Contents

EVALUATING THE DEPTH RESOLUTION OF 3D SENSORS FOR MANUFACTURING AUTOMATION APPLICATIONS

Prem Rachakonda, Geraldine Cheok, Marek Franaszek, Kamel Saidi
Sensing and Perceptions Systems Group
Intelligent Systems Division,
National Institute of Standards and Technology,
Gaithersburg, MD

ABSTRACT: The Intelligent Systems Division (ISD) at the National Institute of Standards and Technology (NIST) is performing research to support the development of documentary standards within ASTM subcommittee E57.23 on Industrial 3D Machine Vision Systems. This subcommittee is addressing the performance evaluation of 3D imaging systems that are used for manufacturing automation and vision guided robotics. The applications of these systems extend to other areas such as robotic bin picking, autonomous vehicles, 3D reconstruction, and entertainment. This paper discusses the activities of one work item (WK73176) that is addressing the evaluation of the depth resolution of a 3D imaging system. The key contributions of this paper include the description of multiple methods to determine the resolution of a 3D imaging system.

KEYWORDS: ASTM E57, 3D imaging systems, documentary standards, depth resolution

1 INTRODUCTION

The Intelligent Systems Division (ISD) at the National Institute of Standards and Technology (NIST) has been involved in the development of documentary standards for 3D imaging systems through ASTM committee E57 on 3D imaging systems since 2006. In 2019 ISD held a workshop at NIST to seek input from industry leaders, practitioners, and researchers from around the world on the topic of standards for 3D perception systems for robotic assembly applications [1]. After this workshop, four topic areas were selected as the ones that will have the most significant impact and four task groups were formed to address them. The four working items are:

1. WK72962 - Standard Test Method for Measuring the Performance of a 3D Perception System Across the Specified Field-of-View
2. WK73176 - Standard Test Methods for Determination of a 3D Perception Systems Point Wise Spatial Resolution
3. WK78941 - Standard Test Method for Performance of Machine Vision systems for Bin-picking Systems
4. WK81247 - Standard Guide for Selection of 3D Vision Technologies for Industrial Applications

Of the four work items, two (WK72962 and WK73176) were established in 2020 and the other two were established in 2022. These four work items are led by industry experts, with support from NIST. The Sensing and Perception Systems Group at NIST is equipped with state-of-the-art instrumentation for reference measurements, various 3D sensors, and an environmentally stable laboratory. The group also has access to various other calibration laboratories, fabrication facilities, and experts in various fields of science, engineering, and mathematics. The rest of this paper will focus primarily on the work of the task group working on WK73176, hereafter referred to as the "resolution task group."

The spatial resolution of a 3D imaging system can be along any of three Cartesian coordinate system directions (Figure 1). Two of the axes (X and Y) are in the plane of the sensor, originating at the sensor's calibrated center, and the third axis, Z is orthogonal to the X-Y plane. Z is typically along the depth direction as shown in Figure 1 (note that, for this sensor, the front bezel is parallel to the sensor plane by design). During the initial meetings of the resolution task group, participants decided to first develop a standard for the resolution of a system along the Z axis (i.e., the depth resolution).

The resolution task group considered several standards [2, 3, 4, 5, 6] and consulted the definition of "resolution" in the international vocabulary of metrology, VIM [7], which states that resolution is the "smallest change in a quantity being measured that causes a perceptible change in the corresponding indication." The group then adapted this definition and drafted the following definition for depth resolution: "the smallest change in depth that causes a perceptible change in the corresponding depth measurement indication." Though spatial resolution in X and Y is also relevant for a 3D imaging system, determining depth resolution is a first step in developing metrics relevant to many 3D vision applications.

Evaluating depth resolution will help end users select 3D sensors for their application as sensors with higher resolution should be able to distinguish smaller features. However, depth resolution determined based on one feature (e.g., a plane) may be very different and optimistic compared to the resolution determined by a non-planar feature that may result in phantom points[*][8] due to edges or other optical phenomena. Resolution metrics based on arithmetic precision[†][9] or pixel spacing are not useful as they may not represent the sensor's ability to distinguish between two features.

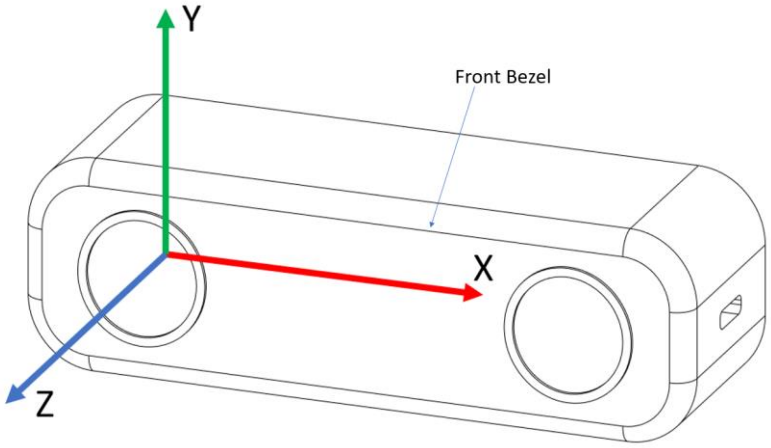


Figure 1: A 3D imaging system with the depth direction along the Z-axis

In this work we explore various methods that were considered to calculate depth resolution and will discuss their capabilities and limitations, as well as their relevance to 3D vision applications. Six different sensors were used and some of their specifications are given in Table

[*] Phantom points are erroneous data points that do not depict actual physical features within the scanned area.
[†] Arithmetic precision is the precision of the floating-point representation of a number.

1‡. These six sensors can be categorized into three sensor types a) active stereo sensors, b) structured-light sensors (time coded), and c) time-of-flight sensors.

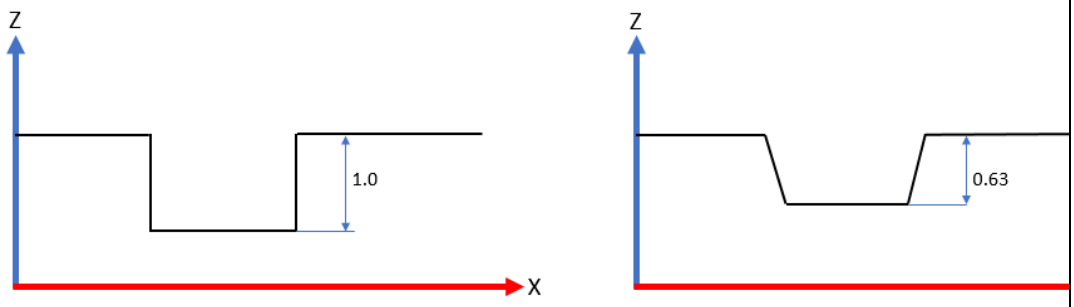
Table 1: Relevant specifications of the sensors used in this work [10]‡

Sensors	Technology	Shutter Type	Frames per sec.	Light Source Wavelength	Depth Image # of Pixels	Depth range (m)
Z1	Structured light	N/A	<1§	White light	1920 × 1080	0.70 – 1.5
E2	Active stereo	Global	<1§	465 nm	1280 × 1024	0.50 – 3.0
D3	Active stereo	Global	30-90	≈ 850 nm	1280 × 720	0.60 – 6.0
L4	Time-of-flight	Global	30	≈ 850 nm	1024 × 768	0.25 – 9.0
K5	Time-of-flight	Rolling	30	≈ 860 nm	1024 × 1024	0.25 – 5.5
R6	Active stereo	Global	30-90	≈ 850 nm	1280 × 720	0.27 – 3.0

Section 2 will review the literature and existing methods for evaluating 3D imaging systems, and will list the sensors used for this work; Section 3 will discuss various methods that were proposed to evaluate sensors, with Section 3.7 describing one method that achieved a general agreement among the resolution task group members; Section 4 will then discuss the test setup, data analysis, and results; and Section 5 will offer concluding remarks.

2 STATE OF THE ART & REVIEW OF LITERATURE

There are a few standards that are used to evaluate some 3D imaging systems [6], two of which are the German guidelines VDI/VDE 2634 part 2 and part 3 [11, 12], which have been the de facto standards for these sensors. The third one is an ISO 10360-13 standard [13] (published in 2021) that extends VDI/VDE 2634 and introduces methods for determining "structural resolution" in its non-mandatory (informative) appendix. ISO 10360-13 defines "structural resolution" as the distance or depth that is 0.63 times the calibrated depth of a feature (e.g., a pit illustrated in Figure 2).



a) Designed pit depth b) Measured pit depth

Figure 2: ISO 10360-13 recommendations for measuring the "structural resolution"

Many 3D imaging device manufacturers refer to the resolution as the smallest distance in a mesh on a surface [14] regardless of geometry or use the spacing of the points in the X and Y axis (non-depth axes) at the focal plane of the sensor. However, the sensor's ability to resolve

‡ Commercial equipment and materials may be identified to specify certain procedures. In no case does such identification imply recommendation or endorsement by the NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

§ There are no specifications for frames/second for these sensors. Empirical frame rate values are used here.

features also depends on the optics used, target properties, etc. Such specifications (based on grid spacing or mesh distance) are not representative of the sensor's real-world resolution.

Gomez et. al [15] describe an artifact with cylindrical holes for qualitative evaluation of 3D sensor resolution. Their quantitative evaluation of the same artifact involves comparison of the hole diameter and hole center-to-center distances from the sensor to the same measurements obtained from a coordinate measuring machine (CMM). Another facet of the same artifact has step features to be used to determine the ability of the sensor to resolve steps.

Neupane et. al [16] evaluate eight 3D sensors for an agronomy application (fruit picking) in outdoor environments (under sunlight). Giancola et. al [17] report evaluations on various sensors and technologies, however lateral resolution (i.e., in the X and Y directions in Figure 1) was specified in terms of the sensor pixel resolution, not the ability to distinguish features. Haggag et. al [18] define resolution in two ways, one as the distance between two unique depth values (as explored in this work), but also as the smallest possible distance between the digital representation of the depth data (bit values). Several other artifacts used for LADARs [19, 20], Computed Tomography [21, 22], and millimeter wave imaging systems [23] were also considered; however, many of these artifacts were better suited for qualitative evaluation of those systems. Gaps remain in the quantitative specification and evaluation of distinguishing features in terms of depth and lateral resolution.

In this work, we explore both the qualitative and quantitative methods and highlight the capabilities and limitations of each method to calculate depth resolution.

3 ARTIFACT DESIGNS AND TEST METHODS EXPLORED

Several artifacts were designed, and tests were conducted by NIST with input from the resolution task group members. This was to understand the performance and data quality of 3D sensors. The next sections will discuss several methods that were considered for this activity. Section 3.7 will describe a method that received consensus in the resolution task group due to the ease of implementation compared to the other methods.

3.1 Gap Artifact

This was one of the first methods suggested in the working group. The idea proposed in this method was a) to distinguish two planes using a statistical test and b) to determine the gap measurement error of the system under test (SUT) relative to the reference instrument (RI) (see Figure 3). To test this method, artifacts were fabricated using 3D printers and laser cutters. Preliminary data from some of the sensors showed that the gap between the two front planes also makes the determination of the depth difficult as it can hinder a clear line of sight from the two camera sensors when a stereo sensor is used. In another setup, a pair of aluminum plates was mounted on translation stages, but the lack of squareness between these two flat plates made it challenging to perfectly align the plate with respect to each other. Gap measurement error is not the same as the depth resolution as the spacing between the plates in the lateral direction influences the ability to distinguish the spacing the longitudinal direction.

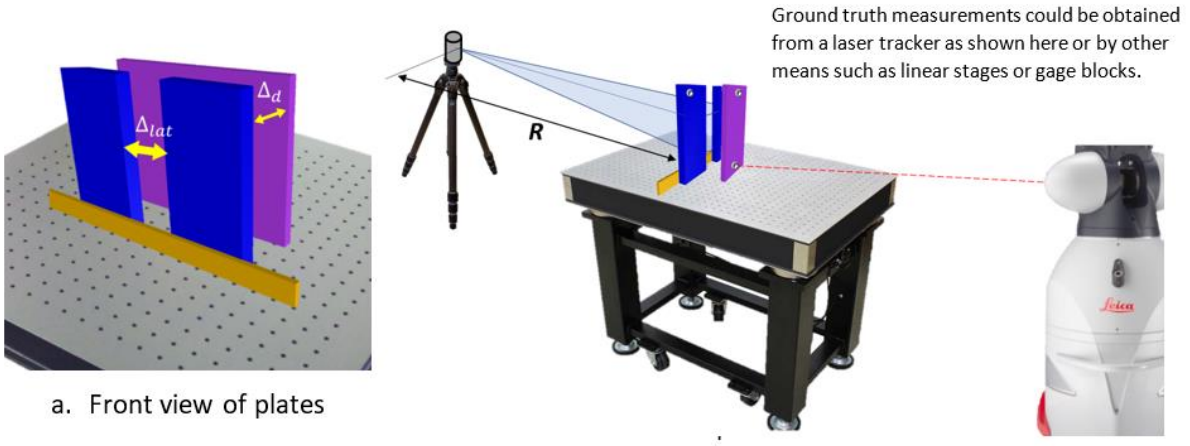


Figure 3: A resolution test using flat planes and reference length measurements[**]

The drawbacks of this test are:

- *the presence of the edges causes spurious points for some systems such as some structured-light sensors (see Figure 4). Exclusion of such points could be complex and may vary from one sensor to another*
- *it is challenging to decouple depth and lateral resolution, as a) the gap in the lateral direction affects the determination of the error in the longitudinal direction and b) the presence of a target behind the front plates has an influence on the data from the front plates; and*
- *data from some of the sensors is highly correlated spatially and the use of statistical tests assumes that there is no such correlation (see section 3.7.1).*

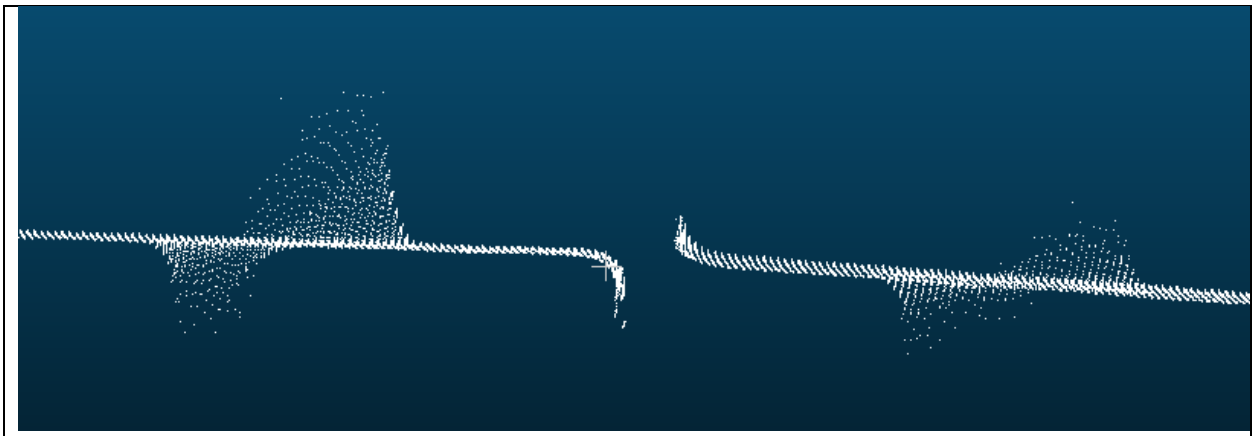


Figure 4: Spurious points at the edges of flat planes as measured by a structured-light system.

3.2 Sinusoidal Artifact

Another artifact developed was a sinusoidal artifact that was designed to have various wavelengths (see Figure 5 and Figure 6) and it was measured using multiple sensors. This was based on the non-mandatory test method suggested in ISO 10360-13 [24].

[**] Figure credit: Felix Thouin, formerly of Airy3D

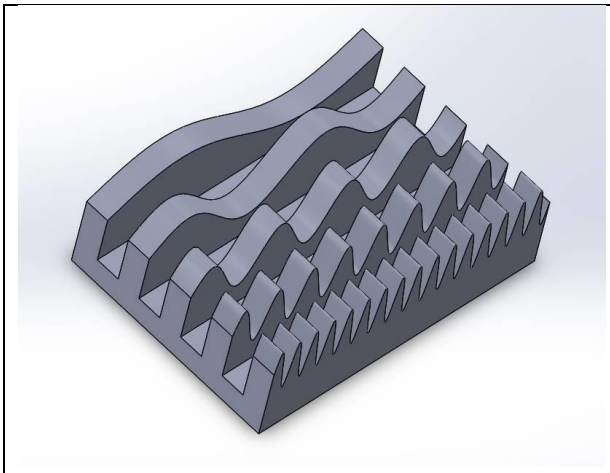


Figure 5: CAD model of sinusoidal artifact

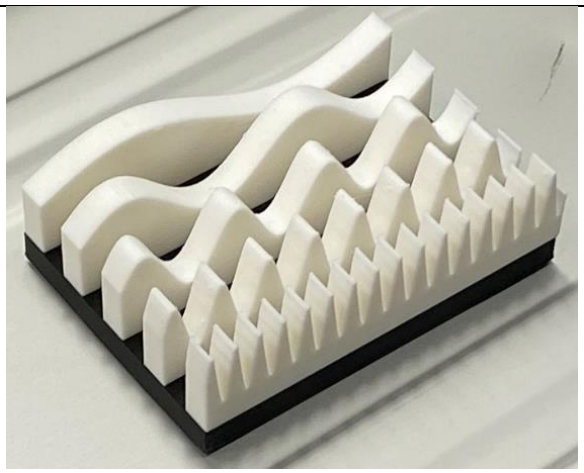


Figure 6: 3D printed sinusoidal artifact

Figure 8 and Figure 11 show the 3D scans of this artifact from two different kinds of sensors. The computer aided design (CAD) model of the target used is shown in Figure 5 and a 3D printed target used for the test is shown in Figure 6. The designed amplitude of the artifact was 6.35 mm, and the wavelengths were 79.8 mm, 39. 9 mm, 19.95 mm, 9.98 mm, and 4.99 mm. For this artifact, the data corresponding to a particular wavelength (19.95 mm) was manually extracted, but no attempt was made to exclude any spurious points other than those belonging to the mounting apparatus. A principal component analysis (PCA) was performed to calculate a rotation matrix (R) in which the dominant features exist. The data was then rotated using the rotation matrix R, resulting in the data's Z-coordinate aligning with the thickness of the stripe of the sinusoidal wave . The illustration in Figure 10 depicts the effect of PCA on data to highlight the feature of interest such as the wavelength or the amplitude. Such transformations can change the orientations. For example, the amplitude in Figure 7 that was in the depth direction is now in the vertical direction after the PCA was performed.

If the Z-coordinate is ignored, the data in the X and Y axes denote the feature of interest, the sine wave. This reduces a 3D problem to a 2D problem. Fast Fourier Transform (FFT) analysis of two waveforms from data from two sensors are shown in Figure 9 and Figure 12, respectively. The performance of these systems varies when the sensors are placed horizontally or vertically with respect to the artifact (see illustration in Figure 7).

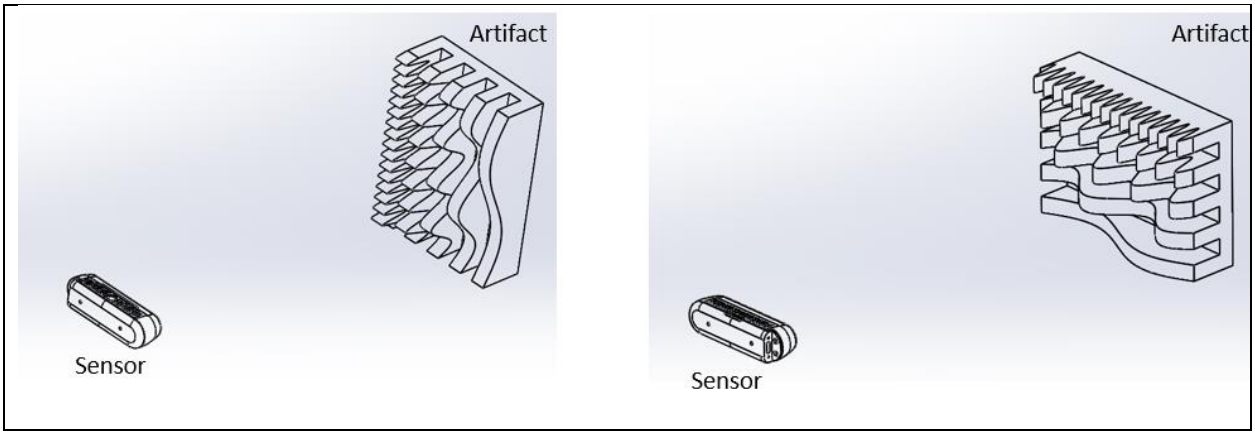


Figure 7: Illustration of the two orthogonal orientations of the artifact with respect to the sensor

Figure 39 and Figure 40 (in the supplementary material) represent scan data from two other sensors.

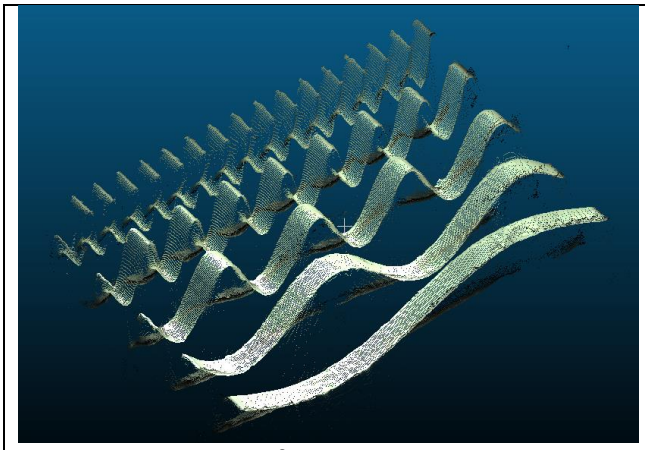


Figure 8: Scan data from Sensor Z1

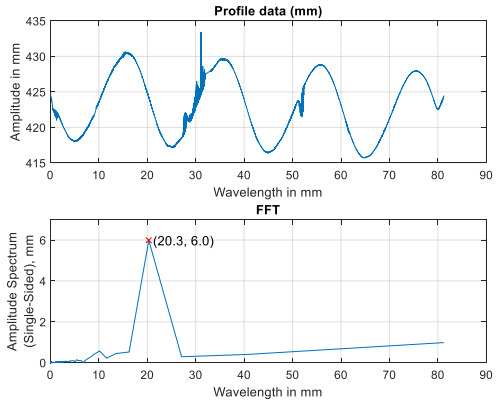


Figure 9: FFT of one sinusoidal feature using Sensor Z1

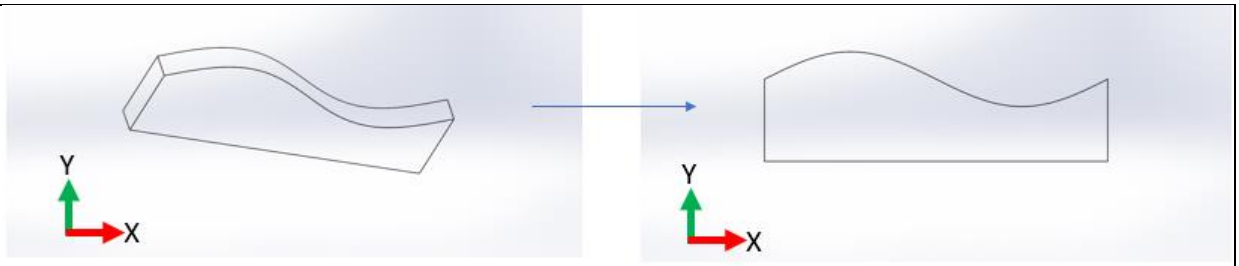


Figure 10: Transforming data using PCA

FFT analysis of the Sensor Z1 data in Figure 9 indicates that the error in the wavelength (difference between the designed and measured wavelengths) is 0.35 mm, and the error in the amplitude (difference between the designed and measured amplitudes) is -0.35 mm. For Sensor E2, the error in the wavelength is 0.95 mm and the error in the amplitude is 0.25 mm (see Figure 12). Note that these error values are with respect to the designed wavelength and not the manufactured feature wavelength. That is, the errors due to the manufacturing process that was used to produce the artifact are not captured in these calculations and can be larger or smaller than the designed dimensions. Therefore, the errors mentioned here can only be used to compare different sensors against a single artifact, but to evaluate a single sensor, ground truth measurements will be necessary .

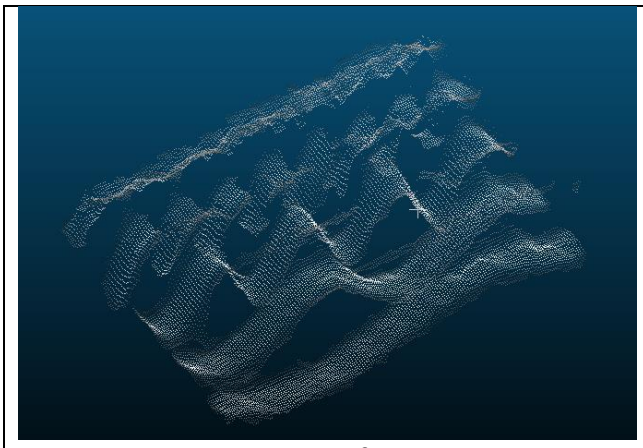


Figure 11: Scan data from Sensor E2

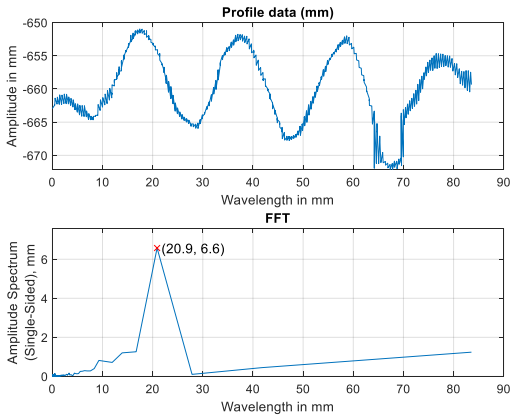


Figure 12: FFT of one sinusoidal feature using Sensor E2

Though the sinusoidal artifact was useful to evaluate sensors qualitatively, deriving a metric for this kind of artifact was challenging. The challenges were in manually segmenting relevant features from the data that are devoid of any spurious points. If the manual segmentation is not done properly, the calculated wavelength of the feature could appear longer. Missing data points due to filtering or invalid points can also lead to erroneous results. In addition, for some sensors (see Figure 12) the data deviates from the designed sinusoidal wavelength, and FFT analysis does not reveal this discrepancy.

3.3 Bessel, Step, and ISO Artifacts

A few other artifact designs were also considered such as the Bessel artifact shown in Figure 13 based on the National Physical Laboratory (NPL) artifact [25] in Figure 14, a step artifact from Airy 3D (Figure 15), and a 3D version of a section of an ISO resolution chart shown in Figure 16 (note that the artifact in Figure 13 has its axis of rotation offset from the Bessel curve's origin). Also note that the premise behind Airy 3D's artifact is very similar to the method recommended by ISO 10360-13 (shown in Figure 2 and described in Section 2).

Figure 13: 3D printed Bessel artifact[††]

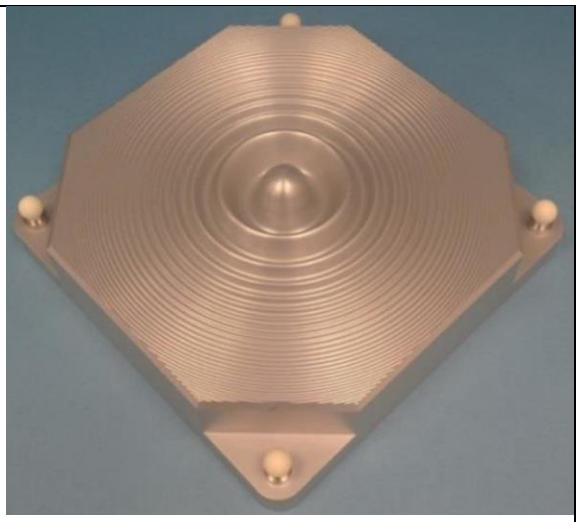


Figure 14: NPL Bessel artifact [26]

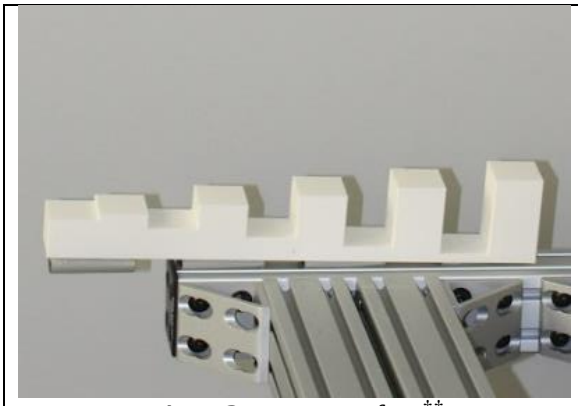


Figure 15: Airy 3D step artifact[‡‡]

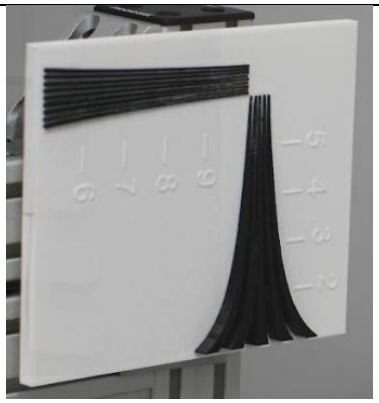


Figure 16: 3D ISO Resolution Wedges[§§]

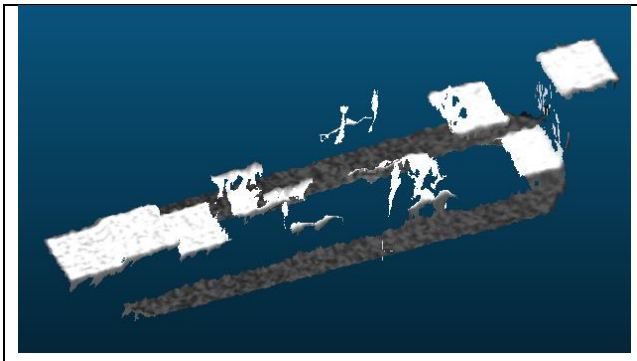


Figure 17: 3D scan of a 3D printed Airy 3D step artifact using Sensor E2

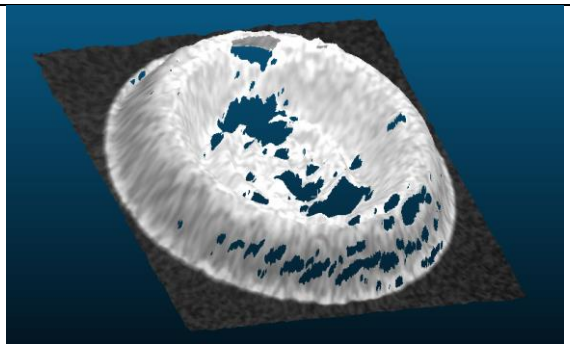


Figure 18: 3D scan of a 3D printed Bessel artifact using Sensor E2

[††] 3D model designed and fabricated at NIST
[‡‡] 3D model courtesy of Airy 3D, fabricated at NIST.
[§§] 3D printed artifact based on ISO-12233, fabricated at NIST.

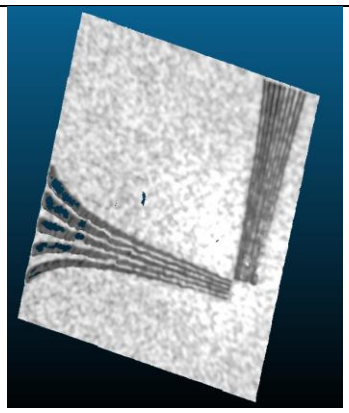


Figure 19: 3D scan of a 3D printed ISO artifact using Sensor E2

The challenges in evaluating 3D sensors using these artifacts were that the sensors could only be evaluated qualitatively, and not quantitatively. Figure 17, Figure 18, and Figure 19 show 3D scans of three of these artifacts using Sensor E2. It can be observed that the non-uniformity of the data on the surfaces of interest make it more challenging for deriving quantitative metrics. Note that several of these artifacts were explored, but there was no consensus among the resolution task group as to metrics. This is because complex target geometries and 3D sensor construction resulted in various data quality issues shown here.

3.4 Number of Unique Values on a Sphere

For this method, a sphere artifact was scanned using various scanners and the unique Z-coordinate values[***] on the surface of the sphere were counted. A higher number of unique Z-coordinate values on the surface of the sphere indicates higher depth resolution. Alternatively, the median difference in Z-coordinate value between two successive unique values is a metric of the resolution of the sensor. The unique values in X, Y, Z axes from its smallest value are shown in Figure 20 and Figure 21 for two sensors (Sensor Z1 and Sensor L4, respectively)[†††]. Note that values are sorted from the lowest to the highest and then the first value is subtracted from each of these values to start from 0. The drawback of this method is that it does not give enough information for an end user to determine if they can distinguish two features from one another. This is because a feature may have any shape and can have more than one data point on it (having more points makes the ability to distinguish features easier). Note that, for systems that apply filters that reduce noise, the number of unique values may be lower than what it would have been if no filters were applied.

[***] Unique values in this scenario are the non-repeating Z-coordinates on the surface of the target.
[†††] Figure 37 and Figure 38 are for two other sensors and are moved to supplementary information for brevity.

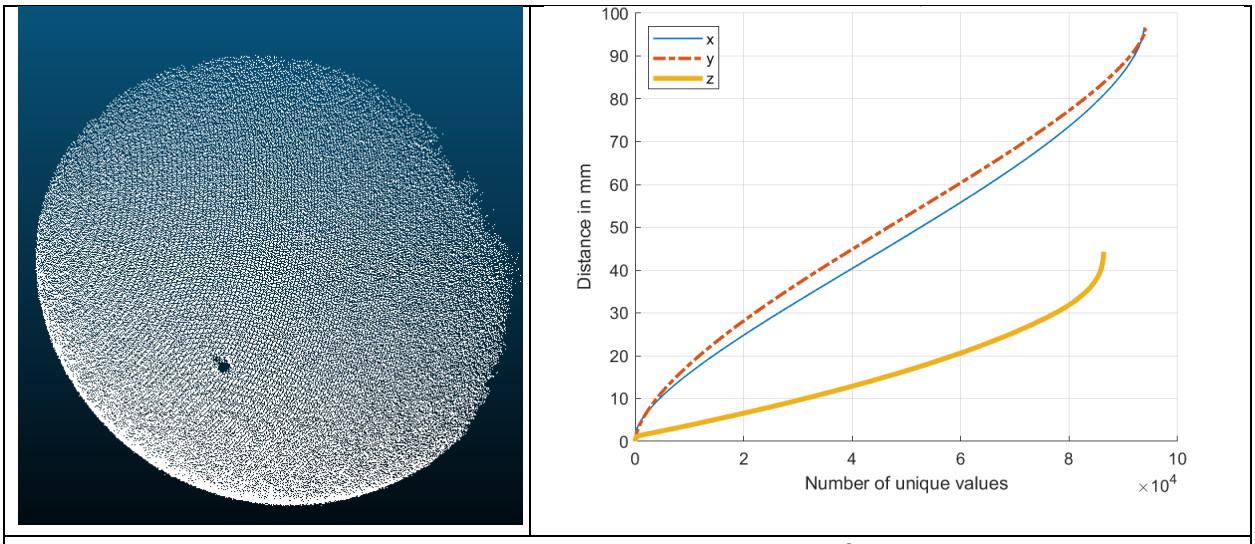


Figure 20: Sensor Z1: 3D scan and the number of unique values

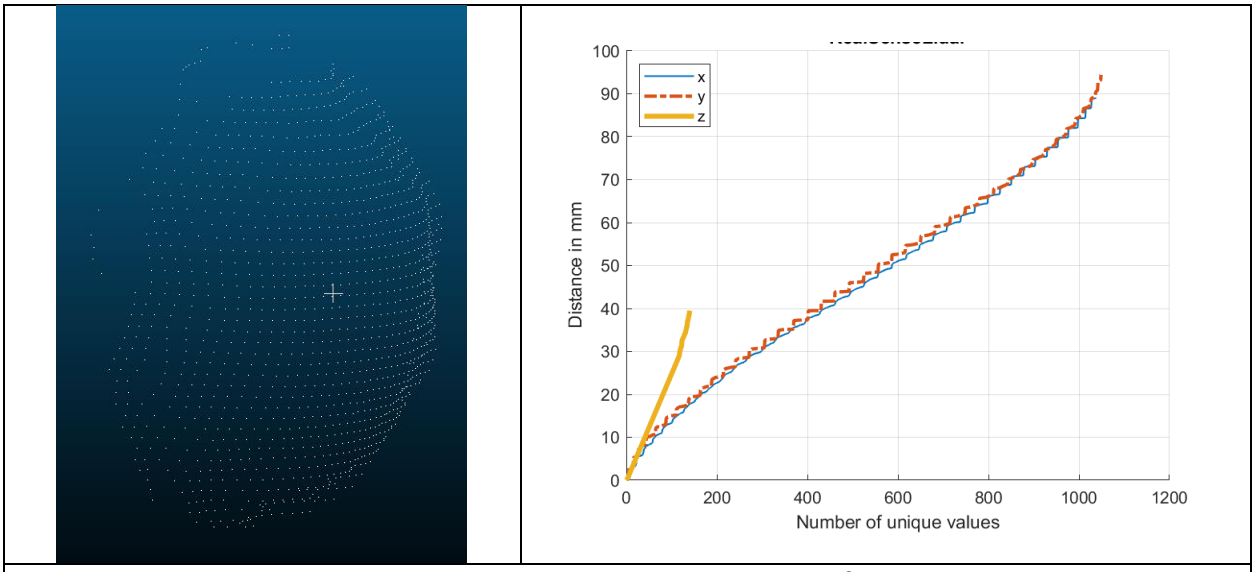


Figure 21: Sensor L4: 3D scan and the number of unique values

3.5 Wedge Artifact

One proposed method to calculate the resolution of a 3D sensor is to use two planes at an angle to one another (physically, this resembles a wedge). It could also be realized using a single planar artifact that could be tilted using precision motion stages. A wedge offers a single monolithic artifact, but same regions cannot be compared (For example, center vs off-center regions can be compared). A slanted plane that can be tilted on motion stages is not monolithic, but the same exact regions can be compared.

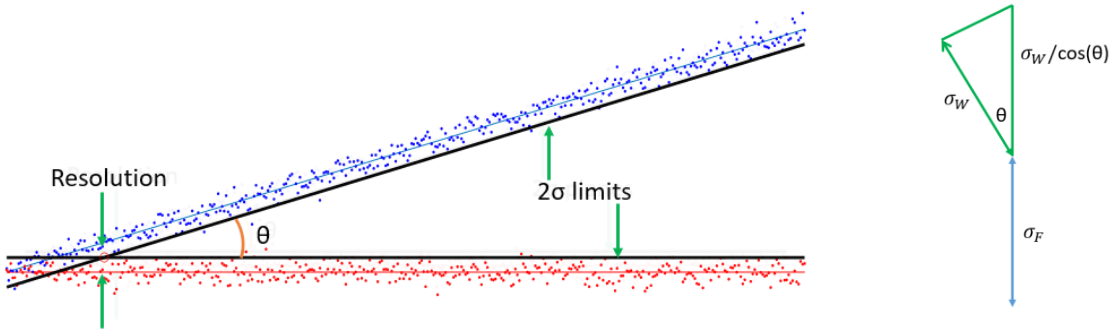
In this method, the depth resolution is defined as the vertical distance between the two planes at which the limits of the standard deviations of the residuals of the measured points from their respective orthogonal plane fits intersect (see Figure 22). If σ_F is the standard deviation of the residuals on the flat plane and σ_W is the standard deviation of the residuals on the angled plane, then the resolution (R_w) may be represented as:

$R_w = n(\sigma_F + \sigma_W/\cos(\theta))$	1

where n is a scalar multiplier (2) to encompass ≈ 95% of the residuals and θ is the angle between the two planes.

For small angles θ < 10°, $\sigma_F \approx \sigma_W$ and $R_w \approx 2n\sigma_F$. However, the standard deviation values represent the deviation of points on any part of the flat plane or the angled plane. Though this is a valid way to define resolution, it does not represent the method in which 3D sensors will be used in object detection tasks for manufacturing applications. In most manufacturing robotic tasks objects have a larger surface area and are represented by more than one point, and the objects often have complex geometries. Some of these geometries can be represented by derived points such as a plane centroid and a sphere center. The variation of these derived points will be much lower than the variation of any single point on the surface of the object. In the case of the wedge artifact, the variation of the centroid of the planar regions will be lower than the variation of any point on its surface. Therefore, the depth resolution calculated by this procedure will be much larger than when using the centroids.

In fact, for the wedge artifact described here the standard deviation of the centroid of a plane can be as low as $\sigma_F/\sqrt{N}$, where N represents the number of points on the plane. This will be true if the residuals are independent (random in nature) and have a normal distribution. That is, a 3D sensor will be able to resolve a feature better if the feature has a greater number of measured points on it. However, the *improvement* in resolving a feature will diminish with increasing number of points (due to the nature of the $1/\sqrt{N}$ factor).



a) Illustration of noise on a wedge artifact, where blue and red lines represent the fitted surfaces of the plane in 2D

b) Geometry for calculating the resolution

Figure 22: Method to use a wedge artifact.

3.6 Metric based on Two Sample T-Test

A two sample T-test involves calculating a test statistic T given by Equation 2 [27], where $\bar{Y}_1$ and $\bar{Y}_2$ are the means of the two sets of data having N_1 and N_2 samples, respectively, and standard deviations s_1 and s_2, respectively. The means of the two datasets are different if $|T| > t_{1-\alpha/2,\,\nu}$, where $t_{1-\alpha/2,\,\nu}$ is the critical value of the T-distribution with ν degrees of freedom for a 5% significance level (95% coverage).

$$T = \frac{\overline{Y_1} - \overline{Y_2}}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \quad (2)$$

Below is the procedure that was developed for this method:

1. A flat plane (see Figure 23) of length L and width W is measured by the sensor at a distance in the center of its field of view. The data is truncated to remove any edge effects and segmented in such a way that the length of the included region (called D1) is $\leq$ L/2 and the width is $\leq$ W/2.
2. The segmented data (D1) is fit to a plane and the residuals from the plane fit represent the noise on the flat plane. Consider this dataset as Y1.
3. The plane is virtually moved by a distance d to generate another dataset $Y_2 = Y_1 - d$ (see Figure 24). 'd' starts from a small value and increments in small steps. For example, 0 µm, 0.1 µm, and so on.
4. A two sample T-test is conducted for datasets Y_1 and Y_2 until the T-test indicates that the samples Y_2 and Y_1 are different.
5. The distance at which the T-test indicates that the samples are different is the resolution of the sensor.

An alternative method is to derive a formula based on the T-test metric. In this test, the data is being replicated at a distance, d_R, where d_R is the theoretical resolution of the system. Given $s_1 = s_2$ and $N_1 = N_2$, d_R is computed as follows:

$$d_R = \sqrt{2} \times k \times \frac{s_1}{\sqrt{N_1}} \quad (3)$$

where, $k = t_{1-\alpha/2,\, \nu-1}$, which can be obtained from many popular software suites such as MATLAB, Python, and Excel, or from the T-table shown in Table 5 in the Supplementary Information. The degree of freedom used is $\nu = 2N_1-3$ to have a slightly larger coverage value than $k = t_{1-\alpha/2,\, 2N-2}$ This is to ensure that the T-test metric can be used to determine that the two populations (Y_1 and Y_2) are different, i.e. resolvable.

Figure 23: A flat plane target (8"x 8" or 20.32 cm x 20.32 cm), with a calibrated flatness of < 17 µm.

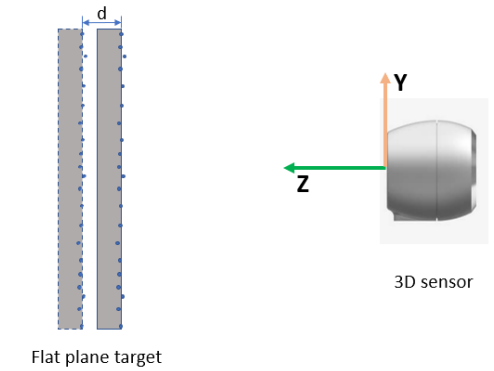


Figure 24: Illustration of a side view of the test involving the flat plane for estimating resolution.

There are three underlying assumptions for this method to be valid [28]:

1. The standard deviation of the samples at both locations are identical;
2. The distribution of the residuals is normal or Gaussian, and
3. There is no correlation between the data in different regions of the plane, i.e., the residuals are random. Note that sometimes the distribution may be Gaussian, but the data has spatial correlations.

For some sensors, the distribution of the samples at different distances can be very different as shown in Figure 25. An F-test[‡‡‡] [29] was conducted to show that not all neighboring datasets have the same standard distribution. The distribution of the residuals can deviate from being Gaussian as shown in Figure 26 and the residuals can be correlated as shown in Figure 27, which are the major drawbacks of this procedure

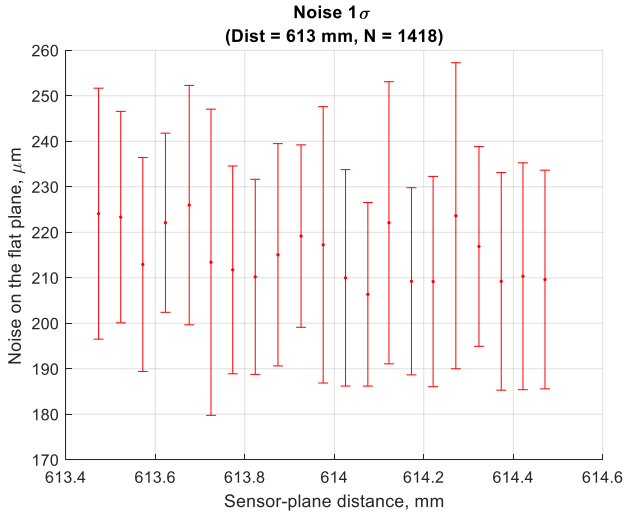


Figure 25: Standard deviation of the residuals from planes at different distances for Sensor R6.

.

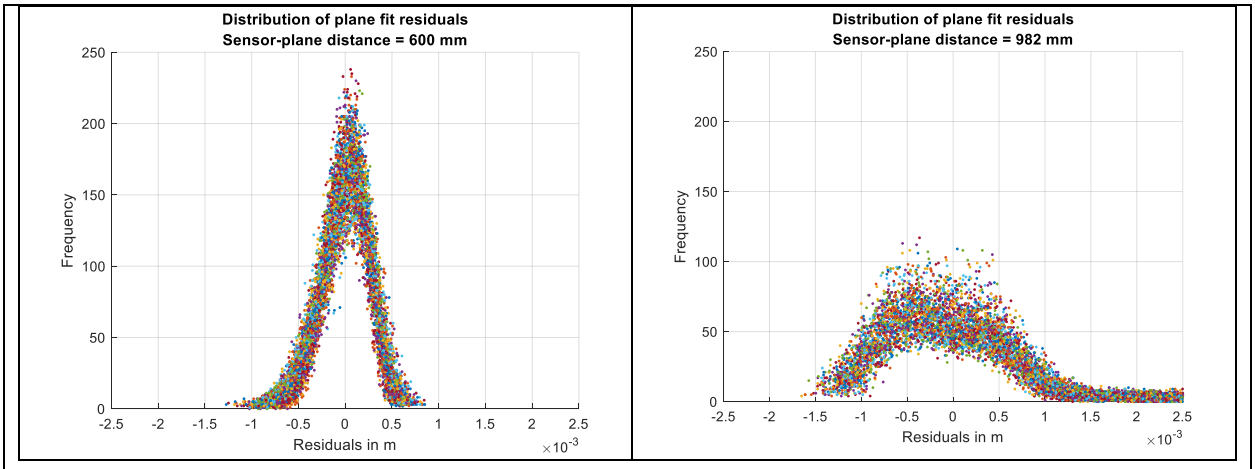


Figure 26: Distribution of the residuals from planes at two distances for Sensor R6 (Different colors represent ≈ 600repeat datasets)

‡‡‡ The two-sample T-test is used to determine if two population means are equal, whereas an F-test is used to test if the variances of two populations are equal.

.

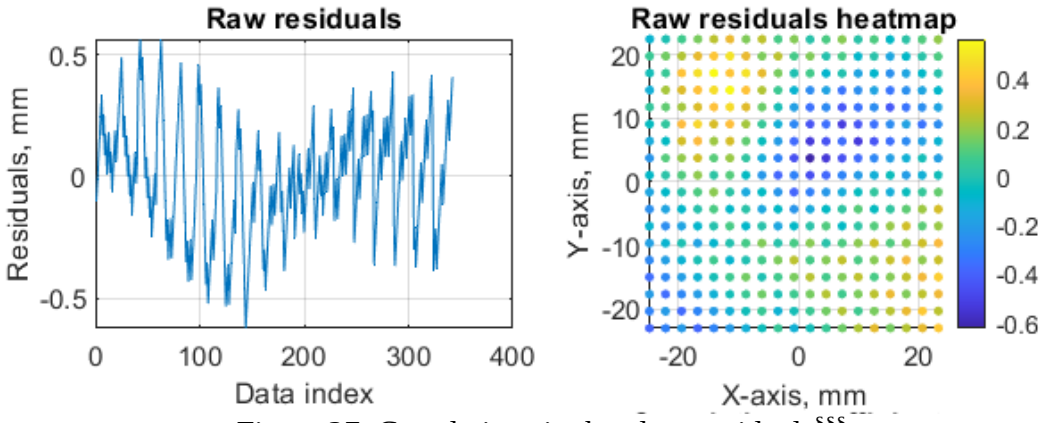


Figure 27: Correlations in the plane residuals[§§§]

3.7 Resolution Metric based on Depth Uncertainty

The definition of depth resolution of a 3D sensor in this work is "the smallest change in depth that causes a perceptible change in the corresponding depth measurement indication". This means that the ability to resolve a target's depth is dependent on the uncertainty in the depth value of that target. To arrive at a metric for depth resolution, it is important to understand the sources of uncertainty in depth evaluation by 3D sensors.

There are several sources of uncertainty in determining the depth using a 3D sensor. Some of these sources are spatial and temporal noise (such as pixel reset noise and shot noise in Time-of-flight sensors [30]), quantization error, sensor pixel grid spacing (for stereoscopic sensors), calibration errors, sensor illumination wavelength, material properties (and thereby reflectance) of the target, distance from the sensor, baseline distance (for triangulation sensors [31]), ambient light intensity/wavelength, and location of the target in the work volume of the sensor. Some of these sources of uncertainty are described below:

3.7.1 Spatial noise

For a flat plane target, spatial noise is the noise observed on the target by fitting the data to a plane (e.g., artifact shown in Figure 23) and calculating the residuals.

3.7.2 Temporal noise

Temporal noise is obtained by taking multiple scans/captures in quick succession, for example, with ~1 second (or less) duration between the successive scans and calculating the variation of the centroids of these datasets. Alternatively, temporal noise could also be calculated per pixel location if Z-depth data is used.

3.7.3 Quantization error

Quantization error is the error due to the representation of continuous depth values by the closest digital value. The reason for quantization errors in 3D sensors is the data precision of the data type used for storing these values and the discrete nature of the sensors/pixels used in some

[§§§] Data index in this figure represents the order in which the data appears in the file and is determined by how the manufacturer exports the data.

systems, such as stereoscopic sensors [32]. In some Time-of-flight 3D sensors, it could also be dependent on the timing electronics. Figure 28 illustrates the possibility of a real point assuming the nearest quantized value. Here the blue points represent the true depth values, and the red point represents the closest quantized value. Figure 29 shows the measured data from one sensor, where the quantization can be observed. Figure 30 shows this effect at multiple distances. Here the mean Z-coordinate data could be a combination of multiple sources of errors, including quantization error and spatial error. It can be observed that there is a significant displacement in the mean value of Z when it is moved from certain positions (e.g., the data at 0.1 mm and 0.15 mm on the X-axis) and the trend reverses after at next few locations, and the pattern repeats to some extent. No additional investigation was conducted to identify the reason for this behavior. If there were no quantization error, the mean of the Z-coordinates is expected to deviate less from the linear fit line and have less systematic effects (like the trend reversal observed).

It can also be observed that at the same location along the X-axis, there are multiple depth values on the Z-axis. These are at different values along the Y-axis not depicted in the figure. The maximum values of the quantization error (Z_q) are the values that the data will assume based on the proximity to the rounded off number representation, which in some cases are integer values.

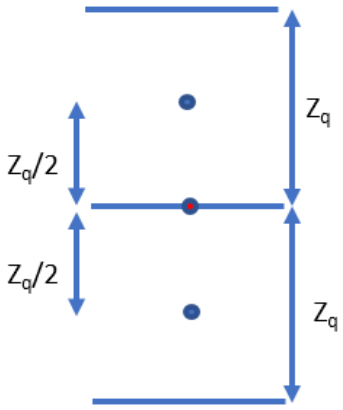


Figure 28: Illustration of quantization of depth values (blue) to their nearest quantized value (red).

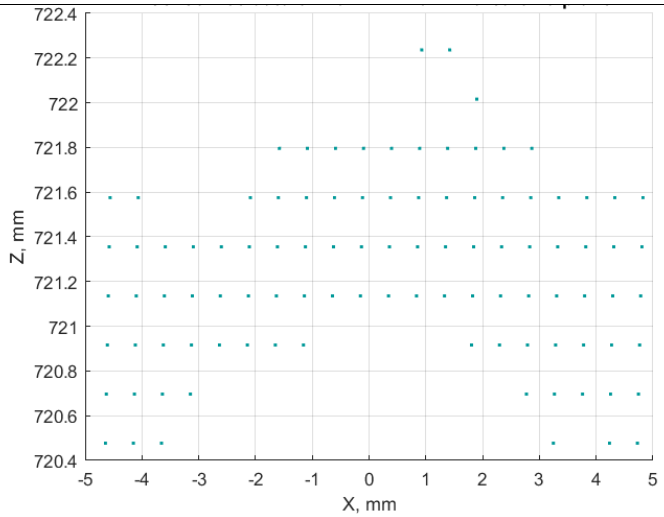


Figure 29: Quantization of depth values to their nearest quantized value that are about 0.22 mm (for Sensor E2 within a 10 mm × 10 mm area).

Because of this quantization error, the depth reported by the sensor will not change linearly with the target's movement. This is because a sensor reports a depth value which is rounded to the nearest quantized number.

3.7.4 Sensor pixel grid spacing for stereoscopic sensors and other similar triangulation sensors.

The uncertainty in depth also decreases with higher sensor resolution as illustrated in Figure 31, and this value remains constant for a sensor where the resolution cannot be changed. The relationships between the depth uncertainty and number of pixels, focal length (f), baseline length (b), and distance (Z) for a specific disparity error (s) are well established for triangulation

sensors (like stereoscopic sensors and structured light sensors) [33, 34] and given in Equation 4. The disparity error (s) (or the correspondence error) depends on the matching algorithm [33] used.

$\delta Z = -\frac{Z^2 s}{fb}$	4

3.7.5 Calibration error

An error in calibration, either due to a limited number of calibration target poses or error in the calibration target itself will result in a depth error. The difference in depth error for small displacements is negligible for the purposes of depth resolution calculation.

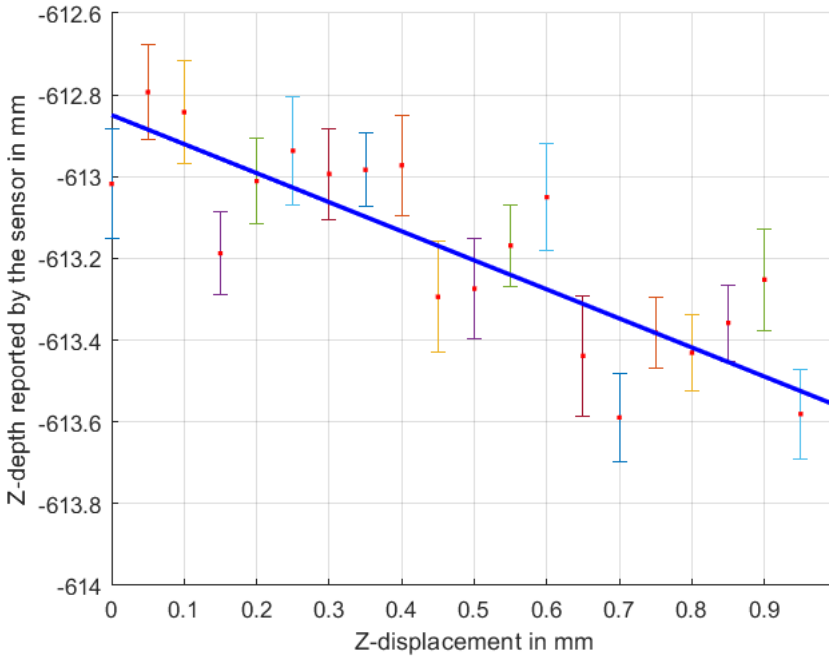


Figure 30: Effect of the quantization error on the Z-coordinate of the centroid (Sensor D3)

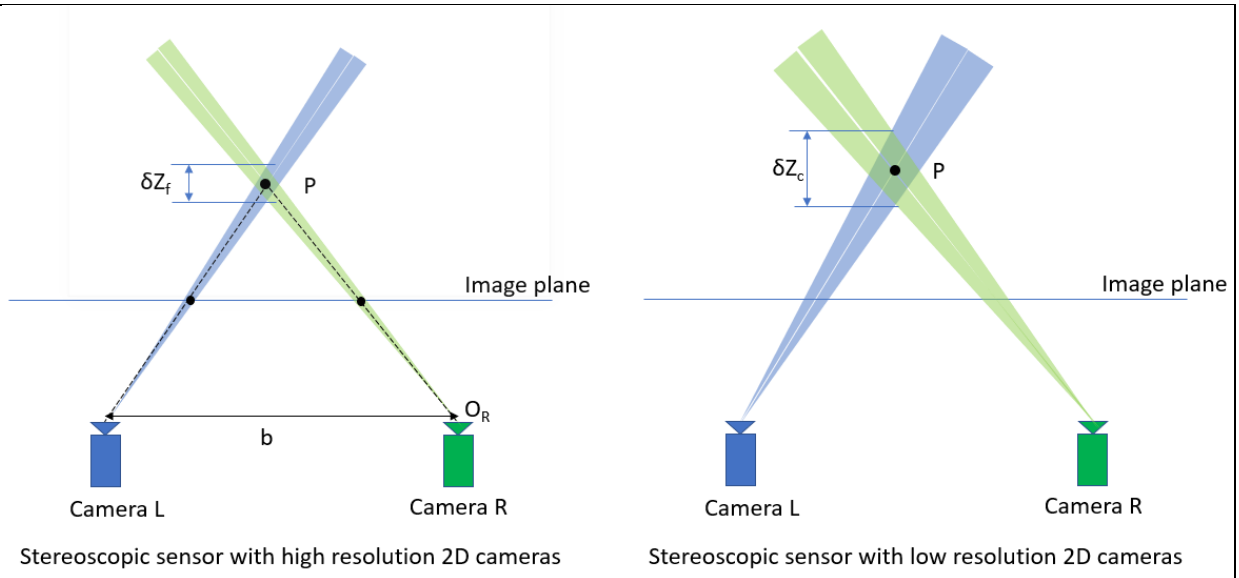


Figure 31: Effect of the sensor resolution on the depth uncertainty.

3.7.6 Effect of other sources of uncertainty

There are many sensors that use a coherent light source (such as an active stereo-sensor) and an incoherent light source (such as a white light structured light sensor) for encoding a surface being scanned. The sensor wavelength is a property of the instrument and affects the performance of the sensor on certain targets (diffuse and specular reflectance of the target's surface). In this work a media blasted aluminum planar target with a matte finish (see Figure 23) was used. It was measured on a spectrophotometer[****] and the reflectance value at 8° incident angle was measured at various wavelengths. Figure 32 shows that reflectance at various sensor wavelengths of interest (from 380 nm to 900 nm).

The location of the target in the field of view of the sensor in all three directions (X, Y, and Z) as well as the ambient light conditions can affect the performance of the system. For the purposes of this work, these values (location and ambient light) were recorded, but not studied. When these properties are held constant, the performance of the sensor varies only due to the sensor noise (spatial and temporal) and the quantization error.

[****] Perkin Elmer Lambda 950 – Courtesy, Heather Patrick, NIST

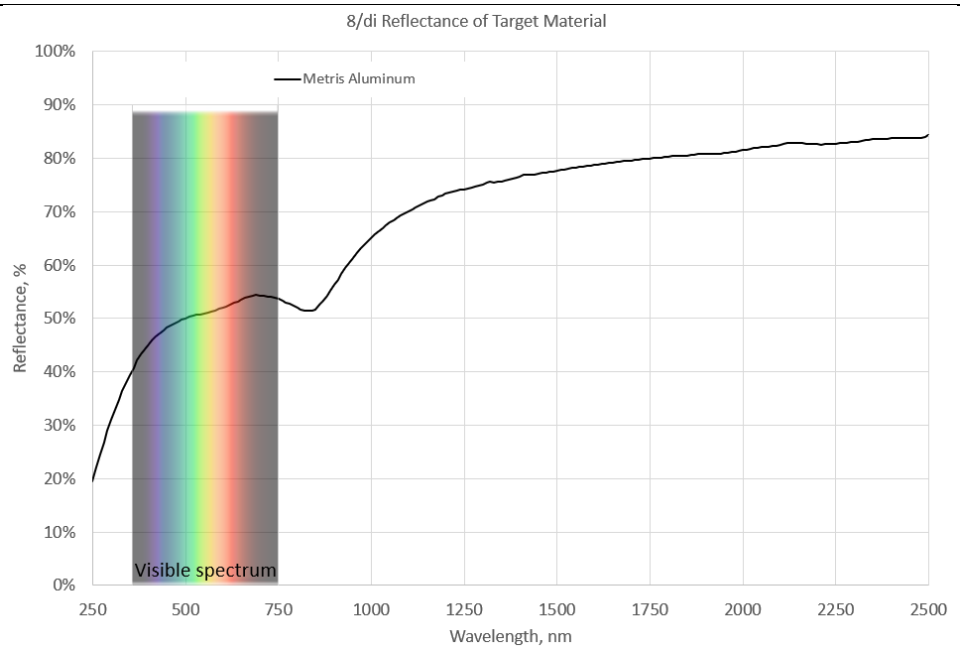


Figure 32: Reflectance of the media blasted aluminum at various wavelengths.

3.7.7 Test procedure

The procedure to conduct a test to evaluate a sensor's depth resolution based on depth uncertainty is given below:

1. Place the target (or the sensor) on a movable stage (coarse and fine) at location #1 in the sensor's working volume along the Z-direction. The setup needs to allow the sensor (or the target) to be moved in the depth (or Z) direction.
2. Obtain N datasets on the flat target ($N \geq 20$), where each dataset corresponds to the point cloud representation of the target.
3. Truncate the data on the target to an area, for example, 50 mm × 50 mm[††††] in the X-Y plane.
4. Calculate the mean of the Z-coordinates for each of the 20 truncated datasets and calculate the standard deviation of these means σ_{zc}. This is the standard uncertainty of the centroid's Z-coordinate and is a measure of the temporal noise
5. For any one of the datasets, calculate the range of the Z-coordinates (Z_r). The selected dataset could be the first of the N datasets.
6. Calculate the number of unique values of the Z-coordinates (Nu) in the above dataset (see section 3.4)
7. The maximum quantization error Z_q is calculated using Equation 5 below

$$Z_q = Z_r/(N_u\text{-}1). \quad (5)$$

8. A point on the flat planar target could exist anywhere between $-0.5Z_q$ and $+0.5Z_q$ and if we assume a uniform distribution, the standard uncertainty due to the quantization error is $0.5Z_q/\sqrt{3} = Z_q/\sqrt{12}$ (see [35]).

[††††] The 50 mm x 50 mm dimension can be changed to reflect the approximate size of the object of interest.

9. The standard uncertainty of locating the Z-coordinate of the centroid at location #1 is given by Equation 6 as the root-sum-square of the individual uncertainties.

$$uc = \sqrt{(\sigma_{zc})^2 + (Z_q/\sqrt{12})^2} \quad (6)$$

When the target moves to location #2 (away from location #1 in the Z-axis) by a distance equal to the yet-to-be-determined resolution of the sensor, the standard deviation of the mean of the Z-coordinates is assumed to be the same value σ_{zc}. The expanded uncertainty of locating the Z-coordinates at both locations is then given by equation 7, where k is the critical value of the t-distribution with degrees of freedom of N-1 for 5% significance level (95% coverage or confidence interval).

$$Uc = k \times \sqrt{2} \times uc \quad (7)$$

The depth (or Z) resolution of the sensor is then the expanded uncertainty of locating the Z-coordinate and is given by equation 8 and is a conservative estimate of the sensor depth resolution.

$$R = k \times \sqrt{2} \times \sqrt{(\sigma_{zc})^2 + (Z_q/\sqrt{12})^2} \quad (8)$$

Note that for some sensors, it is possible to reduce the quantization error value by reducing a parameter (sometimes referred to as the "Depth units"[‡‡‡‡]), thereby lowering the resolution value. However, the effect of the quantization error does not have a 1:1 correspondence (i.e., a 50% reduction in the "Depth unit" does not result in a 50% reduction in the quantization error).

Also, "depth resolution" in this work is dependent on only the temporal noise and the quantization error. Any effect of spatial noise is minimized by averaging. To reduce the effect of spatial noise, the size of the region of interest can be reduced and multiple such regions on the flat target may be evaluated.

4 EXPERIMENTAL SETUP, RESULTS, AND DISCUSSION

Figure 33 shows the experimental setup that was used to test various sensors. The target was a Nikon Reference Plate (media blasted aluminum target, part# 0002829-G) that has a calibrated flatness of less than 17 μm. The sensor was mounted on two linear stages to align the sensor's center with the flat target. The sensor was approximately centered on the target and the sensor's Z-axis and the target's plane normal was nominally parallel[§§§§]. The parallelism was achieved in different ways for different sensors. For some sensors the squareness of the mounting apparatus was used to ensure parallelism. For other sensors, their software suites offered either the depth gradient in the live view (which can be used to indicate parallelism) or displayed the plane's angle with the sensor. The ambient conditions in the laboratory were stable during all the measurements wherein the temperature remained at 20° C ± 1° C and the ambient light was approximately 425 Lux. The data from all the sensors were obtained using each sensors' default settings. Although the performance of these sensors may vary based on these settings, no attempt was made to optimize the settings in the test conditions discussed herein.

[‡‡‡‡] The default "Depth Unit" for Sensor D3 is 1000 microns

[§§§§] The effect of non-parallelism of the target was not studied in this work, but any effect due to non-parallelism is minimized.

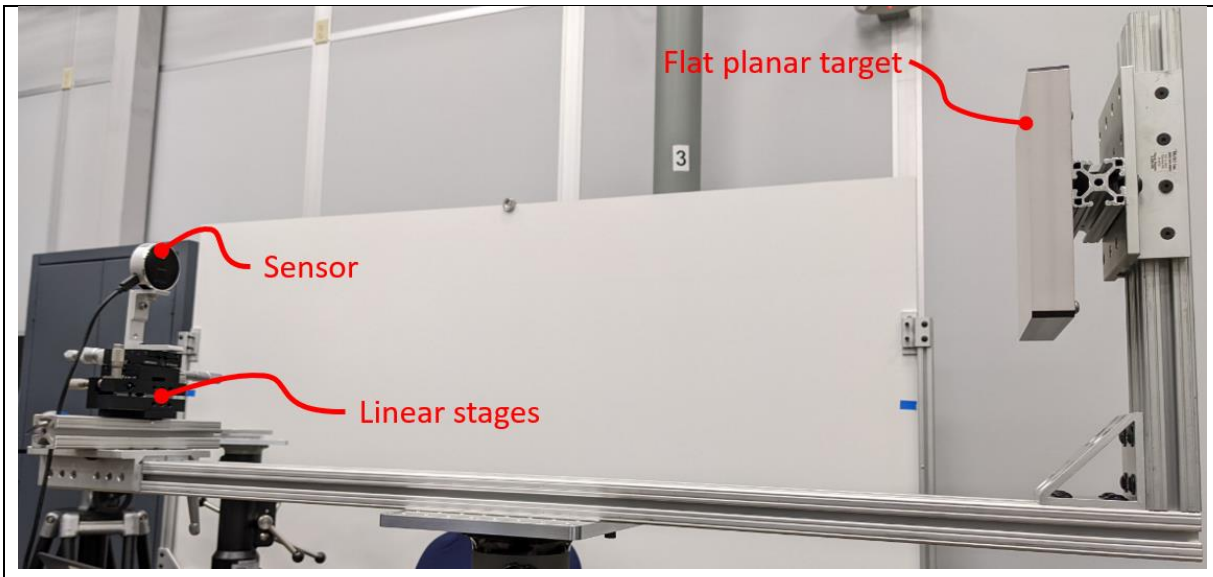
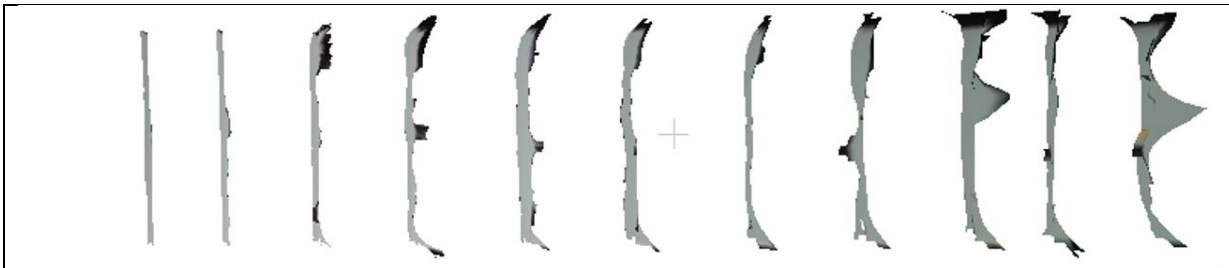


Figure 33: Setup to test various sensors using a flat planar target.



Distance	274.4	345.3	431.4	521.7	627.1	728.2	840.5	944.4	1046.2	1122.0	1211.6
RMS	0.4	0.6	2.7	4.8	5.0	5.0	5.6	5.7	7.5	5.2	8.1

Figure 34: RMS of the residuals from a plane fit of data from Sensor R6 at various distances (values in mm).

It should be noted that for some sensors there were egregious edge effects as shown in Figure 34, so segmentation, which resulted in a relatively lower number of points on the target, was necessary.

Various sensors were used for this work and some relevant specifications of the sensors are given in Table 1. Only some of these sensors offer data as depth images using their software suites directly, but most offer 3D data in PLY (polygon file format [36]) . For that reason, this work only used the 3D data in PLY format.

The data from each sensor was segmented to result in data from a region measuring 50 mm × 50 mm on the target and parallel to the sensor's XY plane (the red area in Figure 35). This was done to exclude any edge data, such as the ones observed in Figure 34.

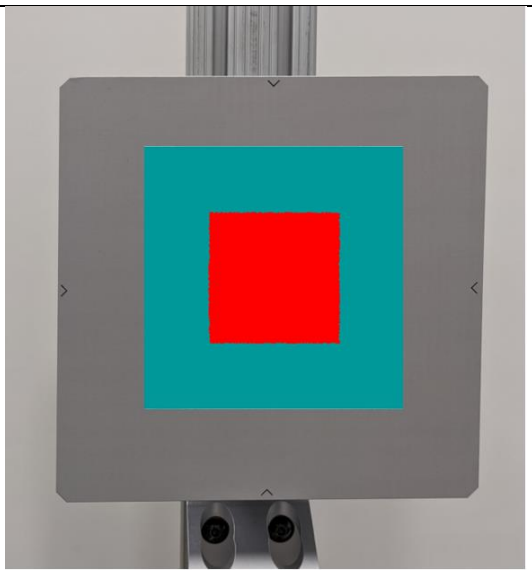


Figure 35: Segmented data (red) from a 3D sensor overlaid on an image of the target.

The data from each sensor was fit to a plane (orthogonal planar fit) and the root-mean-square (RMS) of the residuals from this fit was calculated (and the standard deviations of the residuals are shown in Figure 36).

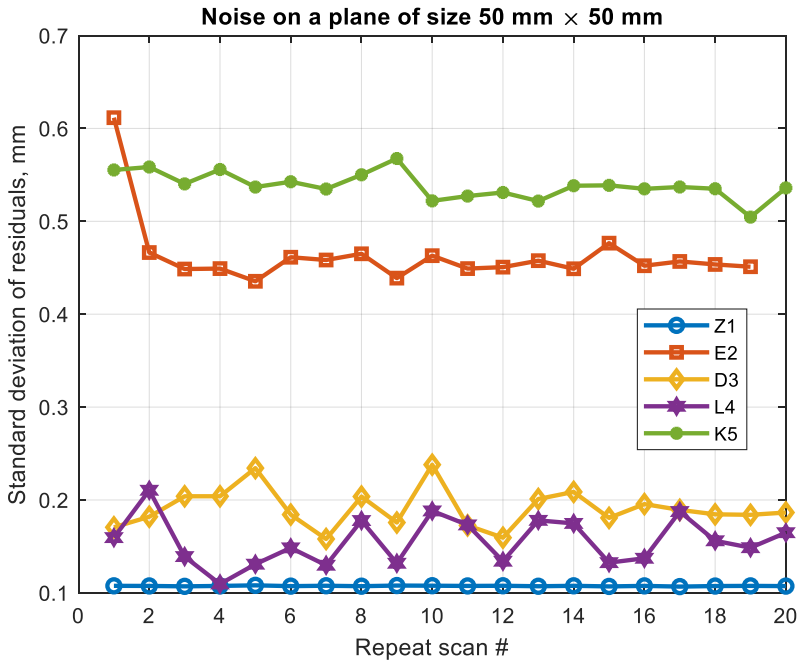


Figure 36: Standard deviations of residuals from plane fits on data from a flat planar target (with a flatness of < 17 μm) captured by five different sensors.

After the quantization errors (Z_q) were found using equation 5 and the standard deviations of the Z- coordinates of the centroids of the segmented data were calculated, the resolution, as

defined in Equation 8, was calculated. The results are given in Table 2. The distances in Table 2 are nominal distances as reported by the sensor and no reference distance measurements were conducted. The distances at which the data was collected was away from the extremities of the sensor range. For Sensor E2, only 19 datasets were collected; however, the coverage factor *k* accounts for this difference when calculating the depth resolution. No data was collected for sensor R6 as it was similar in operating principle as Sensor D3.Note that for sensor Z1, the quantization error was < 0.001 mm.

The value of resolution (*R*) calculated by equation *8* will decrease as the number of datasets collected (repeats) increase and vice versa. This is due to the calculation of *k* which is based on the number of datasets. For example, if acquired datasets were 100, the value of *R* will be 5.2% lower than when the acquired datasets were 20.

Table 2: Noise, quantization error, and resolution of various sensors on an area of 50 mm × 50 mm

	Dist. (mm)	Repeats	Avg. # of pts	Quantization error (mm)	Mean of std. dev. of plane fit residuals (mm)	Std. dev. of plane Z-coordinate of centroids (mm)	Resolution (mm)
Sensor Z1	807.3	20	29043	0.000	0.107	0.016	0.049
Sensor E2	721.3	19	10332	0.220	0.440	0.016	0.194
Sensor D3	507.9	20	501	0.250	0.191	0.441	1.323
Sensor L4	612.9	20	342	0.250	0.156	0.566	1.688
Sensor K5	398.3	20	3772	1.000	0.538	0.053	0.869

The segmentation area was reduced to 25 mm × 25 mm and then to 10 mm × 10 mm and the parameters were calculated again. These new calculations are given in Table 3 and Table 4, respectively.

Table 3: Noise, quantization error, and resolution of various sensors on an area of 25 mm × 25 mm

	Dist. (mm)	Repeats	Avg. # of pts	Quantization error (mm)	Mean of std. dev. of plane fit residuals (mm)	Std. dev. of plane Z-coordinate of centroids (mm)	Resolution (mm)
Sensor Z1	807.3	20	7265	0.000	0.106	0.017	0.049
Sensor E2	721.3	19	2582	0.220	0.436	0.039	0.221
Sensor D3	507.9	20	132	0.250	0.111	0.463	1.388
Sensor L4	612.9	20	90	0.250	0.107	0.602	1.796
Sensor K5	398.3	20	929	1.000	0.465	0.055	0.870

Table 4: Noise, quantization error, and resolution of various sensors on an area of 10 mm × 10 mm

	Dist. (mm)	Repeats	Avg. # of pts	Quantization error (mm)	Mean of std. dev. of plane fit residuals (mm)	Std. dev. of plane Z-coordinate of centroids (mm)	Resolution (mm)
Sensor Z1	807.3	20	1156	0.001	0.106	0.018	0.052
Sensor E2	721.3	19	400	0.220	0.422	0.158	0.504
Sensor D3	507.9	20	16	0.250	0.072	0.458	1.373
Sensor L4	612.9	20	16	0.250	0.075	0.629	1.873

Sensor K5	398.3	20	150	1.000	0.367	0.118	0.923

The resolution for Sensor Z1 for the three segmented areas varied by only 3 μm. For all the segmented regions the quantization error was also nearly consistent, and the temporal noise (standard deviation of the Z-coordinate of the centroids) varied slightly. This could be attributed to higher point density from this sensor and indicates that this sensor could be used for identifying targets smaller than 10 mm x 10 mm.

For some sensors like Z1, E2, and K5 the spatial noise (noise on the plane) was higher than the temporal noise, whereas the reverse was true for sensors D3, L4 and D6. It is not clear why this is the case. Sensors with global shutters typically have higher noise than rolling shutter sensors [37], but that generalization cannot be made here as the sensor sizes and pixel sizes are different among all sensors. The frame rate of all these sensors are fast enough (> 10 frames per second), and the data from Sensor Z1 and Sensor E2 were obtained at their highest possible acquisition rates of about 0.1 frames per second.

One of the overarching questions is whether this method can be used to distinguish a small feature (such as a 4 mm rivet on an automotive part). To address that question, this test could be modified to use data on the plane that is of size 4 mm × 4 mm in the XY plane. This comes with a caveat that the rivet may not have a flat surface and the resolution metric in this case will be an optimistic estimate.

5 CONCLUSION

A variety of artifacts and methods for evaluating the depth resolution of 3D imaging systems were tested. Qualitative evaluation with various artifacts presented in this work was useful for quick understanding of how the sensors work, but it did not offer a concise way to understand the sensors objectively. This is because the factors that influence such a sensor's performance are also dependent on the target size, shape, color, geometry, and the environment in which it is being tested.

Multiple methods to evaluate sensors quantitatively were introduced in this work using a simple geometry – a flat planar target. This was done to reduce the influence of the target-induced errors in evaluating the sensor. Of these methods, the method described in Section 3.7 (using the uncertainty quantification method) has garnered more consensus in the ASTM resolution task group.

6 ACKNOWLEDGEMENTS

The authors would like to thank members of the ASTM E57 work item WK73176 for their valuable contributions and feedback that went into this work. The authors want to specifically thank Adam Pintar, Helen Qiao, and Ya-Shian Li-Baboud of NIST for their technical feedback and manuscript reviews, Heather Patrick of NIST for the reflectance measurements, Felix Thouin (formerly of Airy 3D) for his artifact designs, and John Sweetser of Intel RealSense Group for his in-depth feedback and knowledge about sensor design and evaluation.

7 REFERENCES

[1] Saidi, K., Cheok, G., Qiao, H., Horst, J., Franaszek, M., "Proceedings of the ASTM E57 Workshop on Standards for 3D Perception Systems for Robotic Assembly Applications December 2 – 3, 2019", Gaithersburg, MD https://nvlpubs.nist.gov/nistpubs/ams/NIST.AMS.100-33.pdf

[2] ASTM Standard E1856-13 20021, "Standard Guide for Evaluating Computerized Data Acquisition Systems Used to Acquire Data from Universal Testing Machines" ASTM International, West Conshohoken, PA, www.astm.org. DOI: 10.1520/E1856-13R21

[3] ASTM Standard D7145-05(2015), "Standard Guide for Measurement of Atmospheric Wind and Turbulence Profiles by Acoustic Means" ASTM International, West Conshohoken, PA, www.astm.org. DOI: 10.1520/D7145-05R15

[4] ASTM Standard F1166-21, "Standard Practice for Human Engineering Design for Marine Systems, Equipment, and Facilities" ASTM International, West Conshohoken, PA, www.astm.org. DOI: 10.1520/F1166-21

[5] ASTM Standard E2007-10(2016), "Standard Guide for Computed Radiography" ASTM International, West Conshohoken, PA, www.astm.org.

[6] ASTM Standard E2544-11A(2019)e1, "Standard Terminology for Three-Dimensional (3-D) Imaging Systems," ASTM International, West Conshohoken, PA, www.astm.org.

[7] Definition of resolution. VIM: International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, ISO/IEC Guide 99, 2017. (JCGM 200) https://jcgm.bipm.org/vim/en/4.14.html

[8] Virginia Department of Transportation: Laser Scanning and LiDAR Applications https://www.virginiadot.org/business/resources/LocDes/SurveyManual/Chapter9.pdf

[9] Stan Users Guide: Arithmetic Precision: https://mc-stan.org/docs/2_22/stan-users-guide/arithmetic-precision.html.

[10] A Comprehensive List of 3D Sensors Commonly Leveraged in ROS Development https://rosindustrial.org/3d-camera-survey

[11] VDI-Standard: VDI/VDE 2634 Blatt 2: https://www.vdi.eu/guidelines/vdivde_2634_blatt_2-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/

[12] VDI-Standard: VDI/VDE 2634 Blatt 3: https://www.vdi.eu/guidelines/vdivde_2634_blatt_3-optische_3_d_messsysteme_bildgebende_systeme_mit_flaechenhafter_antastung/

[13] ISO 10360-13:2021 Geometrical product specifications (GPS) — Acceptance and reverification tests for coordinate measuring systems (CMS) — Part 13: Optical 3D CMS https://www.iso.org/standard/74957.html

[14] 3D scanner accuracy and resolution: What you need to know (aniwaa.com) https://www.aniwaa.com/insight/3d-scanners/3d-scanner-accuracy-resolution-basics/

[15] Gomez, M., Vaccaro, C., Maness, C., Lawrence, C., Murray, C., Vargas, L., Caskey G., Evans, C., Schmitz, T., Mechanical Engineering and Engineering Science https://mtrc.utk.edu/wp-content/uploads/sites/54/2019/09/CPM_ext_abstract.pdf

[16] Neupane, C.; Koirala, A.; Wang, Z.; Walsh, K.B. Evaluation of Depth Cameras for Use in Fruit Localization and Sizing: Finding a Successor to Kinect v2. Agronomy 2021, 11, 1780. https://doi.org/10.3390/agronomy11091780

[17] Giancola et al., A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies, Springer Briefs in Computer Science, https://doi.org/10.1007/978-3-319-91761-0_1

[18] H. Haggag, M. Hossny, D. Filippidis, D. Creighton, S. Nahavandi and V. Puri, "Measuring depth accuracy in RGBD cameras," *2013, 7th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2013, pp. 1-7, https://doi.org/10.1109/ICSPCS.2013.6723971

[19] Cheok, G., NISTIR 7266: Proceedings of the 2nd NIST LADAR Performance Evaluation Workshop – March 15 - 16, 2005 https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nistir7266.pdf

[20] G.S. Cheok, A M. Lytle and K.S. Saidi. Status of the NIST 3D imaging system performance evaluation facility. April 17-21 2006. SPIE Defense and Security Symposium. Paper 6214-17 https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=860649

[21] Wang, D. et. al., Measurement of size error in industrial CT system with Calotte cube – Proceedings Volume 9446, Ninth International Symposium on Precision Engineering Measurement and Instrumentation; 94460Q (2015) https://doi.org/10.1117/12.2180600

[22]Muller, P., Coordinate Metrology by Traceable Computed Tomography, PhD Thesis, 2012 https://backend.orbit.dtu.dk/ws/portalfiles/portal/77770178/PhD_thesis_Pavel_M_ller..PDF

[23] David M. Sheen, R. Trevor Clark, and Maurio Grando "Depth resolution evaluation for an active wideband millimeter-wave imaging system", Proc. SPIE 11745, Passive and Active Millimeter-Wave Imaging XXIV, 117450M (12 April 2021); https://doi.org/10.1117/12.2587201

[24] ISO 10360-13:2021 Geometrical product specifications (GPS) — Acceptance and reverification tests for coordinate measuring systems (CMS) — Part 13: Optical 3D CMS https://www.iso.org/standard/74957.html

[25] Dury, Martin & Brown, Stephen & McCarthy, M. & Woodward, Sean. (2015). 3D Optical Scanner Dimensional Verification Facility at the NPL's "National FreeForm Centre". https://www.euspen.eu/knowledge-base/LAMP15119.pdf

[26] Dury, M. R. et al. “3D Optical Scanner Dimensional Verification Facility at the NPL’s “National FreeForm Centre”.” (2015). https://www.euspen.eu/knowledge-base/LAMP15119.pdf

[27] Two-Sample t-Test for Equal Means https://www.itl.nist.gov/div898/handbook/eda/section3/eda353.htm

[28] Diez, D., Cetinkaya-Rundel, M., Barr, C., OpenIntro Statistics, Fourth Edition https://www.openintro.org/book/os/

[29] Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press.

[30] Introduction to the Time-of-Flight (ToF) System Design 2014 https://www.ti.com/lit/ug/sbau219d/sbau219d.pdf

[31] Beraldin, J.A., "Basic theory on surface measurement uncertainty of 3D imaging systems", Proc. SPIE 7239, Three-Dimensional Imaging Metrology, 723902 (19 January 2009); https://doi.org/10.1117/12.804700

[32] Rodríguez, J.J., & Aggarwal, J.K. (1988). Quantization error in stereo imaging. *Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition*, 153-158. https://doi.org/10.1109/CVPR.1988.196229

[33] D. Gallup, J. Frahm, P. Mordohai and M. Pollefeys, "Variable baseline/resolution stereo," *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8, doi: 10.1109/CVPR.2008.4587671. https://people.inf.ethz.ch/pomarc/pubs/GallupCVPR08.pdf

[34] Best-Known-Methods for Tuning Intel® RealSense™ D400 Depth Cameras for Best Performance https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/BKMs_Tuning_RealSense_D4xx_Cam.pdf

[35] Possolo, A., NIST Technical Note 1900: Simple Guide for Evaluating and Expressing the Uncertainty of NIST Measurement Results http://dx.doi.org/10.6028/NIST.TN.1900

[36] Polygon File Format https://www.loc.gov/preservation/digital/formats/fdd/fdd000501.shtml

[37] Rolling vs Global Shutter https://www.photometrics.com/learn/white-papers/rolling-vs-global-shutter

8 SUPPLEMENTARY INFORMATION

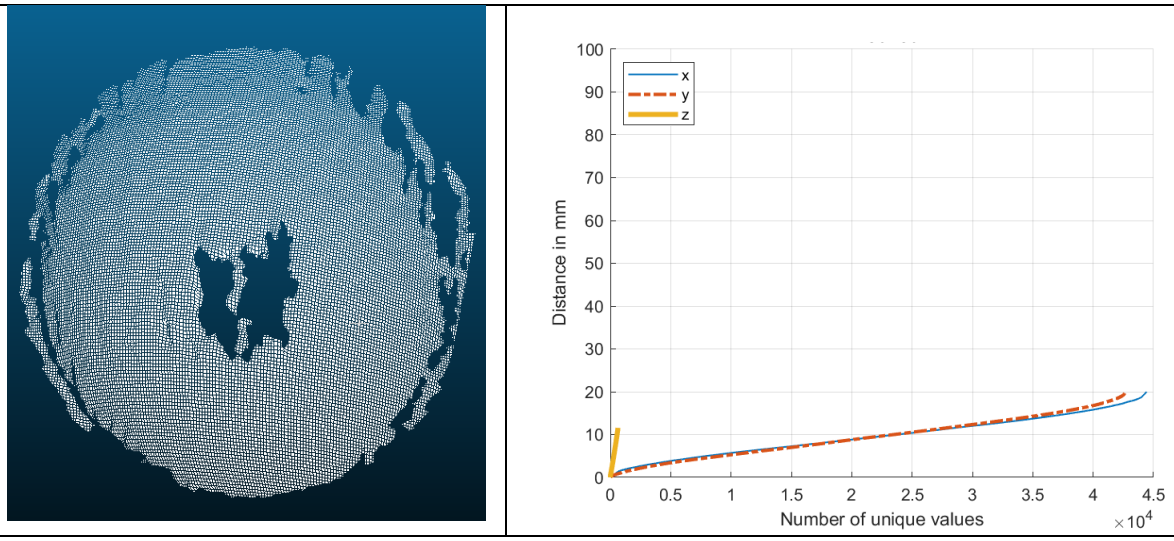


Figure 37: Sensor E2: 3D scan and the number of unique values.

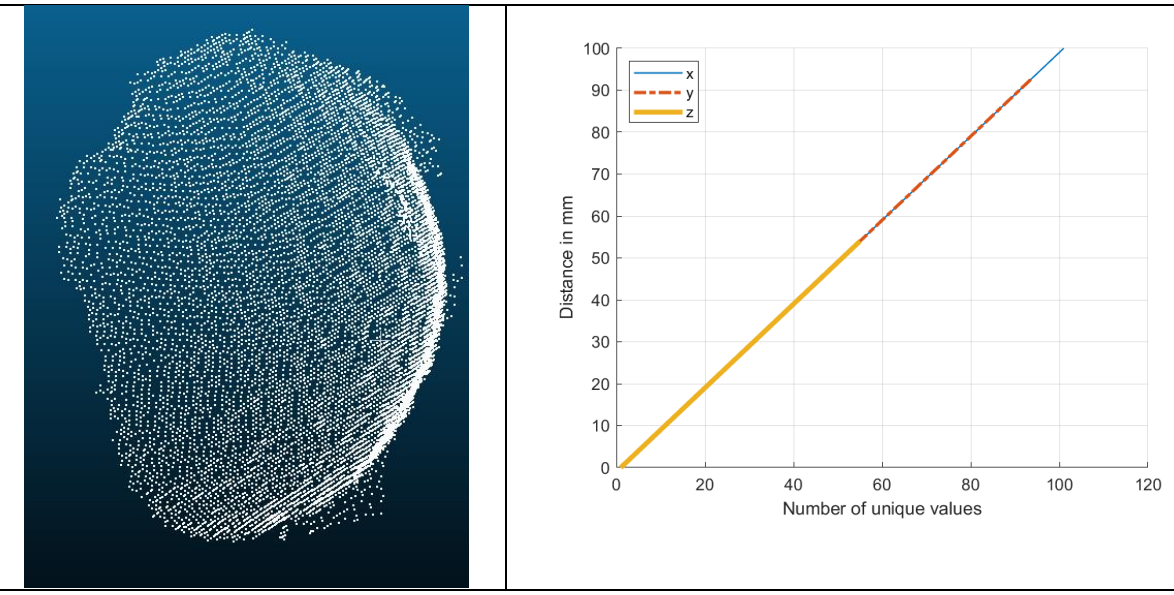


Figure 38: Sensor K5: 3D scan and the number of unique values.

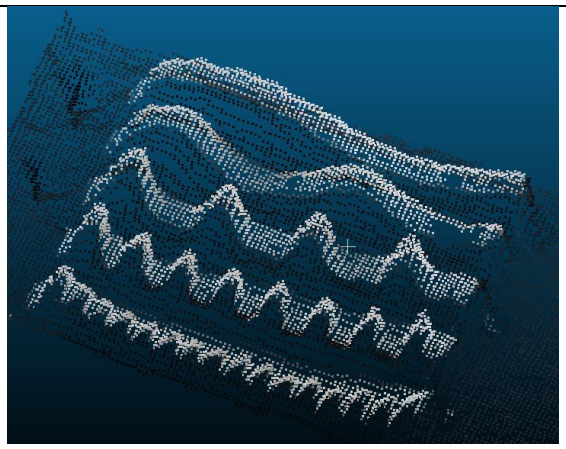


Figure 39: Scan data from Sensor K5

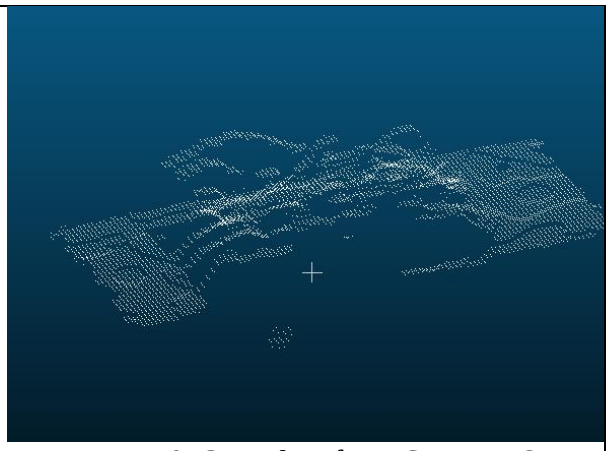


Figure 40: Scan data from Sensor R6

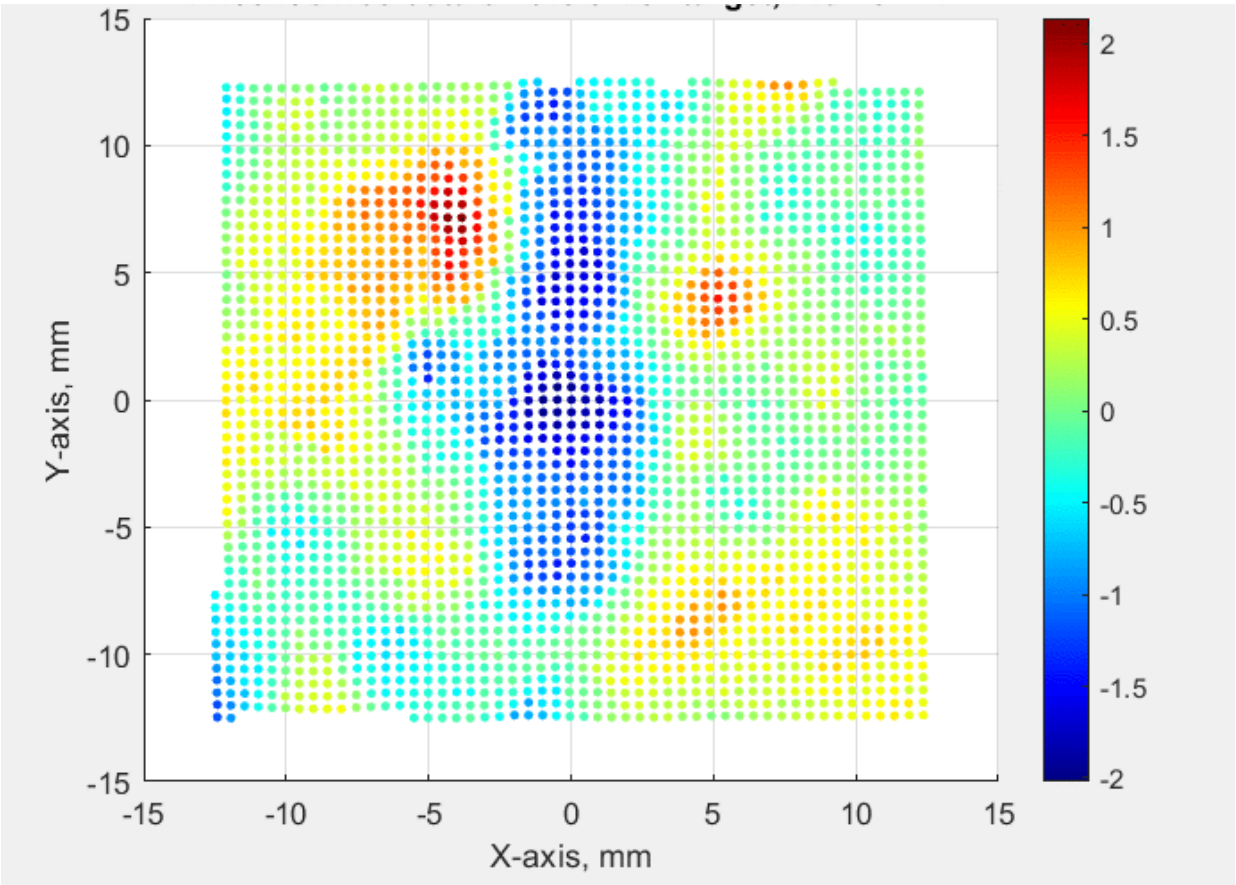


Figure 41: Correlation of residuals from a plane fit for Sensor E2

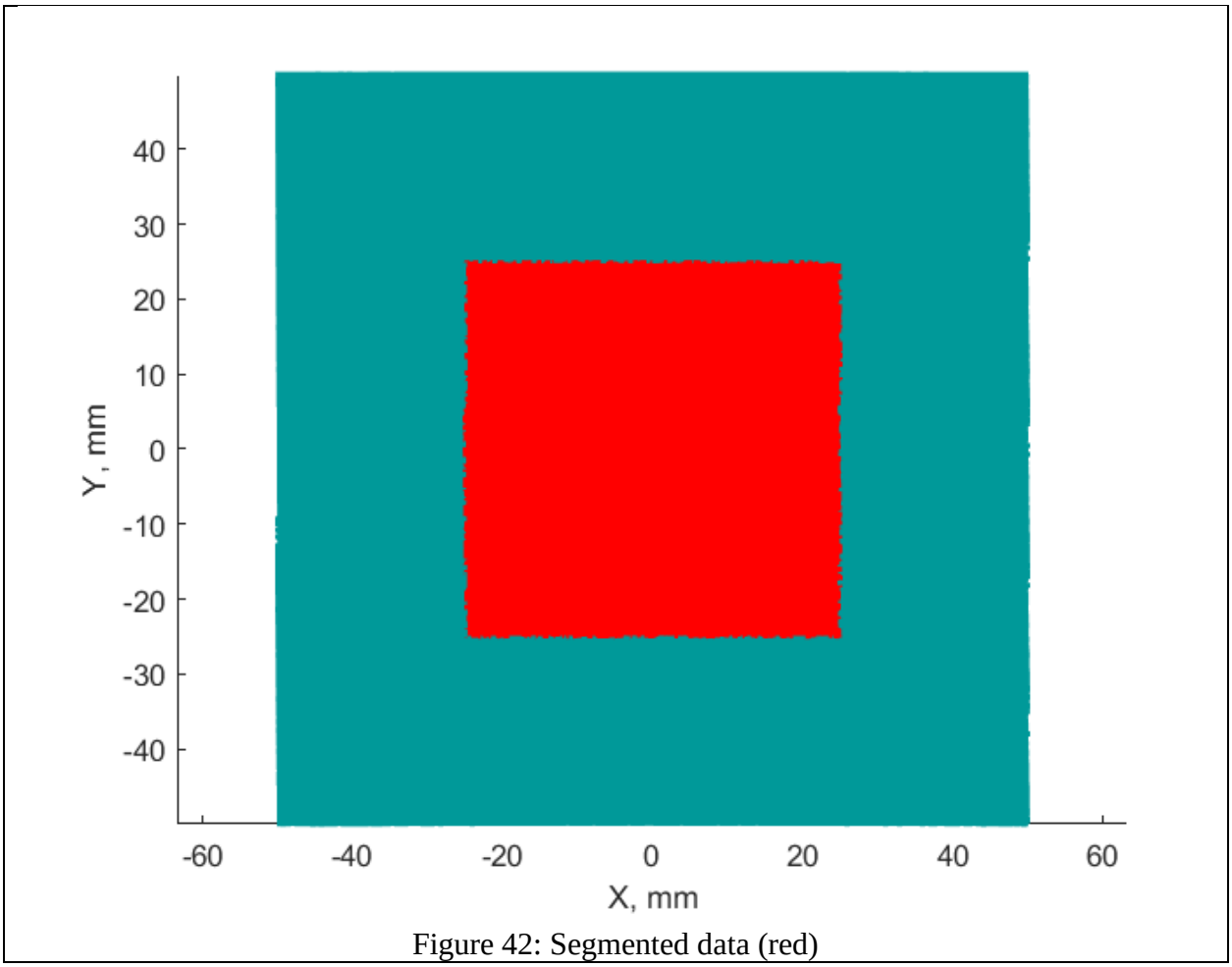


Figure 42: Segmented data (red)

Table 5: Commands in various software to determine *k* for 95% coverage (α = 0.05) and N = 20		
Software	**Command/functions/code**	***k***
MATLAB	alpha = 0.05; N = 20; k = tinv(1-alpha/2,N-1)	2.093024
Octave	pkg load nan; alpha = 0.05; N = 20; k = tinv(1-alpha/2,N-1)	2.093024
Python	from scipy import stats; alpha=0.05; N=20; k=stats.t.ppf(1-alpha/2, N-1); k	2.093024
R	alpha = 0.05; N=20; k = qt(p=alpha/2, df=N-1, lower.tail=FALSE); k	2.093024
Excel	=TINV(0.05,20-1)	2.093024