*Technical Note*

# Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2

**Michal Tölgyessy** \*![ORCID]**, Martin Dekan** ![ORCID]**, Ľuboš Chovanec** ![ORCID] **and Peter Hubinský**

Institute of Robotics and Cybernetics, Faculty of Electrical Engineering and Information Technology STU in Bratislava, Ilkovičova 3, 812 19 Bratislava, Slovakia; martin.dekan@stuba.sk (M.D.); lubos.chovanec@stuba.sk (Ľ.C.); peter.hubinsky@stuba.sk (P.H.)
\* Correspondence: michal.tolgyessy@stuba.sk

**Abstract:** The Azure Kinect is the successor of Kinect v1 and Kinect v2. In this paper we perform brief data analysis and comparison of all Kinect versions with focus on precision (repeatability) and various aspects of noise of these three sensors. Then we thoroughly evaluate the new Azure Kinect; namely its warm-up time, precision (and sources of its variability), accuracy (thoroughly, using a robotic arm), reflectivity (using 18 different materials), and the multipath and flying pixel phenomenon. Furthermore, we validate its performance in both indoor and outdoor environments, including direct and indirect sun conditions. We conclude with a discussion on its improvements in the context of the evolution of the Kinect sensor. It was shown that it is crucial to choose well designed experiments to measure accuracy, since the RGB and depth camera are not aligned. Our measurements confirm the officially stated values, namely standard deviation $\leq 17$ mm, and distance error $< 11$ mm in up to 3.5 m distance from the sensor in all four supported modes. The device, however, has to be warmed up for at least 40–50 min to give stable results. Due to the time-of-flight technology, the Azure Kinect cannot be reliably used in direct sunlight. Therefore, it is convenient mostly for indoor applications.

**Keywords:** Kinect; Azure Kinect; robotics; mapping; SLAM (simultaneous localization and mapping); HRI (human–robot interaction); 3D scanning; depth imaging; object recognition; gesture recognition

## 1. Introduction

The Kinect Xbox 360 has been a revolution in affordable 3D sensing. Initially meant only for the gaming industry, it was soon to be used by scientists, robotics enthusiasts and hobbyists all around the world. It was later followed by the release of another Kinect—Kinect for Windows. We will refer to the former as Kinect v1, and to the latter as Kinect v2. Both versions have been widely used by the research community in various scientific such as object detection and object recognition [1–3], mapping and SLAM [4–6], gesture recognition and human–machine interaction (HMI) [7–9], telepresence [10,11], virtual reality, mixed reality, and medicine and rehabilitation [12–16]. According to [17] there have been hundreds of papers written and published on this subject. However, both sensors are now discontinued and are no longer being officially distributed and sold. In 2019 Microsoft released the Azure Kinect, which is no longer meant for the gaming market in any way; it is promoted as a developer kit with advanced AI sensors for building computer vision and speech models. Therefore, we focus on the analysis and evaluation of this sensor and the depth image data it produces.

Our paper is organized as follows. Firstly, we describe the relevant features of each of the three sensors—Kinect v1, Kinect v2 and Azure Kinect (Figure 1). Then we briefly compare the output of all Kinects. Our focus is not set on complex evaluation of previous Kinect versions as this has been done before (for reference see [18–24]). In the last section, we focus primarily on the Azure Kinect and thoroughly evaluate its performance, namely:

- Warm-up time (the effect of device temperature on its precision)

- Accuracy
- Precision
- Color and material effect on sensor performance
- Precision variability analysis
- Performance in outdoor environment



**Figure 1.** From left to right—Kinect v1, Kinect v2, Azure Kinect.

## 2. Kinects' Specifications

Both earlier versions of the Kinect have one depth camera and one color camera. The Kinect v1 measures depth with the pattern projection principle, where a known infrared pattern is projected onto the scene and out of its distortion the depth is computed. The Kinect v2 utilizes the continuous wave (CW) intensity modulation approach, which is most commonly used in time-of-flight (ToF) cameras [18].

In a continuous-wave (CW) time-of-flight (ToF) camera, light from an amplitude modulated light source is backscattered by objects in the camera's field of view, and the phase delay of the amplitude envelope is measured between the emitted and reflected light. This phase difference is translated into a distance value for each pixel in the imaging array [25].

The Azure Kinect is also based on a CW ToF camera; it uses the image sensor presented in [25]. Unlike Kinect v1 and v2, it supports multiple depth sensing modes and the color camera supports a resolution up to 3840 × 2160 pixels.

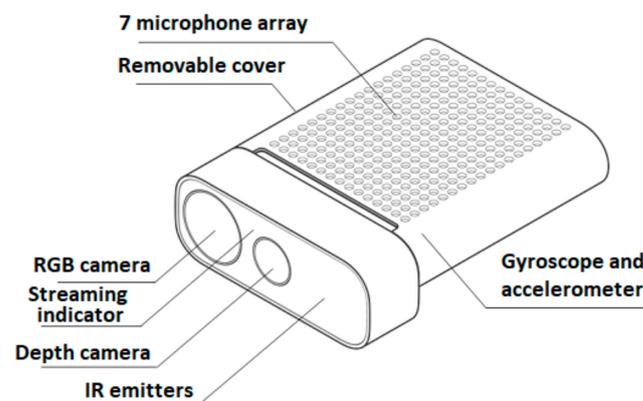The design of the Azure Kinect is shown in Figure 2.



**Figure 2.** Schematic of the Azure Kinect.

Comparison of the key features of all three Kinects is in Table 1. All data regarding Azure Kinect is taken from the official online documentation.

**Table 1.** Comparison of the three Kinect versions.

|  | Kinect v1 [17] | Kinect v2 [26] | Azure Kinect |
|---|---|---|---|
| Color camera resolution | 1280 × 720 px @ 12 fps 640 × 480 px @ 30 fps | 1920 × 1080 px @ 30 fps | 3840 × 2160 px @30 fps |
| Depth camera resolution | 320 × 240 px @ 30 fps | 512 × 424 px @ 30 fps | NFOV unbinned—640 × 576 @ 30 fps NFOV binned—320 × 288 @ 30 fps WFOV unbinned—1024 × 1024 @ 15 fps WFOV binned—512 × 512 @ 30 fps |
| Depth sensing technology | Structured light–pattern projection | ToF (Time-of-Flight) | ToF (Time-of-Flight) |
| Field of view (depth image) | 57° H, 43° V alt. 58.5° H, 46.6° | 70° H, 60° V alt. 70.6° H, 60° | NFOV unbinned—75° × 65° NFOV binned—75° × 65° WFOV unbinned—120° × 120° WFOV binned—120° × 120° |
| Specified measuring distance | 0.4–4 m | 0.5–4.5 m | NFOV unbinned—0.5–3.86 m NFOV binned—0.5–5.46 m WFOV unbinned—0.25–2.21 m WFOV binned—0.25–2.88 m |
| Weight | 430 g (without cables and power supply); 750 g (with cables and power supply) | 610 g (without cables and power supply); 1390 g (with cables and power supply) | 440 g (without cables); 520 g (with cables, power supply is not necessary) |

It works in four different modes-NFOV (narrow field-of-view depth mode) unbinned, WFOV (wide field-of-view depth mode) unbinned, NFOV binned, and WFOV binned. The Azure Kinect has both, a depth camera and an RGB camera; spatial orientation of the RGB image frame and depth image frame is not identical, there is a 1.3-degree difference. The SDK contains convenience functions for the transformation. These two parts are, according to the SDK, time synchronized by the Azure.

## 3. Comparison of all Kinect Versions

In this set of experiments, we focused primarily on the precision of the examined sensors. They were placed on a construction facing a white wall as shown in Figure 3. We measured depth data in three locations (80, 150 and 300 cm), and switched the sensors at the top of the construction so that only one sensor faced the wall during measurement in order to eliminate interferential noise. The three locations were chosen in order to safely capture valid data for approximate min, mid and max range for each of the sensors. A standard measuring tape was used, since in this experiment we measured only precision (repeatability). A total of 1100 frames were captured for every position.



**Figure 3.** Sensor placement for testing purposes.

### 3.1. Typical Sensor Data

Typical depth data of tested sensors is presented in Figure 4.
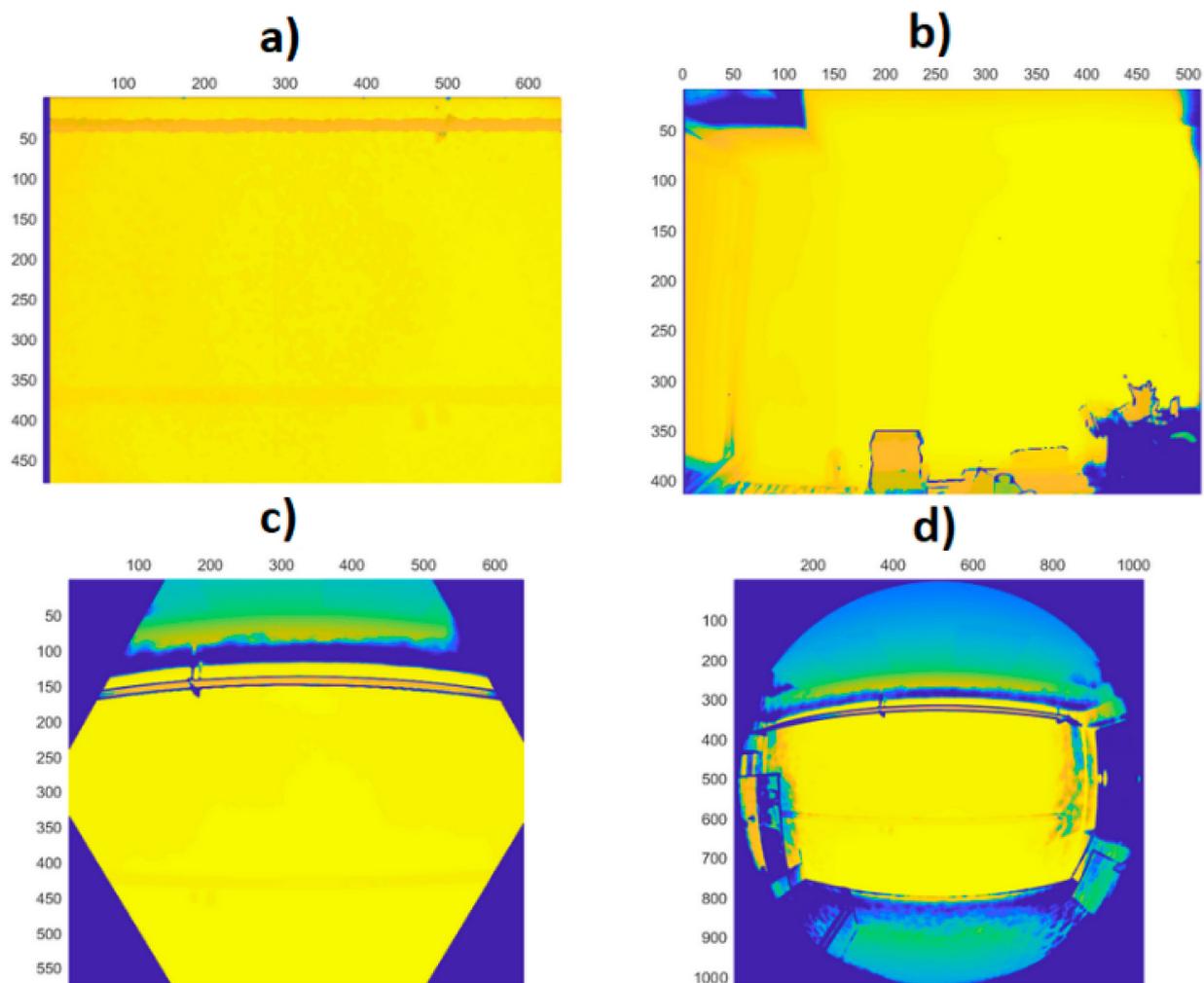
**Figure 4.** Typical data measurements acquired from Kinect v1 (**a**), Kinect v2 (**b**), Azure Kinect in narrow field-of-view (NFOV) binned mode (**c**), and Azure Kinect in wide field-of-view (WFOV) (**d**) sensors (axes represent image pixel positions).

As can be seen, there is a big difference between the first two generations of the Kinect and the new Azure version. The first two cover the whole rectangular area of the pixel matrix with valid data, while the new sensor has hex area for the narrow mode (NFOV), and circular area for the wide mode (WFOV). The data from the Azure Kinect still comes as a matrix, so there are many pixels that are guaranteed to be without information for every measurement. Furthermore, the field of view of the Azure Kinect is wider.

In our experiments we focused on noise-to-distance correlation, and noise relative to object reflectivity.

### 3.2. Experiment No. 1–Noise

In this experiment we focused on the repeatability of the measurement. Typical noise of all three sensors at 800 mm distance can be seen in Figures 5–8. The depicted visualization represents standard deviation for each pixel position computed from repeated measurements in the same sensor position (calculated from distance in mm). For better visual clarity, we limited the maximal standard deviation for each sensor/distance to the value that was located in the area of correct measurements. Extreme values of standard deviation we omitted were caused by one of the following-end of sensor range or border between two objects.
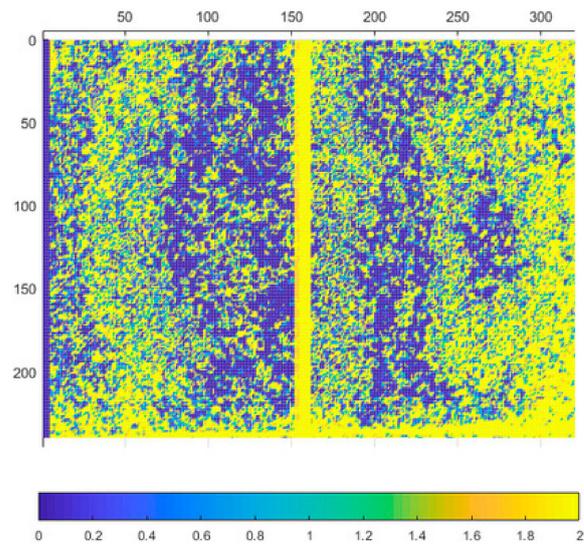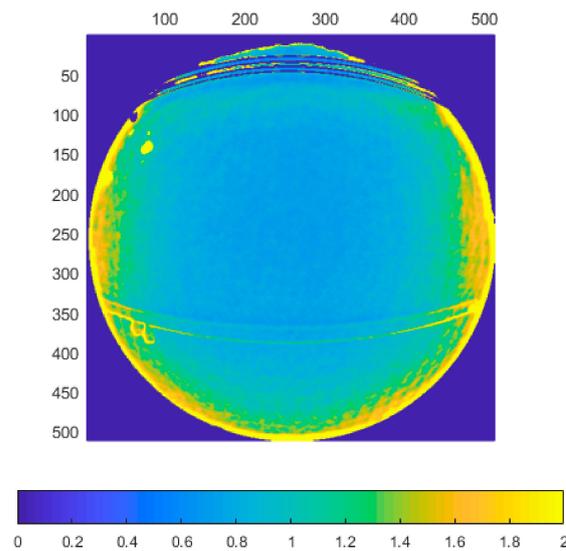
**Figure 5.** Typical depth noise of Kinect v1 in mm (values over 2 mm were limited to 2 mm for better visual clarity). Picture axes represent pixel positions.
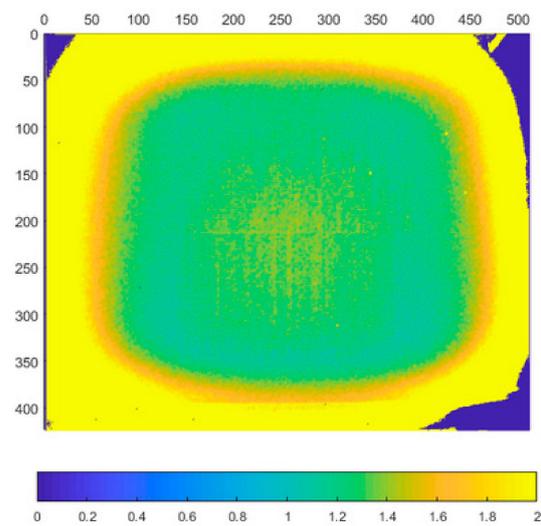


**Figure 6.** Typical depth noise of Kinect v2 in mm (values over 2 mm were limited to 2 mm for better visual clarity). Picture axes represent pixel positions.
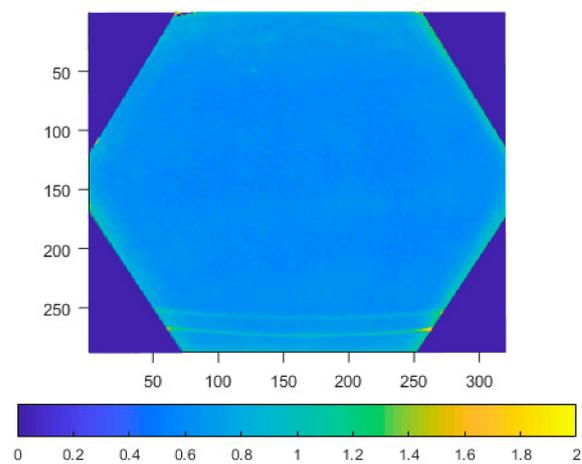


**Figure 7.** Typical depth noise of Azure Kinect in NFOV binned mode in mm (values over 2 mm were limited to 2 mm for better visual clarity). Picture axes represent pixel positions.

**Figure 8.** Typical depth noise of Azure Kinect in WFOV binned mode in mm (values over 2 mm were limited to 2 mm for better visual clarity). Picture axes represent pixel positions.

As can be seen in presented Figures 5–8, the noise of Kinect v2 and Azure Kinect raises at the edges of useful data while the Kinect v1 noise has many areas with high noise. This was expectable as both, Kinect v2 and Azure, work on the same measuring principle.

For the noise to distance correlation, we made measurements at different distances from a white wall. We took an area of 7 × 7 pixels around the center of the captured depth image as this is the area where the measured distance corresponds to the smallest actual Euclidian distance between the sensor chip and the wall; as the wall is perpendicular to the sensor (Figure 9). From these data we calculated the standard deviation for all 49 pixels, and then computed its mean value. From Table 2, it is obvious that Microsoft made progress with their new sensor and the repeatability of the new sensor is much better than the first Kinect generation, and even surpasses the Kinect v2 in 3 out of 4 modes.
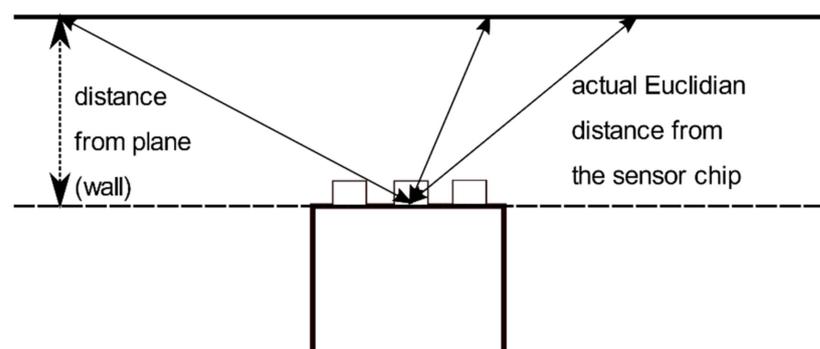


**Figure 9.** Plane vs. Euclidian distance of a 3D point form the sensor chip.

**Table 2.** Standard deviation of all sensors at different distances (mean value in mm).

| | Kinect v1 (320 × 240 px) | Kinect v1 (640 × 480 px) | Kinect v2 | Azure Kinect NFOV Binned | Azure Kinect NFOV Unbinned | Azure Kinect WFOV Binned | Azure Kinect WFOV Unbinned |
|---|---|---|---|---|---|---|---|
| 800 mm | 1.0907 | 1.6580 | 1.1426 | 0.5019 | 0.6132 | 0.5546 | 0.8465 |
| 1500 mm | 3.1280 | 3.6496 | 1.4016 | 0.5800 | 0.8873 | 0.8731 | 1.5388 |
| 3000 mm | 10.9928 | 13.6535 | 2.6918 | 0.9776 | 1.7824 | 2.1604 | 8.1433 |

### 3.3. Experiment No. 2–Noise

The second experiment focused on the noise-to-reflectivity correspondence. We put a white-board and a cork-board table in the sensing area (Figure 10), and then we checked whether it can be seen in the noise data. All sensors were placed in the same position and distance from the wall, but the presented depth image sizes slightly differ, since each sensor has different resolution.
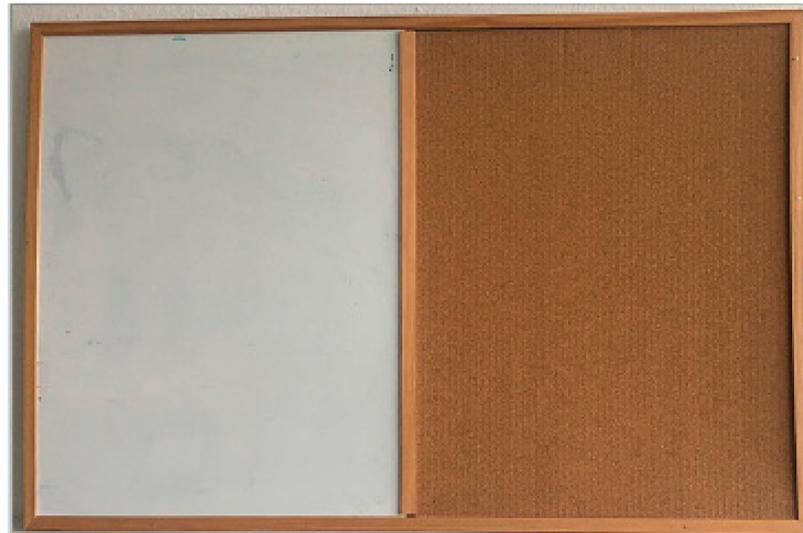


**Figure 10.** Test plate composed of plastic reflective material and cork.

What we expected was, that for Kinect v1, the board should be nonvisible and for Kinect v2, a change in the data should be detectable. The noise data for Kinect v1 and v2 is in Figures 11 and 12.
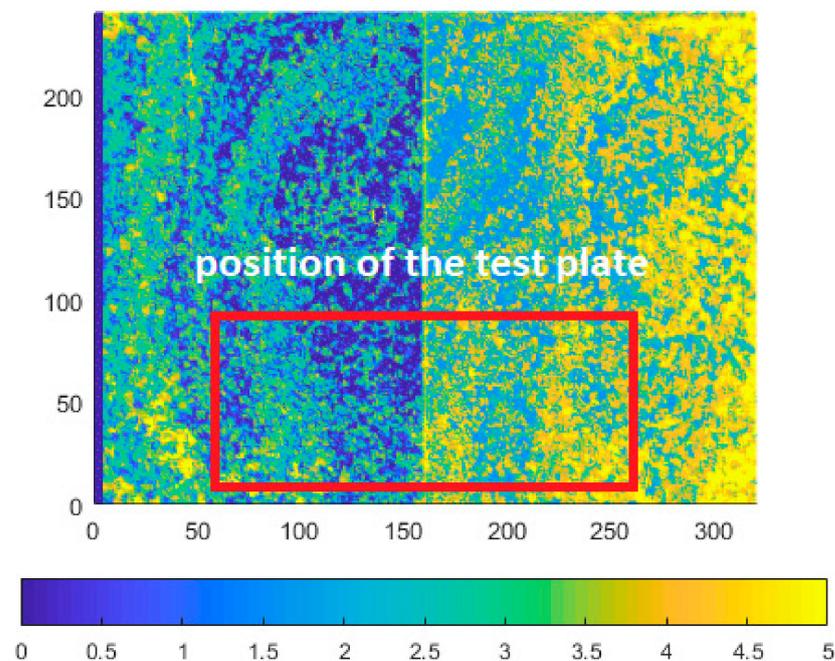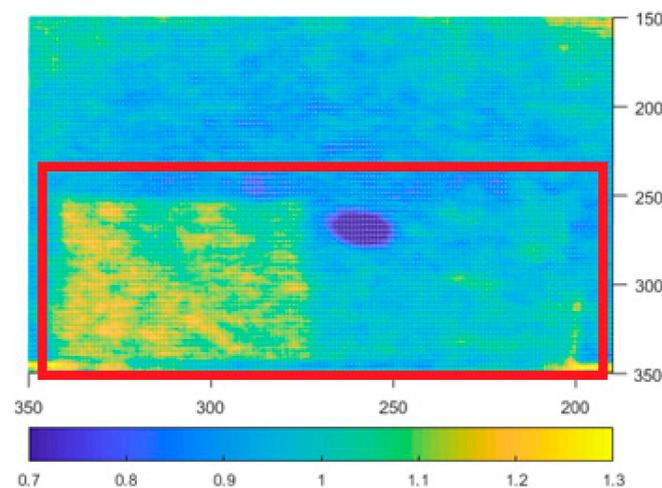


**Figure 11.** Depth noise of Kinect v1 with presence of an object with different reflectivity (values over 5 mm were limited to 5 mm for better visual clarity). Picture axes represent pixel positions.
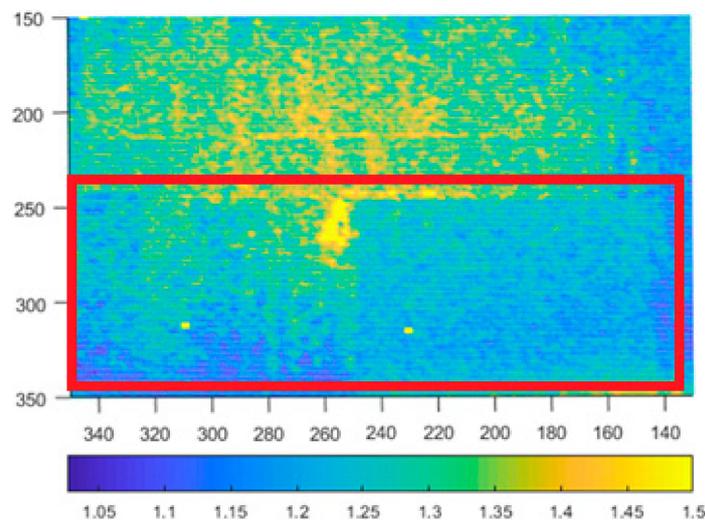
**Figure 12.** Depth noise of Kinect v2 with presence of an object with different reflectivity in the bottom area (values over 1.5 mm were limited to 1.5 mm for better visual clarity). Picture axes represent pixel positions.

As can be seen in Figure 12, there is an outline of the board visible in the Kinect v2 data (for better visual clarity the range of the noise was limited, and only the relevant area of the depth image is shown).

We preformed the same experiment with the Azure Kinect (Figure 13). The board is visible in all 4 modes; therefore, we assume the behavior of Azure in this regard is the same as for Kinect v2. For more information regarding this behavior for Kinect v2 please refer to [26], which deals with this issue in more detail.



**Figure 13.** Depth noise of Kinect Azure with presence of an object with different reflectivity in the bottom area (values over 1.3 mm were limited to 1.3 mm for better visual clarity). Picture axes represent pixel positions.

## 4. Evaluation of the Azure Kinect

The Azure Kinect provides multiple settings for frame rate. It can be set to 5, 15 or 30 hz, with the exception of both the unbinned wide depth mode, and RGB 3072px mode, where the 30 fps is not supported. It has a hardware trigger, with fairly stable frame-rate. We recorded 4500 frames with timestamps and found only small fluctuation ($\pm 2$ ms between frames), which was probably caused by the host PC, not the Azure Kinect itself.

### 4.1. Warm-up Time

The Azure Kinect does not give a stable output value after it is turned on. It needs to be warmed up for some time for the output to stabilize. Therefore, we devised an experiment to determine this time. For this experiment, we put the Azure Kinect roughly 90 cm away from a white wall and started measuring right away with a cold Azure Kinect. We ran the measurement for 80 min and computed the average distance and standard deviation from first 15 s of each minute. From these data we took one center pixel; the results are in Figures 14 and 15. As can be seen, the standard deviation did not change considerably, but the measured distance grew until it stabilized on a value 2 mm higher compared to the starting value. From the results we conclude it is necessary to run the Azure Kinect for at least 60 min to get stabilized output.
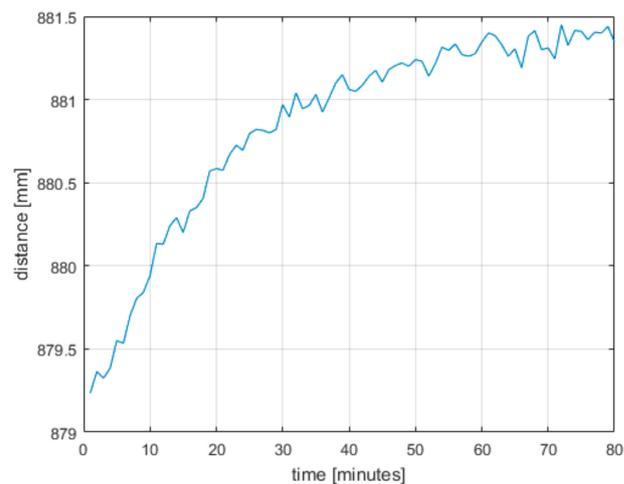


**Figure 14.** Measured distance while warming up the Azure Kinect. Each point represents the average distance for that particular minute.
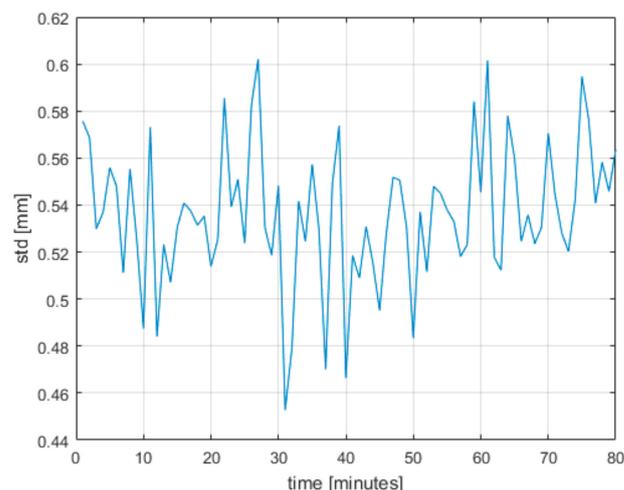


**Figure 15.** Measured standard deviation while warming up the Azure Kinect.

Every other experiment was performed with a warmed-up device.

### 4.2. Accuracy

For accuracy measurements, we mounted a white reflective plate to the end effector of a robotic manipulator—ABB IRB4600 (Figures 16 and 17). The goal was to position the plate in precise locations and measure the distance to this plate. The absolute positioning accuracy of the robot end effector is within 0.02 mm according to the ABB IRB4600 datasheet.
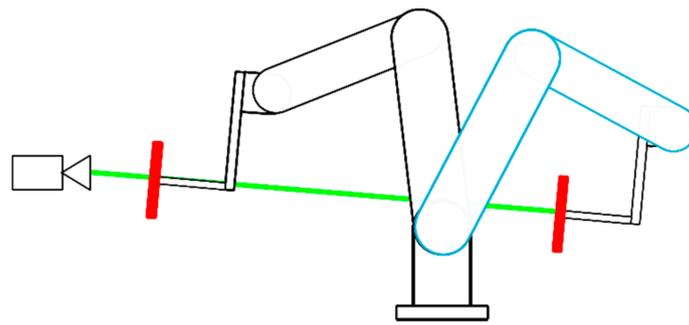
**Figure 16.** Scheme for accuracy measurements using robotic manipulator.



**Figure 17.** Picture of the actual laboratory experiment.

The Azure Kinect depth frame is rotated a few degrees from the RGB frame, thus we had to properly align our measuring plate with this frame. Otherwise, the range error would be distorted. First step in this process was to align the plate with the depth frame; for a coarse alignment we used an external IMU sensor (DXL360s with resolution 0.01 deg). Then, for fine tuning, we changed the orientation of the plate, so that 25 selected depth points reported roughly the same average distance. The average for every point did not deviate more than 0.1 mm (Figure 18).
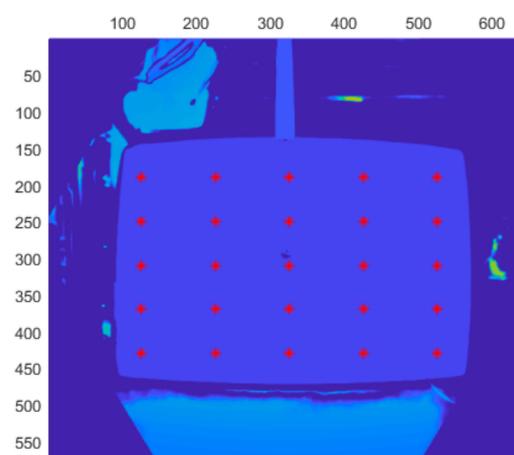


**Figure 18.** Selected depth points for fine tuning of the plate alignment. Picture axes represent pixel positions.

After that, we had to determine the axis along which the robot should move, so that it would not deviate from the depth frame axis (Figure 16). For this purpose, we used both the depth image and the infrared image. At the center of the plane, a bolt head was located. This bolt was clearly visible in depth images at short distances and in infrared images at longer distances. We positioned the robot in such a way, that this bolt was located in the center of the respective image at two different distances. One distance was set to 42 cm and the other to over 2.7 m. These two points defined the line (axis) along which we moved the plate. We assured the bolt was located in the center for multiple measuring positions. Therefore, the axis of robot movement did not deviate from the Z axis of the depth camera for more than 0.09 degrees, which should correspond to sub mm error in reported distance of the Azure Kinect.

Unfortunately, the exact position of the origin of the depth frame is not known. For this reason, we assumed one distance as correct, and all the errors are reported relatively to this position. The originating distance was set to 50 cm, as reported by the unbinned narrow mode of the Azure Kinect (this assumption may be wrong, as it is possible, that the sensor does not report correct distance in any mode or any location). What the reader should take from this experiment is how the accuracy changes with varying distances.

We performed measurements for all 4 modes for distances ranging from 500 mm to 3250 mm with a 250 mm step, and at 3400 mm, which was the last distance our robot could reach. From each measurement, we took 25 points (selected manually similarly to Figure 18) and computed the average distance. The exact position of the points varied for each mode and plate location, but we tried to make them evenly distributed throughout the measurement plate. For the starting 500 mm distance and unbinned narrow mode, the points are the same as shown in Figure 18.

As can be seen in Figure 19, at this range, the distance error does not deviate much, neither with changing modes nor with distances. There is a drop of 8 mm for binned narrow mode, but this is well within datasheet values. If we selected a different distance as the correct one, the error would be even less visible.
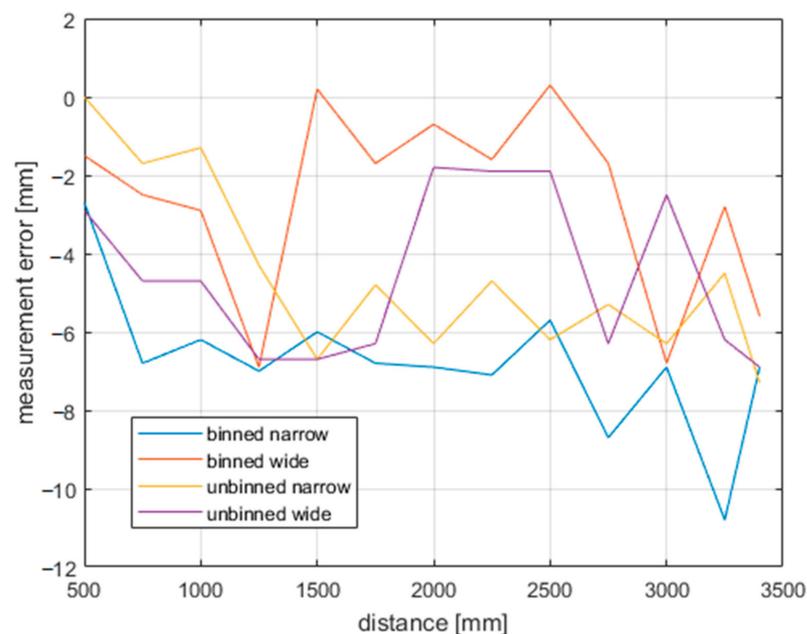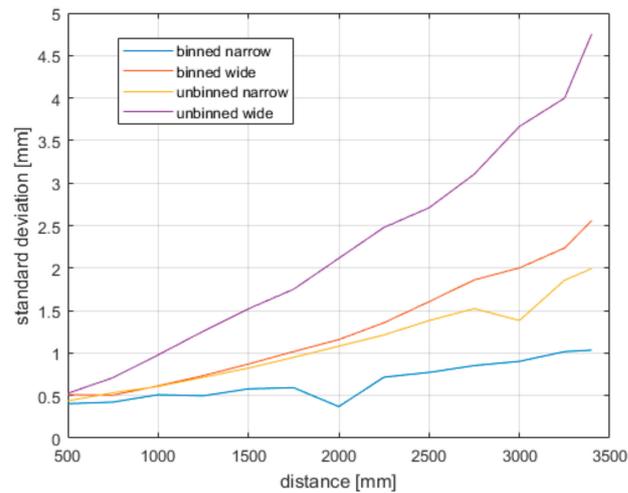


**Figure 19.** Accuracy of the Azure Kinect for all modes.

### 4.3. Precision

To determine the precision (repeatability) we used the same data as for the accuracy measurements; the resulting precision is shown in Figure 20. As expected, the binned versions of both fields of view show much better results.

**Figure 20.** Precision of the Azure Kinect for all modes.

## 4.4. Reflectivity

For this test, our aim was to compare the precision of the Azure Kinect for multiple types of material specimens; therefore, they had different reflectivity properties. The specimens and their layouts are shown in Figure 21; the basic reflectivity features can be seen in the infrared image as shown in Figure 22.



**Figure 21.** Layout of tested specimens: a—felt, b—office carpet (wave pattern), c—leatherette, d—bubble relief styrofoam, e—cork, f—office carpet, g—polyurethane foam, h—carpet with short fibres, i—anti-slippery matt, j—soft foam with wave pattern, k—felt with pattern, l—spruce wood, m—sandpaper, n—wallpaper, o—buble foam, p—plush, q—fake grass, r—aluminum thermofoil.
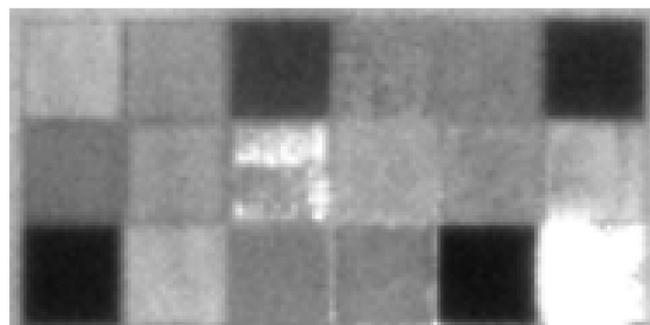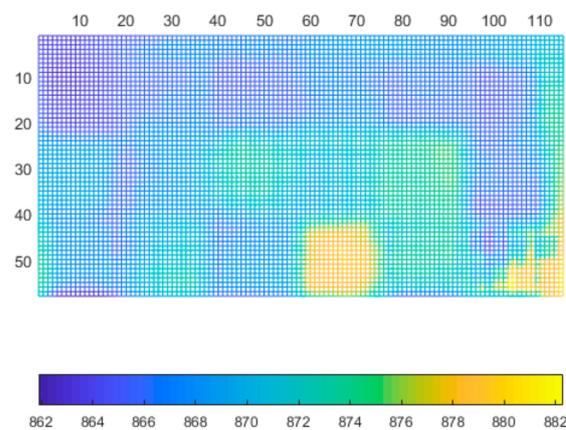


**Figure 22.** Infrared image of tested specimens.

We repeated the measurement 300 times and calculated the average and standard deviation. The results are presented in Figures 23 and 24.
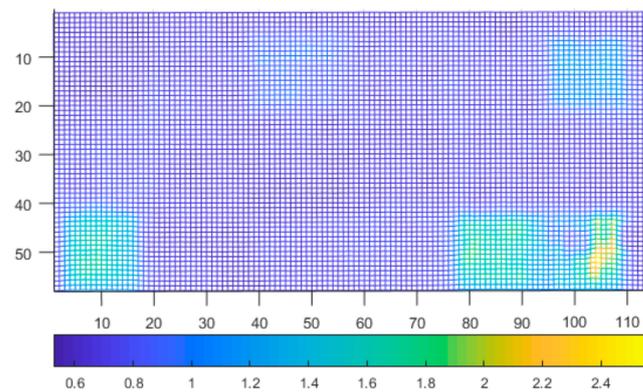
**Figure 23.** Standard deviation of tested specimens.



**Figure 24.** Average distance of tested specimens.

As can be seen, there is a correlation between the standard deviation and reflectivity. The less reflective materials have higher standard deviation. An exception was the aluminum thermofoil (bottom right), which had a mirroring effect, thus resulting in higher standard deviation and average distance. What can be seen in the distance data is, that some types of materials affect the reported distances. These materials are either fuzzy, porous or partially transparent.

### 4.5. Precision Variability Analysis

The Azure Kinect works in 4 different modes, what greatly enhances the variability of this sensor. While the narrow mode (NFOV) is great at higher precision scanning of smaller objects for its small noise and angular range, for many applications such as movement detection, mobile robot navigation or mapping, the wide mode (WFOV) provides unrivaled angular range to other RGBD cameras in this price range. From our experiments, we concluded that in the latter mode the data contains some peculiarities which should be reckoned. As can be seen in Figure 8, the noise of the data raises from the center in every direction, making it considerably higher at the end of the range. We suspected that the source of this rise could be due to one or all of these three aspects:

1. The rise of the noise is due to distance. Even though the reported distance is approximately the same, the actual Euclidian distance from the sensor chip is considerably higher (Figure 9).
2. The sensor measures better at the center of the image. This could be due to optical aberration of the lens.
3. The relative angle between the wall and sensor. This angle changes from center to the edges changing the amount of reflected light back to the sensor, which could affect the measurement quality.

The source of the worsened quality could be either of these factors, even all three combined; therefore, we devised an experiment specifically for each of these factors.

To explore the first aspect, we analyzed the standard deviation computed from data acquired by measuring a wall located 1.2 m away from the sensor as shown in Figure 25. The blue curve is the standard deviation of the distance from the plane; the orange curve is the same curve compensated with the real Euclidian distance from the sensor. When comparing the curves, it is clear that the growing Euclidian distance from the lens has direct impact on the noise. Only the extreme left and right areas indicate growing noise.
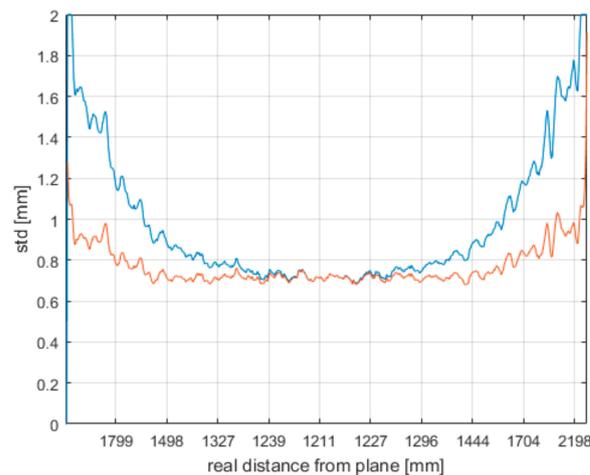


**Figure 25.** Correlation of noise and growing Euclidian distance from the sensor (blue curve—standard deviation of original data; orange curve—original data compensated with real Euclidian distance from the lens).

To test the aberration of the lens, we placed the Azure at one spot 100 cm away from the wall and rotated the sensor about its optical center for every measurement. At each position we made multiple measurements (1100) and calculated the standard deviation. We focused on the same spot on the wall. This way, the distance from the wall at the area of interest did not change; the relative angle between the wall and the area of interest captured by the sensor chip remained constant as well (due to parabolic properties of the lens). The result can be seen in the Table 3.

**Table 3.** Standard deviation at different pixel locations on the chip (mean value in mm).

| Parameter | 54° | 46° | 39° | 30° | 16° | 0° |
|---|---|---|---|---|---|---|
| Std (mm) | 0.5957 | 0.6005 | 0.5967 | 0.6045 | 0.6172 | 0.6264 |

As can be seen, there is no significant difference between measurements, so we concluded there is no drop in quality of the measurement in respect to the angular position. But there is a considerable drop of quality at the edge of the angular range. But this can be seen on all versions of the Kinect.

The third sub-experiment was aimed at the measurement quality with respect to the relative angle between the wall and the sensor; for that we used the same data as in the first sub-experiment. Our point was to focus on 3D points with the same Euclidian distance from the sensor, but with different angle between the wall and the chip (Figure 9).

From all measured positions, we selected three adjacent depth image rows located close to the center of the image and computed the standard deviation. Then we computed the average of each triplet; this resulted in one dataset (comprising of 8 lines-one for each distance) for further processing. For each point of the final dataset, we stored its distance from the wall, angle between the sensor and the wall and its standard deviation (Figure 26).
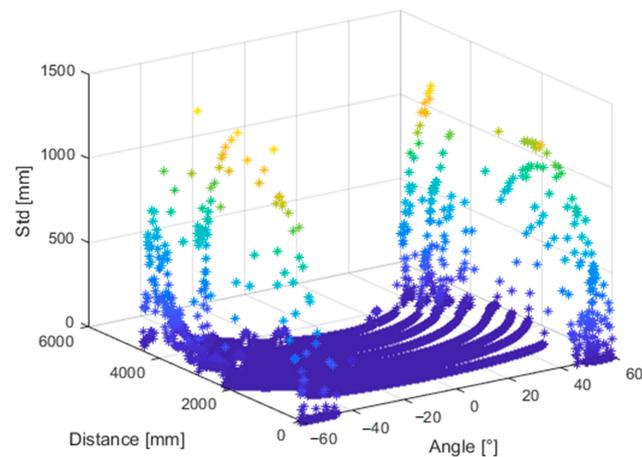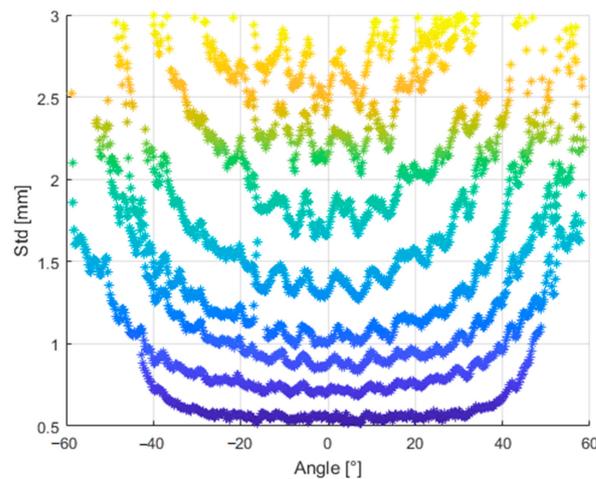
**Figure 26.** Standard deviation of noise of the Kinect Azure with respect to distance from the object (wall) and relative angle between the object and sensor.

Figures 27 and 28 presents the same dataset without extreme values that are irrelevant for further processing. It is clear that there is a correlation between relative angle of measurement and its quality.
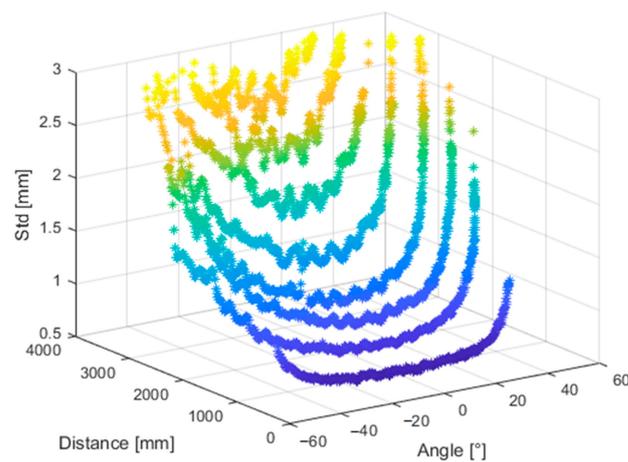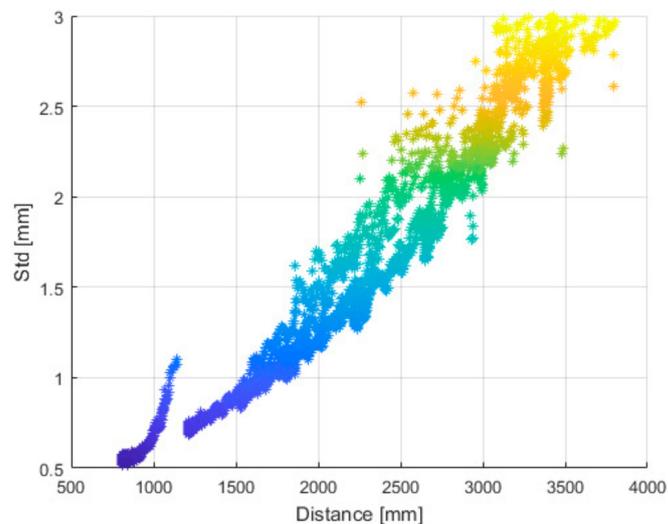


**Figure 27.** Truncated standard deviation of noise of the Kinect Azure with respect to distance from the object and relative angle of the object and sensor.



**Figure 28.** Standard deviation of noise of the Kinect Azure with respect to the relative angle of the object and sensor measured at different distances.

The rotation of previous figure denotes the dependence even more (Figure 29).



**Figure 29.** Standard deviation of noise of the Kinect Azure with respect to distance from the object with variable relative angles of the object and sensor.

To highlight these dependences even more, Figure 30 depicts the standard deviation for particular identical distances with respect to the angle for which we measured.
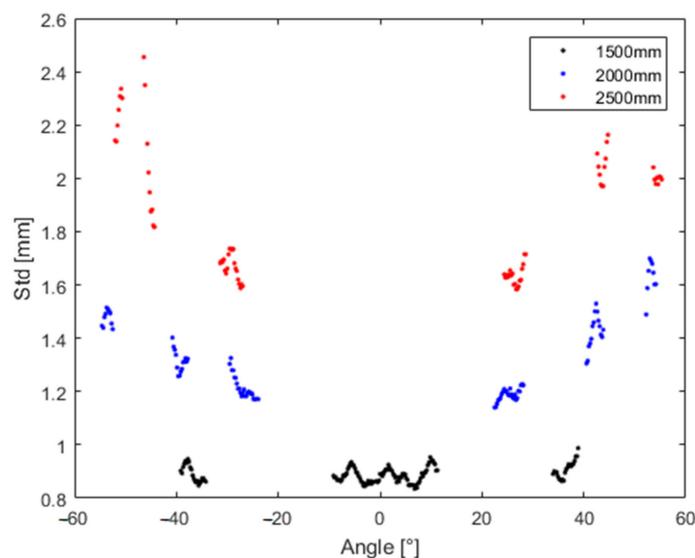


**Figure 30.** Standard deviation for particular identical distances with respect to the angle for which we measured.

Thus, we concluded that the relative angle between the sensor and the measured object plays an important role in the quality of measurement.

### 4.6. Performance in Outdoor Environment

In the final set of experiments, we examined the Azure Kinect's performance in outdoor environment. We designed two different scenarios. In the first one, the sensor was facing the sun directly while scanning the test plate (Figure 31). In the second one, the sun was outside sensor's field of view while shining directly on the test plate (Figure 32). We computed the standard deviation for NFOV and WFOV binned modes on 300 depth images and limited the result to 200 mm for better visibility (Figures 33 and 34 report the results for the first experiment, Figures 35 and 36 report the results for the second experiment).
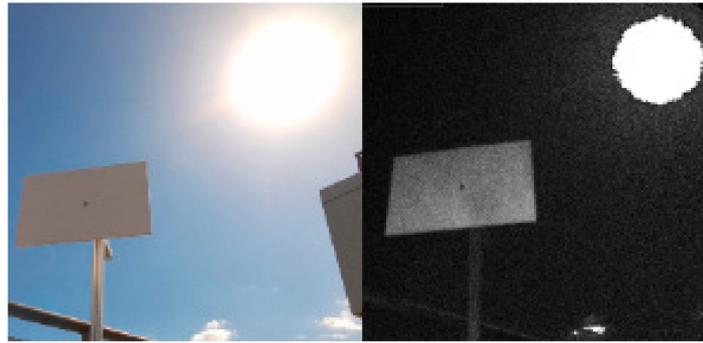
**Figure 31.** RGB and IR image of the first experiment scenario.



**Figure 32.** RGB and IR image of the second experiment scenario.
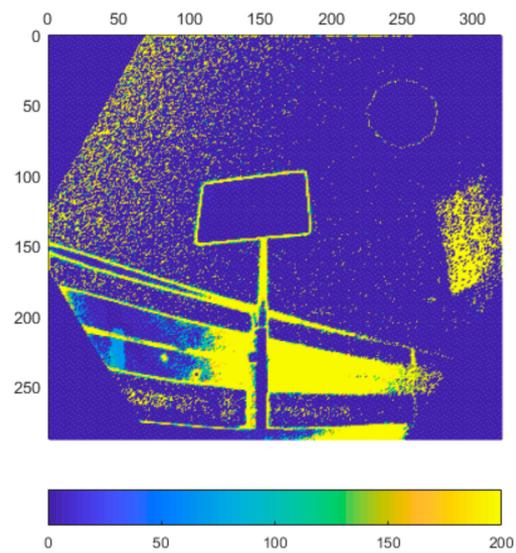


**Figure 33.** Standard deviation of binned NFOV mode limited to 200 mm (experiment 1).

**Figure 34.** Standard deviation of binned WFOV mode limited to 200 mm (experiment 1).



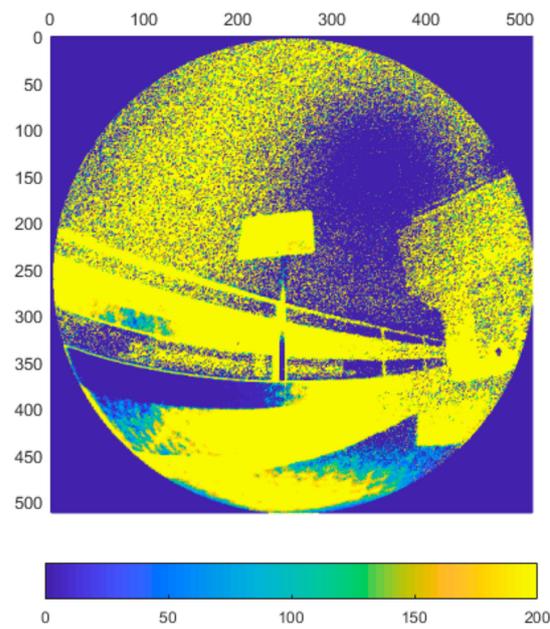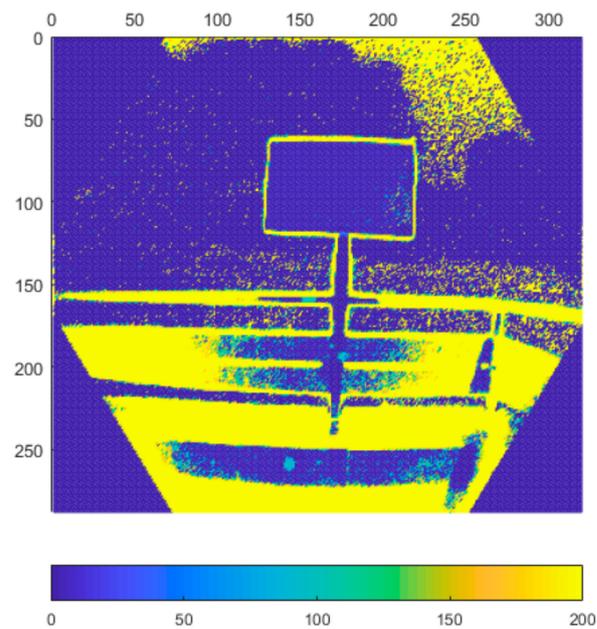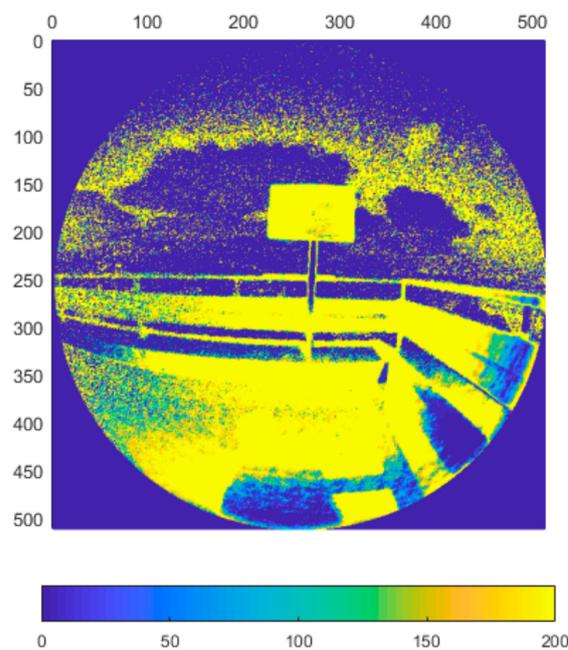**Figure 35.** Standard deviation of binned NFOV mode limited to 200 mm (experiment 2).

**Figure 36.** Standard deviation of binned WFOV mode limited to 200 mm (experiment 2).

As can be seen from the figures, in all measurements there is noise in air around the test plate despite the fact there was no substantial dust present. Both WFOV mode measurements are extremely noisy, what makes this mode unusable for outdoor environment. NFOV mode shows much better results; this is most likely caused by the fact that different projectors are used for each of the modes. Surprisingly, the direct sun itself causes no substantial chip flare outside its exact location in the image.

Even though the NFOV mode gave much better results, the range of acceptable data was only within 1.5 m distance. Even though the test plate shows little noise there were many false measurements (approximately 0.3%), which did not happen in indoor environment at all. With growing distance false measurement count grows rapidly. Therefore, we conclude that even the NFOV binned mode usability in outdoor environment is highly limited.

### 4.7. Multipath and Flying Pixel

As stated in the official online documentation, the Azure Kinect suffers from multipath interference. For example, in corners, the IR light from the sensor is reflected off one wall onto the other. This results in invalidated pixels. Similarly, at the edges of objects, pixels can contain mixed signal from foreground and background. This phenomenon is known as the flying pixel problem. To demonstrate this, we put a plate 4 mm thick in front of a wall and focused on the data acquired around the edges of the plate. As can be seen in Figures 37 and 38, the depth data located at the edges of the plate are inaccurately placed outside the actual object.
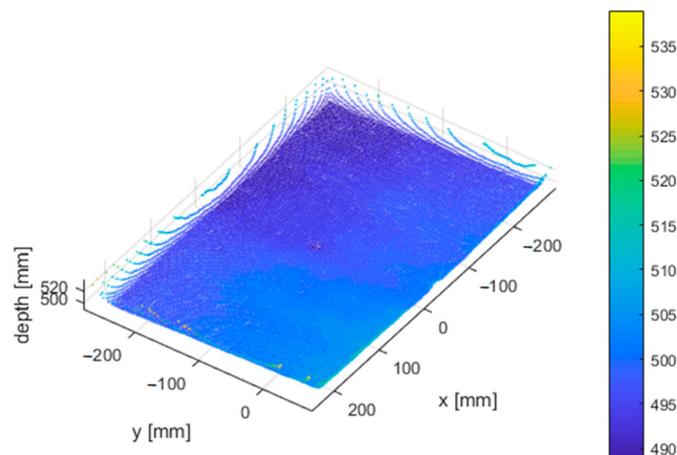
**Figure 37.** Demonstration of the flying pixel phenomenon–fluctuating depth data at the edge of the plate (front view, all values in mm).
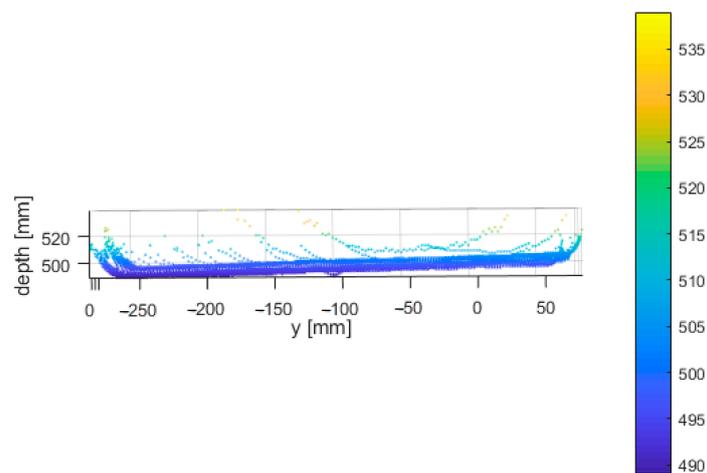


**Figure 38.** Demonstration of the flying pixel phenomenon–fluctuating depth data at the edge of the plate (side view, all values in mm).

## 5. Conclusions

We performed series of experiments to thoroughly evaluate the new Azure Kinect. The first set of experiments put the Azure in context with its predecessors, and in the light of our experiments it can be said that in terms of precision (repeatability) its performance is better than both previous versions.

By examining the warm-up time, we came to the conclusion that it shows the same behavior as Kinect v2 and needs to be properly warmed up for about 50–60 min to give stable output. We examined different materials and their reflectivity. We determined the precision and accuracy of the Azure Kinect and discussed why the precision varies.

All our measurements confirm that both, the standard deviation, and systematic error of the Azure Kinect are within the values specified by the official documentation:

- Standard deviation $\leq$ 17 mm.
- Distance error < 11 mm + 0.1% of distance without multi-path interference

The obvious pros and cons of the Azure Kinect are the following:

PROS

- Half the weight of Kinect v2
- No need for power supply (lower weight and greater ease of installation)
- Greater variability–four different modes
- Better angular resolution

- Lower noise
- Good accuracy

CONS

- Object reflectivity issues due to ToF technology
- Virtually unusable in outdoor environment
- Relatively long warm-up time (at least 40–50 min)
- Multipath and flying pixel phenomenon

To conclude, the Azure Kinect is a promising small and versatile device with a wide range of uses ranging from object recognition, object reconstruction, mobile robot mapping, navigation and obstacle avoidance, SLAM, to object tracking, people tracking and detection, HCI (human–computer interaction), HMI (human–machine interaction), HRI (human–robot interaction), gesture recognition, virtual reality, telepresence, medical examination, biometry, people detection and identification, and more.

## References

1. Elaraby, A.F.; Hamdy, A.; Rehan, M. A Kinect-Based 3D Object Detection and Recognition System with Enhanced Depth Estimation Algorithm. In Proceedings of the 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 1–3 November 2018; pp. 247–252. [CrossRef]
2. Tanabe, R.; Cao, M.; Murao, T.; Hashimoto, H. Vision based object recognition of mobile robot with Kinect 3D sensor in indoor environment. In Proceedings of the 2012 Proceedings of SICE Annual Conference (SICE), Akita, Japan, 20–23 August 2012; pp. 2203–2206.
3. Manap, M.S.A.; Sahak, R.; Zabidi, A.; Yassin, I.; Tahir, N.M. Object Detection using Depth Information from Kinect Sensor. In Proceedings of the 2015 IEEE 11th International Colloquium on Signal Processing & Its Applications (CSPA), Kuala Lumpur, Malaysia, 6–8 March 2015; pp. 160–163.
4. Xin, G.X.; Zhang, X.T.; Wang, X.; Song, J. A RGBD SLAM algorithm combining ORB with PROSAC for indoor mobile robot. In Proceedings of the 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), 19–20 December 2015; pp. 71–74.
5. 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]
6. Ibragimov, I.Z.; Afanasyev, I.M. Comparison of ROS-based visualslam methods in homogeneous indoor environment. In Proceedings of the 2017 14th Workshop on Positioning, Navigation and Communications (WPNC), Bremen, Germany, 25–26 October 2017.
7. Plouffe, G.; Cretu, A. Static and dynamic hand gesture recognition in depthdata using dynamic time warping. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 305–316. [CrossRef]
8. Wang, C.; Liu, Z.; Chan, S. Superpixel-based hand gesture recognition withKinect depth camera. *IEEE Trans. Multimed.* **2015**, *17*, 29–39. [CrossRef]
9. Ren, Z.; Yuan, J.; Meng, J.; Zhang, Z. Robust part-based hand gesture recognition using kinect sensor. *IEEE Trans. Multimed.* **2013**, *15*, 1110–1120. [CrossRef]
10. Avalos, J.; Cortez, S.; Vasquez, K.; Murray, V.; Ramos, O.E. Telepres-ence using the kinect sensor and the nao robot. In Proceedings of the 2016 IEEE 7th Latin American Symposium on Circuits & Systems (LASCAS), Florianopolis, Brazil, 28 February–2 March 2016; pp. 303–306.
11. Berri, R.; Wolf, D.; Osório, F.S. Telepresence Robot with Image-Based Face Tracking and 3D Perception with Human Gesture Interface Using Kinect Sensor. In Proceedings of the 2014 Joint Conference on Robotics: SBR-LARS Robotics Symposium and Robocontrol, Sao Carlos, Brazil, 18–23 October 2014; pp. 205–210.
12. Tao, G.; Archambault, P.S.; Levin, M.F. Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis. In Proceedings of the 2013 International Conference on Virtual Rehabilitation (ICVR), Philadelphia, PA, USA, 26–29 August 2013; pp. 164–165.
13. Satyavolu, S.; Bruder, G.; Willemsen, P.; Steinicke, F. Analysis of IR-based virtual reality tracking using multiple Kinects. In Proceedings of the 2012 IEEE Virtual Reality (VR), Costa Mesa, CA, USA, 4–8 March 2012; pp. 149–150.
14. Gotsis, M.; Tasse, A.; Swider, M.; Lympouridis, V.; Poulos, I.C.; Thin, A.G.; Turpin, D.; Tucker, D.; Jordan-Marsh, M. Mixed realitygame prototypes for upper body exercise and rehabilitation. In Proceedings of the 2012 IEEE Virtual Reality Workshops (VRW), Costa Mesa, CA, USA, 4–8 March 2012; pp. 181–182.
15. Heimann-Steinert, A.; Sattler, I.; Otte, K.; Röhling, H.M.; Mansow-Model, S.; Müller-Werdan, U. Using New Camera-Based Technologies for Gait Analysis in Older Adults in Comparison to the Established GAITRite System. *Sensors* **2019**, *20*, 125. [CrossRef] [PubMed]
16. Volák, J.; Koniar, D.; Hargas, L.; Jablončík, F.; Sekel'Ova, N.; Durdík, P. RGB-D imaging used for OSAS diagnostics. *2018 ELEKTRO* **2018**, 1–5. [CrossRef]
17. Guzsvinecz, T.; Szucs, V.; Sik-Lanyi, C. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. *Sensors* **2019**, *19*, 1072. [CrossRef] [PubMed]
18. 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]
19. Smisek, J.; Jancosek, M.; Pajdla, T. 3D with Kinect. In *Consumer Depth Cameras for Computer Vision*; Springer: Lomdom, UK, 2013; pp. 3–25.
20. Fankhauser, P.; Bloesch, M.; Rodriguez, D.; Kaestner, R.; Hutter, M.; Siegwart, R. Kinect v2 for mobile robot navigation: Evaluation and modeling. In Proceedings of the 2015 International Conference on Advanced Robotics (ICAR), Istanbul, Turkey, 27–31 July 2015; pp. 388–394.
21. Choo, B.; Landau, M.; Devore, M.; Beling, P. Statistical Analysis-Based Error Models for the Microsoft KinectTM Depth Sensor. *Sensors* **2014**, *14*, 17430–17450. [CrossRef] [PubMed]
22. Pagliari, D.; Pinto, L. Calibration of Kinect for Xbox One and Comparison between the Two Generations of Microsoft Sensors. *Sensors* **2015**, *15*, 27569–27589. [CrossRef] [PubMed]
23. 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]

24. Zennaro, S.; Munaro, M.; Milani, S.; Zanuttigh, P.; Bernardi, A.; Ghidoni, S.; Menegatti, E. Performance evaluation of the 1st and 2nd generation Kinect for multimedia applications. In Proceedings of the 2015 IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6.
25. 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; pp. 94–96.
26. Wasenmüller, O.; Stricker, D. Comparison of Kinect V1 and V2 Depth Images in Terms of Accuracy and Precision. In Proceedings of the 13th Asian Conference on Computer Vision, Taipei, Taiwan, 20−24 November 2016; pp. 34–45.