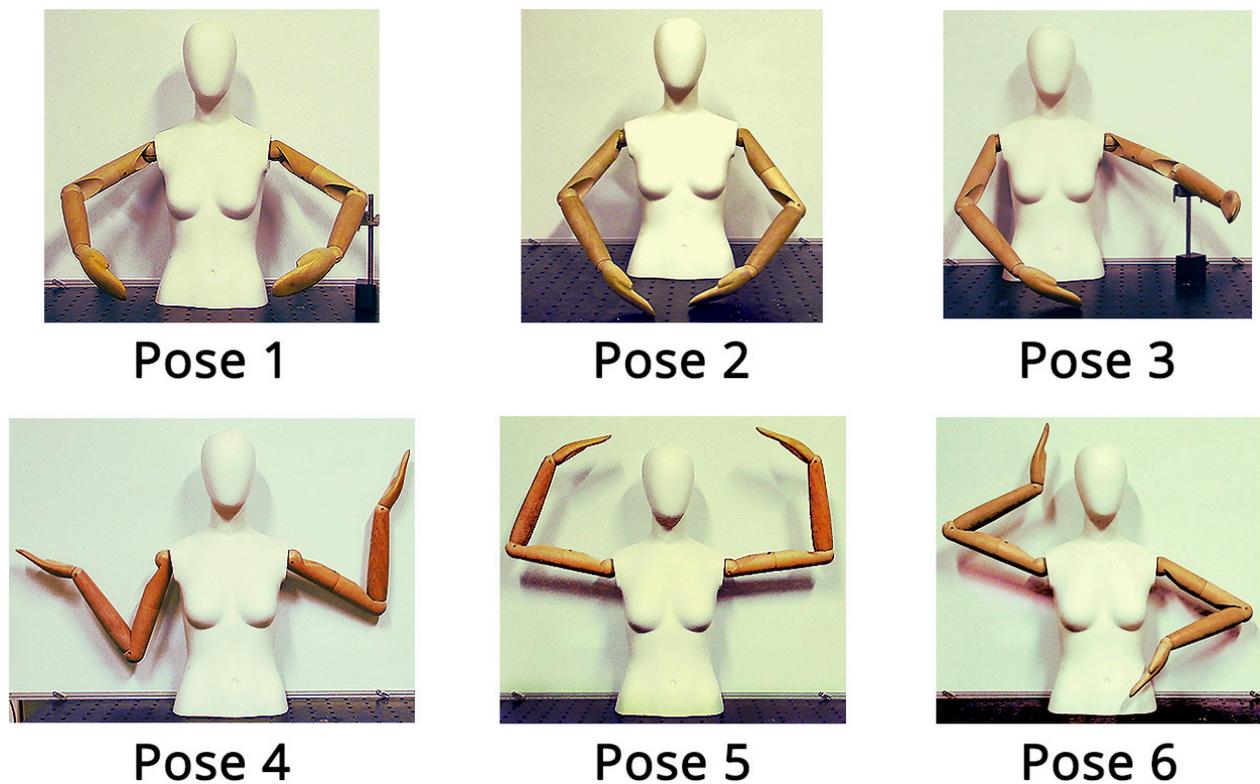


Figure 9. Dummy poses considered for the analysis as in [43].

In addition to the reference values obtained from the Konica Minolta, the results were compared with the values obtained by Kinect v2. Figure 10 shows the resulting values of the angles of interest for each mannequin configuration. From these bar plots, it is evident that Basler ToF 640 (red bar) and Kinect v2 (orange bar) are the ones that achieve the worst performance. In contrast, Basler Blaze 101 (green bar) and Azure Kinect (blue bar) show a very similar trend with Konica Minolta (purple bar, golden standard). Excluding the performance of Kinect v2, the easiest poses to measure are poses 4 and 5, while for poses 1, 2, and 3, the three cameras show values that are sometimes very different from each other. This depends on the pose itself: poses 1, 2, and 3 span along the z -axis more than the other, resulting in higher occlusions.

To better analyze the results, the differences between the resulting values and the corresponding angle measured by the Konica Minolta are shown in bar plots in Figure 11. For most configurations, the Kinect v2 measurements are noticeably different from the reference ones acquired with Konica Minolta, sometimes reaching values of more than -30 deg (measured angle higher than reference) and more than 50 deg (measured angle lower than reference). The β_L angle measured by Basler ToF 640 is the most different from the reference for all configurations, achieving errors ranging from more than -20 deg up

to more than 30 deg. In comparison, Basler Blaze 101 performs better, achieving errors of -10 deg up to 25 deg. Azure Kinect measurements are typically the same as those computed from Konica Minolta, with errors ranging from -10 deg up to 12 deg, mostly occurring in correspondence of angles γ_L and γ_R . These two angles consider the arm and the hand segments, which are difficult to measure accurately since the point clouds are less dense. These tests showed that Kinect Azure's performance in reconstructing human bodies is notably improved compared to Kinect v2, resulting in lower errors. It is worth noting that this analysis has been performed without the aid of the Kinect Azure SDK, which computes the human skeleton using a skeletonization algorithm. Instead, the body segments have been extracted from the point clouds by using PolyWorks and approximated as the principal component vector resulting from the PCA analysis. As a result, each body segment was considered as a vector facilitating the estimation of angles between them.

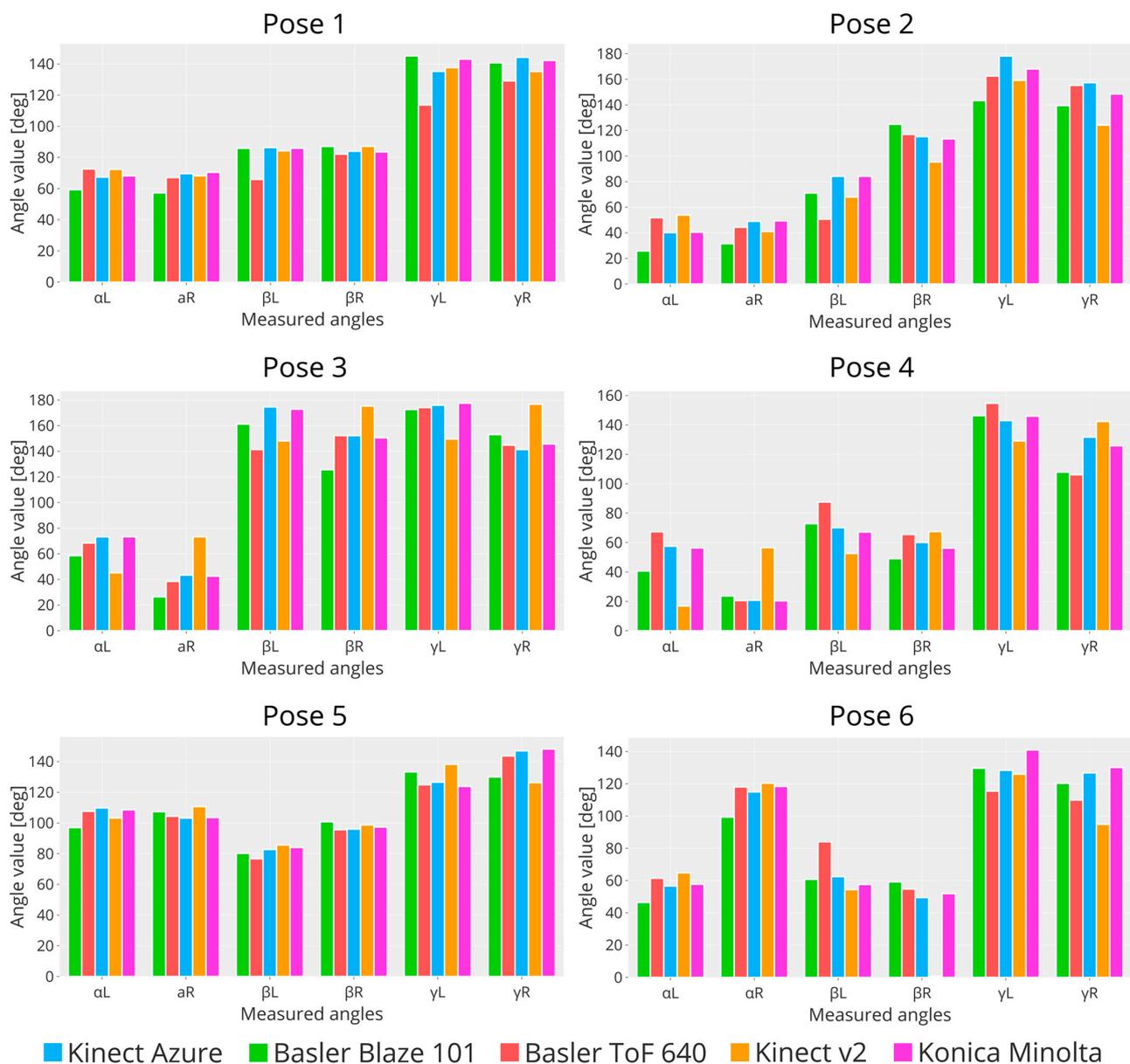


Figure 10. Graphs showing the measured angles of interest for the six poses.

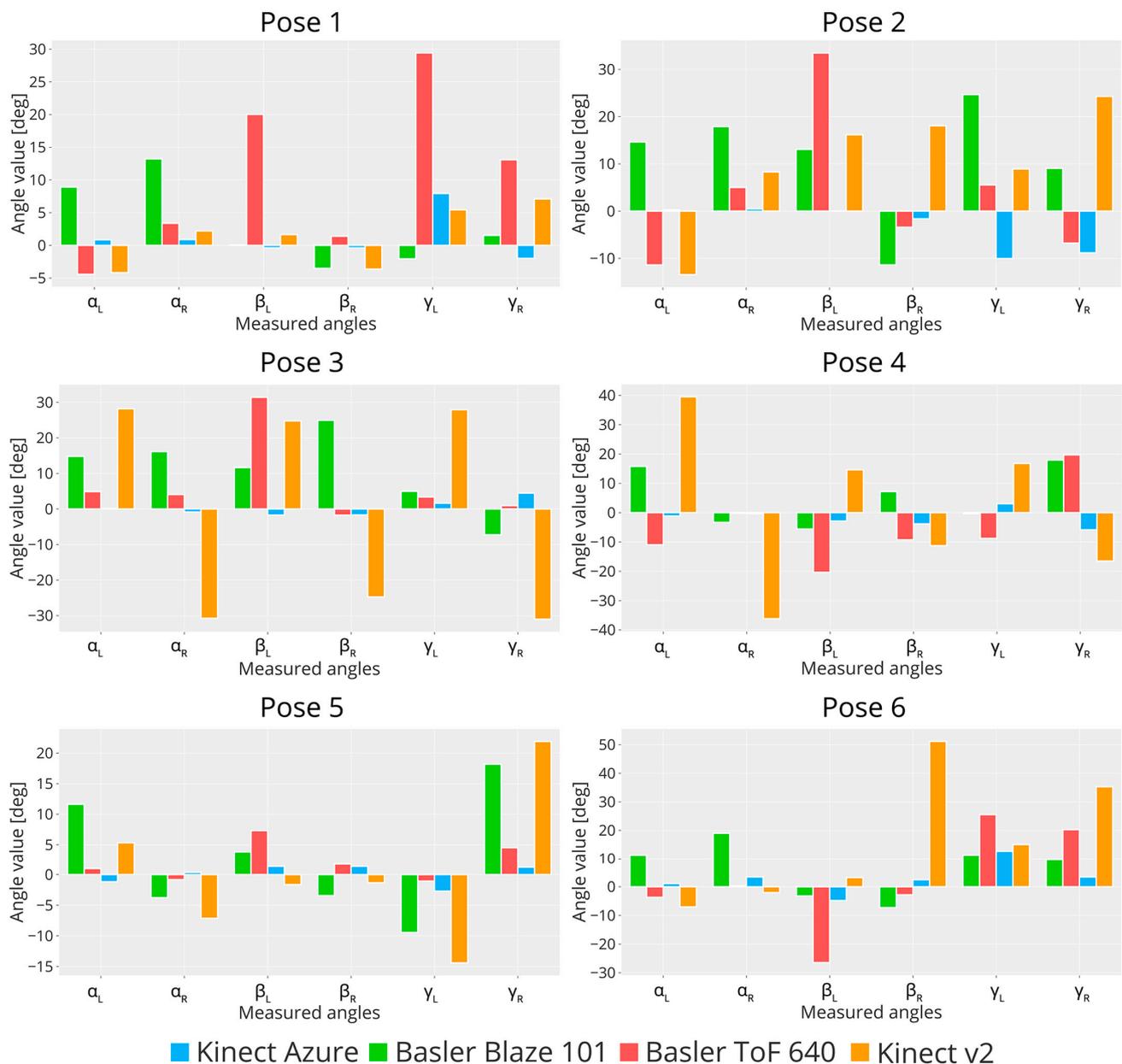


Figure 11. Bar plots representing the measuring error achieved by the three sensors and the old Kinect v2 for the six poses considered. The error is computed by subtracting the angle value obtained from Konica Minolta from the camera's measure.

6. Conclusions

This paper details an experimental procedure to assess the metrological characteristics of ToF cameras with respect to the typical error sources of this technology. This contribution is especially important to robustly compare results among devices. The procedure was conducted on a consumer-end device (Kinect Azure) and two industrial ones (Basler Blaze 101 and Basler ToF 640). Although comparing cameras belonging to different worlds may seem counter-intuitive, this choice is motivated by the necessity of new solutions for modern industry, which is moving towards innovative environments where both humans and machine collaborate.

Error sources such as temperature influence on depth measurement, depth distortion, depth amplitude errors, temporal errors, and overall depth measurement uncertainty were evaluated for the three devices in different experiments. A summary of the developed evaluation procedure can be found in Table 2 for quick reference. The results of each

camera are summarized in Table 3 in comparison with the corresponding datasheet and relevant references.

Table 2. Summary of the proposed ToF camera evaluation protocol.

	Warm-Up Time	Depth Amplitude	Depth Distortion	Temporal Error	Overall Uncertainty
Env. conditions	Ensure optimal temperature (i.e., 24 °C) Ensure constant illumination without natural light interference				
Reference target	Opaque target with verified planarity especially in the central region				
Hardware set-up	Fixed distance D_n Turn off camera before experiment for at least 4 h	Camera mounted on support at fixed height Ensure camera perpendicularity with respect to the target Define a set of distances D_n in the optimal working range according to the camera datasheet			
Data acquisition	1 depth frame or point cloud every 10 s at 30 fps	30 depth frames or point cloud at each D_n at 30 fps			
Data analysis	Group frames belonging to 5 min time windows (30 frames total) Extract 15×15 ROI around the central pixel Compute mean depth μ_t and standard deviation σ_t (Equation (1))	Extract target \forall frame \in to each D_n Compute error ε_n (Equations (2)–(5))	Extract only the depth value of the central pixel \forall frame \in to each D_n Compute error ε_n (Equations (2)–(5))	Extract only the depth value of the central pixel \forall frame \in to each D_n Compute deviation σ_m (Equation (6)) Compute linear regression and check R^2	Extract 20×20 ROI around the central pixel Use all data points inside ROI Compute linear regression and check R^2
Data correction	Apply bias correction and obtain d_t^* (Equation (1))	Ensure that ε_n is the relative error not the absolute depth		//	Remove outliers before applying linear regression
How to visualize	X-axis: time [s] Y-axis: d_t^* with corresponding σ_t [mm] Optional: secondary y-axis showing relative error [%]	IR image and Depth error X-axis: x coordinate [px] and [mm], respectively Y-axis: y coordinate [px] and [mm], respectively Show color bar	X-axis: distance D_n [mm] Y-axis: ε_n [mm]	X-axis: μ_m [mm] Y-axis: σ_m [mm] Show linear regression line	X-axis: D_n [mm] Y-axis: D_m [mm] Show linear regression line

The presence of flying pixel problems and multipath errors was observed in the point cloud acquisitions; however, they were not quantified in this study. From these results, we may conclude that the best and worst performing cameras among the three are Kinect Azure and Basler ToF 640, respectively, while Basler Blaze 101 achieves comparable results with respect to Kinect Azure.

Since the metrological characterization of error sources may not be enough to determine the right device for a target application, we proposed an example in which the cameras are used for 3D reconstruction. The first experiment involved a cylindrical object, and the aim was to correctly estimate its diameter from the point cloud at different heights and distances from the cameras. However, from this test, the point clouds of the two industrial cameras resulted in being too noisy to be acceptable for evaluation. This was

probably because they are best suited for objects with regular shapes placed at shorter distances from the light emitter. It is worth mentioning that from the error source characterization alone, it was not possible to predict such inconsistent behavior, thus highlighting the need for a standard experimental procedure for the assessment of ToF cameras' 3D reconstruction capabilities.

Table 3. Summary of the error sources influence resulting from our experiments in comparison with relevant references and cameras' datasheets. Asterisks refer to data obtained by our procedure. DS refers to data found in the device's datasheet.

	Warm-Up Time	Depth Amplitude	Depth Distortion	Temporal Error	Overall Uncertainty
Kinect Azure *	Not needed	2 to 5 mm	−18 to 10 mm	0.8 to 3.2 mm	13 mm
Kinect Azure DS	Not provided	Not provided	$<11 \text{ mm} + 0.1\% D_n$	$\leq 17 \text{ mm}$	Not provided
Kinect Azure [1]	50 min	Not provided	−7 to 0 mm	0.5 to 2 mm	Not provided
Kinect Azure [2]	Not provided	−2 to 0 mm	1.1 to 12.7 mm	0.6 to 3.7 mm	Not provided
Basler Blaze 101 *	50 min	3 to 7 mm	−11 to 9 mm	0.5 to 3 mm	6 mm
Basler Blaze 101 DS	20 min	−5 to 5 mm	Not provided	$<2 \text{ mm}$	Not provided
Basler ToF 640 *	50 min	−10 to −6 mm	−48 to 76 mm	0.6 to 3.8 mm	13 mm
Basler ToF 640 DS	20 min	−10 to 10 mm	Not provided	$\leq 8 \text{ mm}$	Not provided

The second experiment was aimed at estimating the capability of the cameras to reconstruct human bodies. This is especially useful for healthcare applications and for human activity monitoring and safety in industrial workspaces. The target object was a mannequin representing the human upper body with movable arms to simulate a variety of human poses. The reconstruction results of each camera were compared with Kinect v2 and with an industrial digitizer Konica Minolta (gold standard). In conclusion, the performance of Kinect Azure and Basler Blaze 101 are usually comparable with the reference except for some tricky poses where occlusions interfere with the measurement. The difference between the two is probably due to the typical application for which the cameras are developed. Industrial cameras perform best with smaller objects and have higher point cloud density to better reconstruct surface defects in controlled environments, while consumer-end cameras are typically used in unstructured environments with a variety of ambient conditions to reconstruct bigger objects and bodies.

As a further development, we aim to expand the characterization procedure by adding tests on different surfaces in terms of both materials' opacity and color. Moreover, the 3D reconstruction experimental procedure proved to be necessary as well to define which camera is best suited for the task; hence, we aim to rigorously standardize it as well in the future by also taking into consideration objects of different shapes and materials, a full-body mannequin, and human subjects. In this way, researchers and practitioners may conduct a thorough metrological investigation of their sensor of choice.

Author Contributions: Conceptualization, S.P.; Formal analysis, S.P.; Investigation, C.N. and M.Z.; Methodology, S.P.; Software, C.N. and M.Z.; Supervision, M.L. and M.D.C.; Validation, C.N., A.L. and M.L.; Writing—original draft, S.P. and C.N.; Writing—review and editing, A.L., M.L. and M.D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union's Horizon 2020 research and innovation programme, via an Open Call issued and executed under Project EUROBENCH (grant agreement No. 779963).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors wish to thank Gabriele Coffetti and Cristian Fracassi for their help and support during this research work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Giancola, S.; Valenti, M.; Sala, R. State-of-the-Art Devices Comparison. In *A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies*; SpringerBriefs in Computer Science; Springer: Cham, Switzerland, 2018; pp. 29–39.
2. Horaud, R.; Hansard, M.; Evangelidis, G.; Ménier, C. An overview of depth cameras and range scanners based on time-of-flight technologies. *Mach. Vis. Appl.* **2016**, *27*, 1005–1020. [[CrossRef](#)]
3. Rapp, H.; Frank, M.; Hamprecht, F.A.; Jähne, B. A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time-of-flight-cameras. *Int. J. Intell. Syst. Technol. Appl.* **2008**, *5*, 402–413. [[CrossRef](#)]
4. Frank, M.; Plaue, M.; Rapp, H.; Koethe, U.; Jähne, B.; Hamprecht, F.A. Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. *Opt. Eng.* **2009**, *48*, 013602.
5. He, Y.; Liang, B.; Zou, Y.; He, J.; Yang, J. Depth Errors Analysis and Correction for Time-of-Flight (ToF) Cameras. *Sensors* **2017**, *17*, 92. [[CrossRef](#)] [[PubMed](#)]
6. Sarbolandi, H.; Lefloch, D.; Kolb, A. Kinect range sensing: Structured-light versus Time-of-Flight Kinect. *Comput. Vis. Image Underst.* **2015**, *139*, 1–20. [[CrossRef](#)]
7. Wasenmüller, O.; Stricker, D. Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. In *Computer Vision—ACCV 2016 Workshops*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2017; Volume 10117, pp. 34–45.
8. Corti, A.; Giancola, S.; Mainetti, G.; Sala, R. A metrological characterization of the Kinect V2 time-of-flight camera. *Robot. Auton. Syst.* **2016**, *75*, 584–594. [[CrossRef](#)]
9. He, Y.; Chen, S. Recent Advances in 3D Data Acquisition and Processing by Time-of-Flight Camera. *IEEE Access* **2019**, *7*, 12495–12510. [[CrossRef](#)]
10. Chen, S.; Yi, J.; Ding, H.; Wang, Z.; Min, J.; Wu, H.; Cao, S.; Mu, J. 3D Object Reconstruction with Kinect Based on QR Code Calibration. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 27–29 June 2020; pp. 459–463.
11. He, H.; Wang, H.; Sun, L. Research on 3D point-cloud registration technology based on Kinect V2 sensor. In Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 1264–1268.
12. Shen, B.; Yin, F.; Chou, W. A 3D Modeling Method of Indoor Objects Using Kinect Sensor. In Proceedings of the 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 9–10 December 2017; Volume 1, pp. 64–68.
13. Zhao, Y.; Carraro, M.; Munaro, M.; Menegatti, E. Robust multiple object tracking in RGB-D camera networks. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 6625–6632.
14. Jiao, J.; Yuan, L.; Tang, W.; Deng, Z.; Wu, Q. A Post-Rectification Approach of Depth Images of Kinect v2 for 3D Reconstruction of Indoor Scenes. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 349. [[CrossRef](#)]
15. Chen, Y.; Zhang, B.; Zhou, J.; Wang, K. Real-time 3D unstructured environment reconstruction utilizing VR and Kinect-based immersive teleoperation for agricultural field robots. *Comput. Electron. Agric.* **2020**, *175*, 105579. [[CrossRef](#)]
16. Oliver, A.; Kang, S.; Wünsche, B.C.; MacDonald, B. Using the Kinect as a Navigation Sensor for Mobile Robotics. In Proceedings of the 27th Conference on Image and Vision Computing New Zealand, Dunedin, New Zealand, 26–28 November 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 509–514.
17. Popov, V.L.; Ahmed, S.A.; Shakev, N.G.; Topalov, A.V. Detection and Following of Moving Targets by an Indoor Mobile Robot using Microsoft Kinect and 2D Lidar Data. In Proceedings of the 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 18–21 November 2018; pp. 280–285.
18. Lai, C.C.; Su, K.L. Development of an intelligent mobile robot localization system using Kinect RGB-D mapping and neural network. *Comput. Electr. Eng.* **2018**, *67*, 620–628. [[CrossRef](#)]
19. Henry, P.; Krainin, M.; Herbst, E.; Ren, X.; Fox, D. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *Int. J. Robot. Res.* **2012**, *31*, 647–663. [[CrossRef](#)]
20. Caruso, L.; Russo, R.; Savino, S. Microsoft Kinect V2 vision system in a manufacturing application. *Robot. Comput.-Integr. Manuf.* **2017**, *48*, 174–181. [[CrossRef](#)]
21. Rodriguez-Garavito, C.H.; Camacho-Munoz, G.; Álvarez-Martínez, D.; Cardenas, K.V.; Rojas, D.M. 3D Object Pose Estimation for Robotic Packing Applications. In *Applied Computer Sciences in Engineering. WEA 2018. Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2018; Volume 916, pp. 453–463.
22. Nascimento, H.; Mujica, M.; Benoussaad, M. Collision Avoidance Interaction Between Human and a Hidden Robot Based on Kinect and Robot Data Fusion. *IEEE Robot. Autom. Lett.* **2021**, *6*, 88–94. [[CrossRef](#)]

23. Pasinetti, S.; Nuzzi, C.; Lancini, M.; Fornaser, A.; Docchio, F.; Sansoni, G. Development and characterization of a safety system for robotic cells based on multiple Time of Flight (TOF) cameras and point cloud analysis. In Proceedings of the 2018 IEEE International Workshop on Metrology for Industry 4.0 and IoT, Brescia, Italy, 16–18 April 2018.
24. Halme, R.-J.; Minna, L.; Kämäräinen, J.; Pieters, R.; Latokartano, J.; Hietanen, A. Review of vision-based safety systems for human-robot collaboration. *Procedia CIRP* **2018**, *72*, 111–116. [[CrossRef](#)]
25. Palmieri, P.; Melchiorre, M.; Scimmi, L.S.; Pastorelli, S.; Mauro, S. Human Arm Motion Tracking by Kinect Sensor Using Kalman Filter for Collaborative Robotics. In *Advances in Italian Mechanism Science; IFToMM ITALY 2020; Mechanisms and Machine Science*; Springer: Cham, Switzerland, 2021; Volume 91, pp. 326–334.
26. Nuzzi, C.; Ghidini, S.; Pagani, R.; Pasinetti, S.; Coffetti, G.; Sansoni, G. Hands-Free: A robot augmented reality teleoperation system. In Proceedings of the 2020 17th International Conference on Ubiquitous Robots (UR), Kyoto, Japan, 22–26 June 2020.
27. Sankar, S.; Tsai, C.-Y. ROS-Based Human Detection and Tracking from a Wireless Controlled Mobile Robot Using Kinect. *Appl. Syst. Innov.* **2019**, *2*, 5. [[CrossRef](#)]
28. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.-E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [[CrossRef](#)]
29. Carraro, M.; Munaro, M.; Burke, J.; Menegatti, E. Real-time marker-less multi-person 3d pose estimation in rgb-depth camera networks. *arXiv* **2017**, arXiv:1710.06235.
30. Nuzzi, C.; Pasinetti, S.; Pagani, R.; Ghidini, S.; Beschi, M.; Coffetti, G.; Sansoni, G. MEGURU: A gesture-based robot program builder for Meta-Collaborative workstations. *Robot. Comput.-Integr. Manuf.* **2021**, *68*, 102085. [[CrossRef](#)]
31. Torres, S.H.M.; Kern, M.J. 7 DOF industrial robot controlled by hand gestures using microsoft kinect v2. In Proceedings of the 2017 IEEE 3rd Colombian Conference on Automatic Control (CCAC), Cartagena, Colombia, 18–20 October 2017; pp. 1–6.
32. Ganguly, B.; Vishwakarma, P.; Biswas, S.; Rahul. Kinect Sensor Based Single Person Hand Gesture Recognition for Man-Machine Interaction. In *Computational Advancement in Communication Circuits and Systems; Lecture Notes in Electrical Engineering*; Springer: Singapore, 2020; Volume 575, pp. 139–144.
33. Roy, G.; Bhuiya, A.; Mukherjee, A.; Bhaumik, S. Kinect Camera Based Gait Data Recording and Analysis for Assistive Robotics-An Alternative to Goniometer Based Measurement Technique. *Procedia Comput. Sci.* **2018**, *133*, 763–771. [[CrossRef](#)]
34. Pasinetti, S.; Fornaser, A.; Lancini, M.; De Cecco, M.; Sansoni, G. Assisted Gait Phase Estimation Through an Embedded Depth Camera Using Modified Random Forest Algorithm Classification. *IEEE Sens. J.* **2020**, *20*, 3343–3355. [[CrossRef](#)]
35. Pasinetti, S.; Nuzzi, C.; Covre, N.; Luchetti, A.; Maule, L.; Serpelloni, M.; Lancini, M. Validation of Marker-Less System for the Assessment of Upper Joints Reaction Forces in Exoskeleton Users. *Sensors* **2020**, *20*, 3899. [[CrossRef](#)] [[PubMed](#)]
36. Mettel, M.R.; Alekseev, M.; Stocklów, C.; Braun, A. Safety Services in Smart Environments Using Depth Cameras. In *Ambient Intelligence: Aml 2017; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2017; Volume 10217, pp. 80–93.
37. Butaslac, I.I.; Luchetti, A.; Parolin, E.; Fujimoto, Y.; Kanbara, M.; De Cecco, M.; Kato, H. The Feasibility of Augmented Reality as a Support Tool for Motor Rehabilitation. *Int. Conf. Augment. Real. Virtual Real. Comput. Graph.* **2020**, 12243, 165–173.
38. Luchetti, A.; Parolin, E.; Butaslac, I.; Fujimoto, Y.; Kanbara, M.; Bosetti, P.; De Cecco, M.; Kato, H. Stepping over Obstacles with Augmented Reality based on Visual Exproprioception. In Proceedings of the 2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Recife, Brazil, 9–13 November 2020.
39. Albert, J.A.; Owolabi, V.; Gebel, A.; Brahms, C.M.; Granacher, U.; Arnrich, B. Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study. *Sensors* **2020**, *20*, 5104. [[CrossRef](#)] [[PubMed](#)]
40. Tölgyessy, M.; Dekan, M.; Chovanec, L.; Hubinský, P. Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2. *Sensors* **2021**, *21*, 413. [[CrossRef](#)] [[PubMed](#)]
41. Choo, B.; Landau, M.; DeVore, M.; Beling, P.A. Statistical Analysis-Based Error Models for the Microsoft Kinect™ Depth Sensor. *Sensors* **2014**, *14*, 17430–17450. [[CrossRef](#)]
42. Kurillo, G.; Hemingway, E.; Cheng, M.-L.; Cheng, L. Evaluating the Accuracy of the Azure Kinect and Kinect v2. *Sensors* **2022**, *22*, 2469. [[CrossRef](#)]
43. Pasinetti, S.; Hassan, M.M.; Eberhardt, J.; Lancini, M.; Docchio, F.; Sansoni, G. Performance Analysis of the PMD Camboard Picoflexx Time-of-Flight Camera for Markerless Motion Capture Applications. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 4456–4471. [[CrossRef](#)]
44. Crenna, F.; Rossi, G.; Palazzo, A. Measurement of human movement under metrological controlled conditions. *Acta Imeko* **2015**, *4*, 48–56. [[CrossRef](#)]
45. Hussmann, S.; Knoll, F.; Edeler, T. Modulation Method Including Noise Model for Minimizing the Wiggling Error of TOF Cameras. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 1127–1136. [[CrossRef](#)]
46. Bamji, C.S.; Mehta, S.; Thompson, B.; Elkhatib, T.; Wurster, S.; Akkaya, O.; Payne, A.; Godbaz, J.; Fenton, M.; Rajasekaran, V.; et al. IMpixel 65nm BSI 320MHz demodulated TOF Image sensor with 3µm global shutter pixels and analog binning. In Proceedings of the 2018 IEEE International Solid—State Circuits Conference—(ISSCC), San Francisco, CA, USA, 11–15 February 2018.
47. Foix, S.; Alenya, G.; Torras, C. Lock-in time-of-flight (ToF) cameras: A survey. *IEEE Sens. J.* **2011**, *11*, 1917–1926. [[CrossRef](#)]
48. Fürsattel, P.; Placht, S.; Balda, M.; Schaller, C.; Hofmann, H.; Maier, A.; Riess, C. A Comparative Error Analysis of Current Time-of-Flight Sensors. *IEEE Trans. Comput. Imaging* **2016**, *2*, 27–41. [[CrossRef](#)]

49. Joint Committee for Guides in Metrology (JCGM). Guide to the Expression of Uncertainty in Measurement (GUM). 2008. Available online: <https://www.bipm.org> (accessed on 6 October 2022).
50. Kahlmann, T.; Remondino, F.; Ingensand, H. Calibration for increased accuracy of the range imaging camera swissranger. In Proceedings of the ISPRS Commission V Symposium 'Image Engineering and Vision Metrology', Dresden, Germany, 25–27 September 2006; Volume 36, pp. 136–141.
51. Chiarabando, F.; Chiarabando, R.; Piatti, D.; Rinaudo, F. Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera. *Sensors* **2009**, *9*, 10080–10096. [[CrossRef](#)]
52. Steiger, O.; Felder, J.; Weiss, S. Calibration of time-of-flight range imaging cameras. In Proceedings of the 2008 15th IEEE International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1968–1971.
53. Imiya, A.; Kawamoto, K. Learning dimensionality and orientations of 3D objects. *Pattern Recognit. Lett.* **2001**, *22*, 75–83. [[CrossRef](#)]
54. Kiyomitsu, K.; Kakinuma, A.; Takahashi, H.; Kamijo, N.; Ogawa, K.; Tsumura, N. Volume measurement of the leg with the depth camera for quantitative evaluation of edema. In *Advanced Biomedical and Clinical Diagnostic and Surgical Guidance Systems XV*; SPIE BiOS: Bellingham, WA, USA, 2017.
55. Fornaser, A.; Tomasin, P.; De Cecco, M.; Tavernini, M.; Zanetti, M. Automatic graph based spatiotemporal extrinsic calibration of multiple Kinect V2 ToF cameras. *Robot. Auton. Syst.* **2017**, *98*, 105–125. [[CrossRef](#)]
56. Martinek, M.; Grosso, R.; Greiner, G. Optimized Canonical Coordinate Frames for 3D Object Normalization. In Proceedings of the Vision, Modeling and Visualization (VMV), Magdeburg, Germany, 12–14 November 2012.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.