BOOTSTRAP METRIC FOR QUANTIFYING THE DEPTH RESOLUTION OF 3D SENSORS

Prem Rachakonda[1*], Hariharan Iyer[2], Marek Franaszek[1], Kamel Saidi[1]

[1]Intelligent Systems Division,
National Institute of Standards and Technology,
Gaithersburg, MD, USA.

[2]Statistical Engineering Division,
National Institute of Standards and Technology,
Gaithersburg, MD, USA.

ABSTRACT: Depth resolution of a 3D imaging system or 3D sensor is defined as the smallest change in physical depth that causes a detectable change in the corresponding measured or derived depth. The ability of a sensor to resolve depth on a particular target is correlated with the sensor noise on that target. This work presents a novel metric for calculating the depth resolution of a 3D sensor.

KEYWORDS: 3D imaging systems, sensor parameters, depth resolution, Bootstrapping, standards, ASTM E57

1 INTRODUCTION

The adoption of 3D sensors is growing rapidly across various engineering fields, such as manufacturing automation, robotic bin picking, assembly, reverse engineering, and inspection. One of the key parameters that defines a 3D sensor's performance is depth resolution, which is often undefined or poorly defined. This is mainly because the depth resolution of a 3D sensor is influenced by multiple factors, including sensor noise, sensing principle, target characteristics (e.g., reflectance, texture, and geometry), processing algorithm, and test procedures.

Building on the previous research presented in 2023 [1], the authors have investigated alternative methods to quantify depth resolution to contribute to the ASTM WK73176 [2] working group. This work presents an update to those efforts by introducing a novel metric based on the Bootstrap method [3].

2 BACKGROUND OF THIS WORK

The depth direction of a 3D imaging system or sensor is defined as the direction that is perpendicular to the sensor's reference plane, as defined by the manufacturer, and passes through the sensor's origin. This definition applies to a variety of sensors, including depth cameras, stereoscopic sensors, and scanning or spinning lidars. For sensors that use a spherical coordinate system (such as the spinning lidars), the reference plane can be arbitrary but must still pass through the sensor's origin.

The depth resolution of a 3D imaging system is defined as the smallest change in physical depth that causes a detectable change in the corresponding measured or derived depth. Several stakeholders recommended the need for such a parameter during a NIST workshop [4]. Subsequently, an ASTM work item, WK73176, was initiated to develop a procedure to calculate depth resolution. It is titled "Standard Test Methods for Determination of a 3D Perception Systems Point Wise Spatial Resolution[†]". The purpose of such a standard is to reduce confusion among

[*] Corresponding author: prem.rachakonda@nist.gov

[†] Note that the scope of the standard has been limited to depth direction only and, as such, will not address resolution in the non-depth direction.

end-users, integrators, and researchers about the performance of these systems, enable apples-to-apples comparison, and help end-users to make an informed choice.

Rachakonda et al. [1] described a metric for calculating depth resolution, and an uncertainty calculation for this metric was also developed [5]. However, some 3D sensors present unique challenges due to the correlation between depth data and increased sensor noise, as well as significant quantization (data binning). Data binning introduces nonlinearity, violating standard Gaussian assumptions. That is, the data cannot be described by the typical assumptions of conventional statistical methods, indicating a non-parametric characteristic. To address these characteristics, we explored various statistical methods through extensive experimentation and developed a novel metric to calculate depth resolution. This new depth resolution (R_B) metric is based on measuring the displacement of a planar target in finite steps and calculating the distance at which the datasets differ statistically. The setup and the calculation are discussed next.

3 SETUP TO CALCULATE DEPTH RESOLUTION

The setup for the proposed metrics is similar to that described in [1] and is reproduced here. Figure 1 shows an illustration of the setup with a target placed in front of a sensor so that the sensor's reference plane and the target's front plane are parallel.

A media-blasted aluminum flat planar artifact with a calibrated flatness of less than 17 µm was used as the target, offering a near-Lambertian surface. The sensor was mounted on two linear stages to manually align the sensor with the flat target. The sensor was approximately centered on the target, and the sensor's reference plane and the target's flat surface were aligned to be nominally parallel.[‡] 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 either displayed the depth gradient in the live view (which can indicate parallelism) or the plane's angle relative to the sensor.

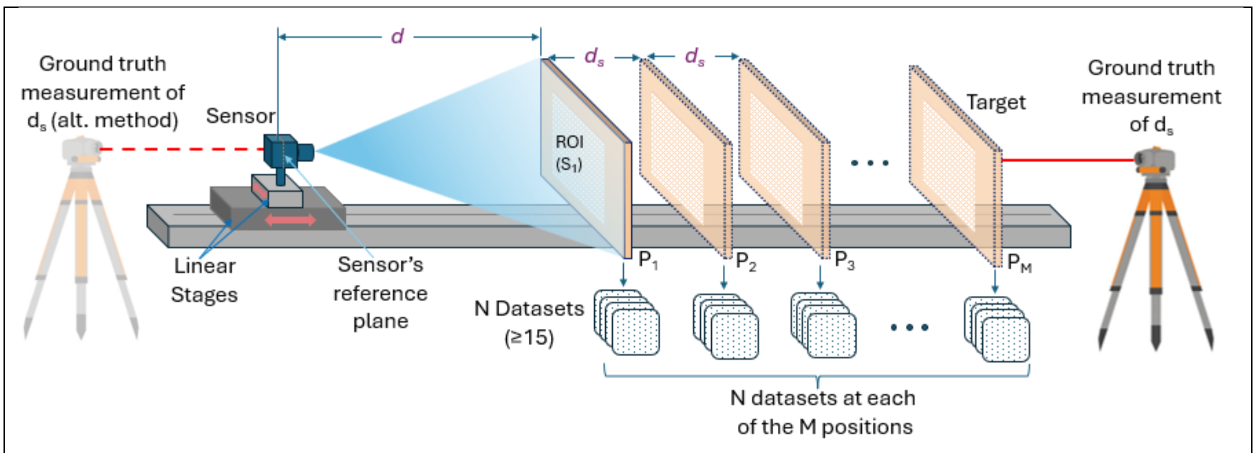

Figure 1: Setup to test various sensors using a flat planar target (not to scale)

The ambient conditions in the laboratory remained stable throughout all measurements, with the temperature at 20 °C ± 1 °C and the ambient light at approximately 425 Lux. Data from the sensor was obtained using the sensor's default settings. Although sensor performance may vary

‡ To implement this method, the end-user must either identify a plane on the sensor housing that is parallel to the sensor's reference plane or obtain this information from the manufacturer. The effect of the target's non-parallelism was not studied in this work.

with their settings [6], no attempt was made to optimize them under the test conditions discussed here.

A critical aspect of these measurements is to measure the sensor's displacement with low uncertainty. This could be achieved using a translation stage with built-in optical encoders, a pre-calibrated stage, or an apparatus that uses ground-truth instruments such as a laser tracker or laser radar to measure the displacement of the target or the stage.

4 THE BOOTSTRAP METHOD FOR CALCULATING RESOLUTION

The method described by Rachakonda et al. [1], though convenient, results in a very conservative estimate of the sensor's ability to distinguish between two sets of depth values at different distances. The current investigation explored several statistical techniques, primarily because the sensor data was non-parametric. One such metric is the Bootstrap metric and is described next.

4.1 Depth resolution based on the Bootstrap method:

The depth resolution based on the Bootstrap method is determined by measuring the displacement of the translation stage that moves the sensor relative to the target[§]. This method identifies the displacement between target positions at which the datasets become statistically distinguishable. The following steps outline the calculation of this metric.

1. Set up and estimate the noise:
    a. Target Placement: Position the target at a distance '*d*' from the sensor, ensuring that the target's front plane is perpendicular to the sensor's optical axis, or parallel to the sensor's reference plane.
    b. Data Acquisition: Collect a minimum of 15 datasets ($N \geq 15$) from the sensor to ensure reliable statistical analysis.
    c. Data Segmentation: Segment each of the collected datasets to isolate a region of interest (S_1) on the target, excluding any edge points. The size of the segmented region (e.g., 10 mm × 10 mm or 50 mm × 50 mm) is user-defined, but it should not include any edge points. Further, an additional condition can be imposed on the segmented data to ensure a minimum number of points (e.g., 30 per dataset).
    d. Standard Deviation Calculation: Compute the standard deviation (σ_1) of the Z-coordinates of the segmented point cloud S_1 from the first dataset.
2. Calculate the stepping distance:
    a. Calculate the stepping distance (d_s) for the target movement based on the standard deviation (σ_1) obtained earlier. A possible approach is to set d_s to a fraction of σ_1, such as $d_s = t \times \sigma_1$ (where t can be a small number, say 0.4). Another possible approach is to calculate d_s based on the uncertainty of the ground truth measurement method. The choice of d_s has further implications.
        i. If d_s is too small, it may be at the limit of the translation stage's stepping capability, or the capability of the method used to measure the movement of the stage (ground truth measurement). In such a case, t must be increased.
        ii. If d_s is too large, it will result in a value for depth resolution that is larger than the sensor's capability. In such a case, t must be decreased.

[§] Some sensors output negative depth values with the zero-point being at the sensor origin; thus, for consistency, all the values need to be converted to positive depth values.

iii. In either case, the value of $t = 0.4$ is a reasonable starting point.

3. Capture data and segment:
   a. Now, capture a minimum of N datasets ($N \geq 15$) from the sensor when the target is at a distance 'd', and denote it as position P_1.
   b. Move the target away from the sensor by a distance d_s
   c. Repeat steps 3a and 3b M-1 times ($M \geq 20$).
   d. At the end of this step, $N \times M$ datasets would have been captured.
   e. Segment each of the datasets using the procedure in 1.c

Algorithm 1: Pseudo code to calculate the depth resolution

```
FUNCTION bootstrap_resolution(ZC, ds, dREF)
//input: //Arrays use 1-based indexing
//ZC – array of mean of segmented data at each successive distance. ZC is of size M×N ;
//ds - stepping distance;
//dREF – array of distances measured using a ground truth instrument (alternative method); dREF is of size  M×1
  // --- Initialization ---
  nBoot ← 1000             // Number of bootstrap iterations
  alpha ← 0.05             // Significance level for hypothesis testing
  M ← size(ZC,1)           // Total number of target positions and data segments
  reference ← ZC[1]        // First segment used as the baseline
  H ← array of size M      // Stores binary results (0: same, 1: different)

  // Bootstrap Significance Testing
  FOR i FROM 1 TO M DO
    comparison ← ZC[i]
    boot_diffs ← empty list of size nBoot

    FOR b FROM 1 TO nBoot DO
      sample1 ← Resample(reference) with replacement
      sample2 ← Resample(comparison) with replacement
      boot_diffs[b] ← mean(sample2) - mean(sample1) // ΔZBi,bootstrap distribution of the difference in depth
    END FOR

    p_value ← proportion of boot_diffs where (value <= 0) // Calculate p-value: proportion of values ≤ 0
    IF p_value <= alpha THEN
      H[i] ← 1   // Datasets are statistically different ; H is an array with one-based index
    ELSE
      H[i] ← 0   // Datasets are statistically the same
    END IF
  END FOR

// Address edge cases, failures and calculate resolution
  last_zero_index ← find last index i where H[i] == 0
  Resolution ← NaN //Initialize
  IF last_zero_index is EMPTY THEN
    RAISE ERROR "Resolution cannot be calculated: Unknown error"
  ELSE IF sum(H) == 0 THEN
    RAISE ERROR "Displacement step size (ds) is too small: No difference detected"
  ELSE IF last_zero_index == 1 THEN
    RAISE ERROR "Displacement step size (ds) is too large"
  ELSE IF (M - last_zero_index ) <3 THEN //Need at least 3 '1's after the last '0'
    RAISE ERROR "Displacement step size (ds) is too large or needs three '1's "
  ELSE
    Resolution ← last_zero_index * ds
    //Resolution ← dREF[last_zero_index+1]- dREF[1] //Alternative method
  END IF

  RETURN Resolution //RB
END FUNCTION
```

4. Compare data and find distinctly different datasets.
    a. Now, using the bootstrap method that is described in Algorithm 1, compare the segmented dataset at P_1 with the datasets at P_i ($i=1$ to M)[**]. The bootstrap method determines if two datasets are statistically similar or different. The outcome of this bootstrap comparison can be represented as a binary array, H, where each element of the array is H_i, where
        - $H_i=0$ indicates that the datasets are not statistically different (i.e., they are similar)
        - $H_i=1$ indicates that the datasets are statistically different

The resolution (R_B) is then calculated based on the sequence of 0s and 1s in the array H. In this method, resolution is the minimum distance at which datasets differ consistently at the statistical level. This can be determined by finding the first point at which the array H becomes consistently 1, at least three times.

- Example 1: If H = [0 0 0 1 1 1 1], the datasets become consistently different at the 4th element (or target position), or the last zero is at position 3; so, the resolution $R_B = 3*d_s$.
- Example 2: If H = [0 0 1 0 0 1 0], the datasets do not become consistently different, and the resolution is not clearly defined. In such a case, step #3 needs to be repeated by collecting data for a longer displacement, d_s.

5 RESULTS AND DISCUSSION

The setup described in Figure 1 was used to acquire data on the target at ~613 mm from the sensor, truncated to a 10 mm × 10 mm region along the X and Y directions (orthogonal to the depth direction, Z). Figure 2 shows the bootstrap distribution of the difference in depth (ΔZ_{Bi}, vertical lines). The 0s and 1s next to these lines show whether the data at each location were statistically the same (0) or different (1) from the dataset at position P_1. Based on this, the depth resolution is calculated using Algorithm 1.

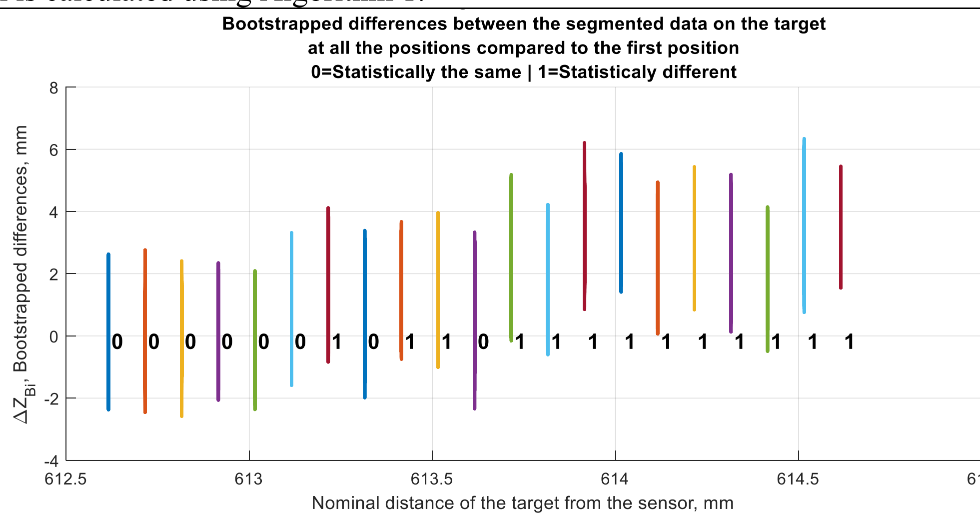

Figure 2: Bootstrapped differences between the segmented data on the target at all the positions compared to the first position

[**] Here, P_1 is compared with itself when i=1 and is not necessary but is retained for procedural convenience.

Further, we can calculate the depth resolution at subsequent locations if there are enough datasets. Figure 3 shows plots for five different parameters. The R_F plot refers to the method described by Rachakonda et al.[1] and shown in equation (1), where σ_{zc} is the standard deviation of the mean of the Z-coordinates, Z_q is the statistical mode of the inter-layer spacing in the Z-direction (also called quantization error), and k is the critical value of the t-distribution with degrees of freedom of $N\text{-}1$ for 5% significance level (95% coverage or confidence interval).

$$R_F = k \times \sqrt{2} \times \sqrt{(\sigma_{zc})^2 + (Z_q / \sqrt{12})^2} \qquad (1)$$

The metric R_P in Figure 3 refers to the range of values that represent 95% of the mean depths at each location, and the R_B plot represents the depth resolution as calculated by the Bootstrap method (Algorithm 1). Note that the plots for R_F and the R_P are based on single datasets at each location, whereas the plot for R_B is based on pairs of datasets. Further, R_P is a metric that does not assume any underlying distribution, whereas R_F assumes that the data follows a Gaussian distribution.

R_F, Z_q, and R_P metrics are relevant in the context of 3D sensors for determining whether specific target parameters (such as flatness, waviness, etc.) can be calculated. For example, if the inter-layer spacing $Z_q = 1$ mm, one cannot determine the flatness of that target to a value lower than 1 mm.

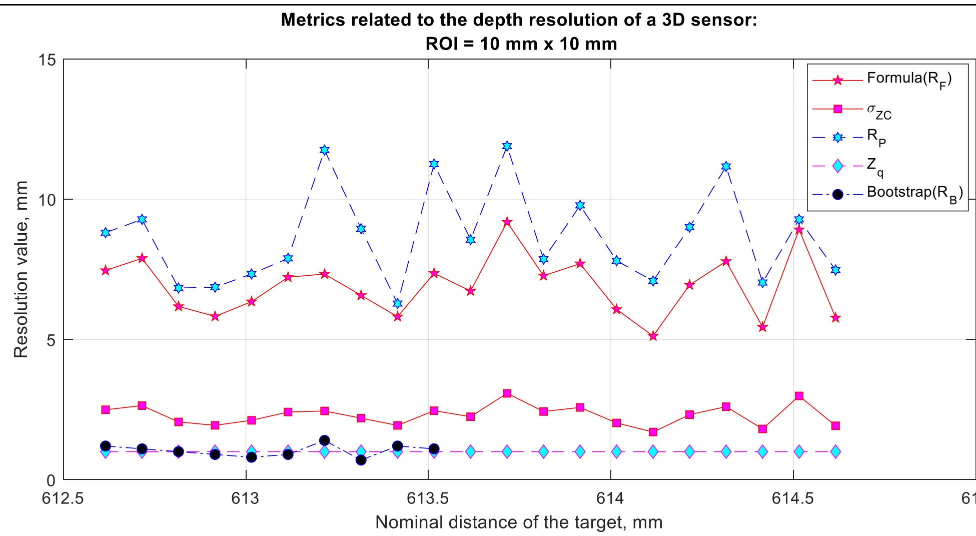

Figure 3: Comparison of different metrics related to depth resolution of 3D sensors

The Bootstrap method addresses several issues that limit the use of other statistical tests, such as the t-test or the Mann-Whitney U test [7]. The Bootstrap method does not require equal variances and is not affected by data quantization or binning. The role of the Bootstrap method is to assess whether the bootstrapped differences between locations (e.g., P_2 vs. P_1, P_3 vs. P_1, etc.) differ from 0. The datasets are deemed different or distinguishable if the bootstrapped differences are less than 0, with a probability of $\leq$ 5%.

Figure 2 shows the bootstrapped differences, where a '0' represents that the datasets are statistically the same, and a '1' represents that the datasets are statistically different. The distance between the target positions at which the datasets are distinguishable is the sensor's depth resolution (calculated using Algorithm 1). This value (R_B) is defined only at the target's location 'd' from the sensor and within a specific region of interest (ROI) on the near-Lambertian target.

Changes in the target-sensor distance, ROI, or target surface characteristics will affect these metrics.

Despite its merits, implementing the Bootstrap method poses physical challenges for sensors with exceptionally low noise levels. The primary constraints are hardware-related: translation stages must be capable of extremely fine increments (e.g., 0.1 μm to 1.0 μm for some sensors), and the ground-truth metrology system must offer uncertainty levels four times lower than those of the sensor being tested, a requirement that may be difficult to meet.

6 SUMMARY

Determining the resolution of a 3D sensor is a complex task due to the numerous variables introduced by the sensor itself, the target being measured, the measurement procedures, and other environmental factors. To simplify the problem, the task was narrowed to determining the sensor's depth resolution only. NIST staff investigated various statistical approaches and developed a Bootstrap-based metric as a potential solution in support of the ASTM work item, WK73176. This metric is particularly well-suited to the non-parametric nature of sensor data, providing a robust and reliable means of evaluating the depth resolution of 3D sensors.

7 ACKNOWLEDGEMENTS

The authors gratefully acknowledge the valuable input offered by the members of the ASTM E57.23 subcommittee and David Allen from the NIST Physical Measurement Laboratory.

8 REFERENCES

[1] Rachakonda, P. , Cheok, G. , Franaszek, M. and Saidi, K. (2023), Evaluating the depth resolution of 3D sensors for manufacturing automation applications, Measurement, , https://doi.org/10.1016/j.measurement.2023.113307, (Accessed September 29, 2025)

[2] ASTM WK73176: New Test Methods for Determination of a 3D Perception Systems Point Wise Spatial Resolution https://www.astm.org/membership-participation/technical-committees/workitems/workitem-wk73176 (Accessed March 3, 2026)

[3] Efron, B., & Hastie, T. (2016). Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge: Cambridge University Press. doi:10.1017/CBO9781316576533

[4] Saidi, K. , Cheok, G. , Qiao, G. , Horst, J. and Franaszek, M. (2020), Proceedings of the ASTM E57 Workshop on Standards for 3D Perception Systems for Robotic Assembly Applications December 2-3, 2019, Advanced Manufacturing Series (NIST AMS), National Institute of Standards and Technology, Gaithersburg, MD, https://doi.org/10.6028/NIST.AMS.100-33 (Accessed February 9, 2026)

[5] Rachakonda, P. , Saidi, K. , Franaszek, M. , Qiao, H. , Sweetser, J. , Khatoonabadi, A. , Dechow, D. and Rhodes, W. (2024), Towards the development of standards and performance metrics for 3D imaging systems, Proceedings of the SPIE Defense + Commercial Sensing 2024, National Harbor, MD, US, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=957648 (Accessed March 3, 2026)

[6] Pierce, T. and Rachakonda, P. (2025), Investigating the factors that influence 3D stereo depth sensor noise, SPIE Defense + Commercial Sensing 2025 conference, Kissimmee, FL, US, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=959923 (Accessed February 9, 2026)

[7] Chicco D, Sichenze A, Jurman G. A simple guide to the use of Student's t-test, Mann-Whitney U test, Chi-squared test, and Kruskal-Wallis test in biostatistics. BioData Min. 2025 Aug 20;18(1):56. doi: 10.1186/s13040-025-00465-6 (Accessed February 9, 2026)