Anatomic Scanning Using the Structure Sensor 3 and the Structure SDK

Paulo E. X. SILVEIRA *, Anton TOKAR, Dmitry GLADYSHEV, Jane MULLIGAN, Ravi SHAH
Structure, Boulder, CO, USA

https://doi.org/10.15221/25.17

Abstract

We present the Structure Sensor 3, an RGB-D stereo vision sensor optimized for anatomic scanning, along with the Structure Software Development Kit (SDK), a sophisticated SDK capable of precise 3D reconstruction using a range of depth sensors. We discuss features that set the Structure Sensor 3 apart from other sensors, including its operation at a wavelength that enables the filtering of sunlight, its ease of integration and use with a wide range of iOS devices, its robustness, accuracy and multiple presets, including unique ones that benefit from improvements in the signal-to-noise-ratio (SNR), frame rate and latency made possible by pixel binning. We also discuss the sensor's unique ability to filter pixels in the depth map based on its confidence threshold.

Next, we discuss how the Structure SDK was expanded and improved to support the Structure Sensor 3 while maintaining backward compatibility with Structure's long legacy of 3D sensors while being capable of easily integrating future sensors. We then discuss the Structure SDK's ability to effectively support 3rd-party sensors, such as Apple's TrueDepth and LiDAR sensors. Finally, we discuss the Structure SDK's ability not only to be easily integrated into iOS apps but also future plans to expand its use to other operating systems.

Keywords: 3D body scanning, 3D reconstruction, anthropometrics, computer vision, SDK

1. Introduction

Three-dimensional (3D) RGB-D sensing [1] has become a cornerstone technology across disciplines ranging from healthcare to industrial design to manufacturing and more. By enabling precise digital representations of the human body and its environment, precise 3D reconstruction [2-3] has opened the door to new forms of diagnosis, personalized therapies and patient-specific device fabrication. Despite this progress, the requirements of clinical and field applications, including robustness, portability, accuracy, and seamless integration, have remained a challenge for existing mobile depth sensing solutions.

The Structure Sensor 3 was originally conceived and developed to address these unmet needs. As the third generation of the original Structure Sensor launched in 2014, it embodies a decade of iterative advancement in mobile 3D scanning. With enhanced robustness, greater resistance to ambient illumination, extended operational lifetime and unique capabilities such as binning and depth confidence threshold filtering, the Structure Sensor 3 delivers a step change in performance for anatomical capture. These characteristics extend its utility from controlled laboratory environments into demanding real world clinical and healthcare applications.

In parallel, the Structure SDK has matured over the past decade into a comprehensive platform for 3D reconstruction and application development. Designed to abstract away the complexity of low-level sensor operation, the SDK provides developers with efficient access to real-time streaming, reconstruction, mesh processing, and anthropometric measurements. Importantly, it maintains backward compatibility with earlier Structure devices while extending support to third-party sensors, including Apple's LiDAR and TrueDepth systems. The SDK not only empowers developers to build advanced applications but also establishes a forward-looking foundation for future sensing technologies.

This paper presents the innovations introduced by the Structure Sensor 3 and Structure SDK. We first detail the sensor's design improvements and experimentally validate critical performance metrics, such as depth accuracy, robustness and noise characteristics. We then describe the new SDK features that enhance object scanning and augmented reality (AR) applications, including mesh export and continuous tracking. Together, these contributions demonstrate how the combination of robust hardware with a flexible software framework advance the state of the art in anatomic scanning and accelerate adoption in healthcare and beyond.

* paulo@structure.io; https://www.structure.io/

Figure 1 shows a few sample scans using the Structure Sensor 3 and the Structure SDK: in the top left we have a color-rendered scan of a female right foot, size 7.5, showing rich detail and good color rendering (166k faces). In the bottom left we see a prime example of the Structure Sensor 3's ability to scan outdoors, in a sunny day by scanning a large rock (about 1.4m wide, 296k faces). In Section 2.1 we discuss the sensor features that make such outdoor scans possible. On the right we see the sensor's ability to perform full-body scans, showing the color-rendered scan of a large mannequin (1.94m tall, 240k faces).

Figure 1: Sample scans performed using the Structure Sensor 3.
Top left: female right foot, size 7.5 (US). Bottom left: Large rock (1.4m wide), captured outdoors during daylight.
Right: full-body scan of a large mannequin (1.94m tall).

This paper is organized as follows: the Structure Sensor 3 is described in Section 2, along with a list of key differences compared to its immediate predecessor, the Structure Sensor Pro. That description is followed by experimental verification of some critical parameter metrics, such as sensor noise, along with a description on how to benefit from some of the new features available in the Structure Sensor 3 for improved scanning, including binning, the use of aggregate windows of different sizes, and the optimized use of the confidence threshold.

Section 3 describes the Structure SDK and its use as a state-of-the-art 3D reconstruction software capable of supporting multiple sensors. An overall description of the SDK architecture is followed by a more detailed description of new methods specifically for object scanning and augmented reality (AR) applications. Following that, the availability of foot anthropometrics in the SDK is also described.

Section 4 presents the main conclusions of the paper, including future plans for the SDK and sensors.

2. The Structure Sensor 3

Originally developed by Occipital Inc. in 2014, the Structure Sensor revolutionized mobile scanning by making it affordable and easy-to-use [4]. Followed by the launch of the Structure Sensor Mark II in 2019 and the Structure Sensor Pro in 2021, Occipital continued to innovate and gradually cater to the needs of healthcare applications, demonstrating along the way its ability to surpass the performance of far more costly sensors in medical applications [5-6].

In August 2022, Structure split from Occipital and became a new company dedicated to the support of healthcare applications. This dedication culminated with the launch, in September 2024, of the Structure Sensor 3, the first sensor designed from the ground up with the needs of the healthcare community in mind. Because it is consistently more robust – passing standardized drop tests – it upholds its calibration significantly longer than its predecessors, as evidenced by its robust depth coverage. Its battery life has also been extended compared to that of the Structure Sensor Mark II, making the Structure Sensor 3 overall a best-in-class choice for the rigors of daily use.

Figure 2 shows a 3D rendering of the Structure Sensor 3, highlighting the position of the infrared (IR) cameras, the laser projector, status LED and connector. For increased robustness, the Structure Sensor 3 benefits from an all-aluminum, black-anodized enclosure. To maintain compatibility with its predecessor, the new sensor retains the same baseline distance between cameras (75mm) and the cameras have nearly the same field of view (FOV) of 50° horizontal and 59° vertical (74.3° diagonal).

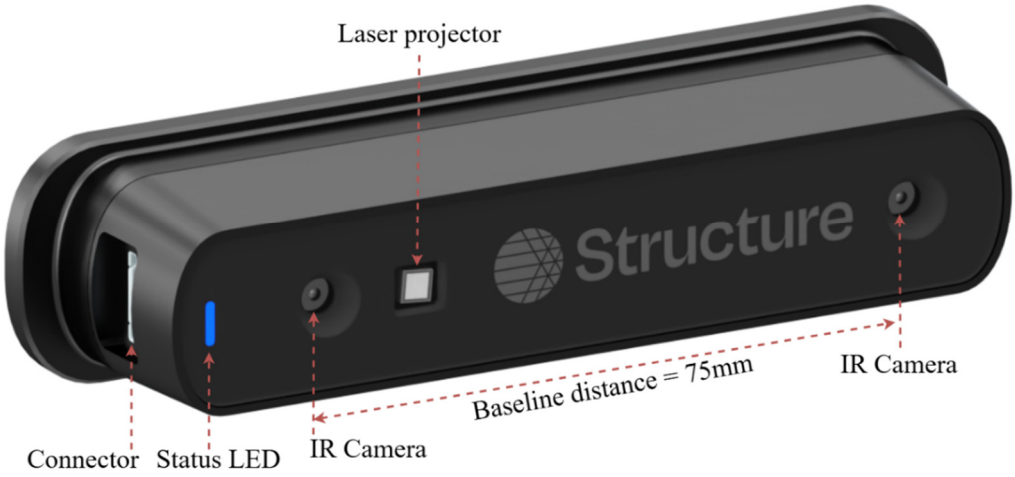

Figure 2: Three-dimensional rendering of the Structure Sensor 3, showing the location of its connector, status LED, laser projector and IR cameras with a baseline distance of 75mm.

Table 1 compares key specifications of the Structure Sensor 3 with those of the Structure Sensor Pro, highlighting key differences that stand the new sensor apart, such as its increased operational time, improved outdoor operation, robustness and processing power. Some of those key specifications are discussed in more detail next.

Because the Structure Sensor 3 closely matched the baseline distance and FOV of the Structure Sensor Pro, its depth sensing resolution also matches that of that sensor when used in full-resolution, as shown by the blue line in Figure 3. The orange line depicts the depth resolution of the Structure Sensor 3 when used in 2x2 binned mode (more on that in Section 2.4 below). The red dashed line highlights the depth resolution of 1mm, achieved at a distance of 0.8m in binning mode, for reasons that will become clear in Section 2.6.

For reasons that will become clear in Section 2.1, there is no theoretical limit to the maximum range of the sensor. In practice, however, the loss in depth resolution and the ability of the laser projector to illuminate far objects limit its maximum range to no more than 15m, with 12m being a more practical limit. The closest range depends on the mode: in full-resolution it is limited to 33cm with a depth resolution of 0.08mm, while in binned mode the minimum range can be as close as 25cm with a depth resolution of 0.1mm. Note that even though the full-resolution mode provides the sensor with the highest depth resolution over all distances, the binned mode enables one to get closer to the body part being scanned and, in so doing, almost matches the depth resolution of the sensor in full-resolution mode.

Table 1: Comparison between the Structure Sensor 3 and the Structure Sensor Pro.

Changes / Improvements	**Structure Sensor Pro**	**Structure Sensor 3**
Improved outdoor operation	825nm center wavelength.	940nm center wavelength – improved sunlight rejection.
Increased operational time	650mAh Battery capacity, 750mA current consumption in operation.	1250mAh Battery capacity, 830mA current consumption in operation.
Increased battery lifetime	1C max. battery discharge rate.	2C max. battery discharge rate.
Calibration Robustness	Thinner chassis. Adaptor plate attached after calibration.	Stiffer chassis, combined backer and adaptor plate, attached before calibration.
Physical Robustness	Survives 1 meter drop tests in shipping box, with air pillows.	Survives 1 meter drop tests in the original retail package.
Reduced number of parts, simplified manufacturing	Additional screws and parts. Additional manufacturing step.	Combined backer and adaptor plate
Processing power	NU3000 processor: 660MHz. 128 MB LPDDR2.	NU4000 processor: 1GHz, 1.8 TOPs, CNN co-processor. 1GB LPDDR4.
Binning mode	N/A	Enables wider scanning range and improved SNR.
Ingress Protection	N/A	IP53 compliant

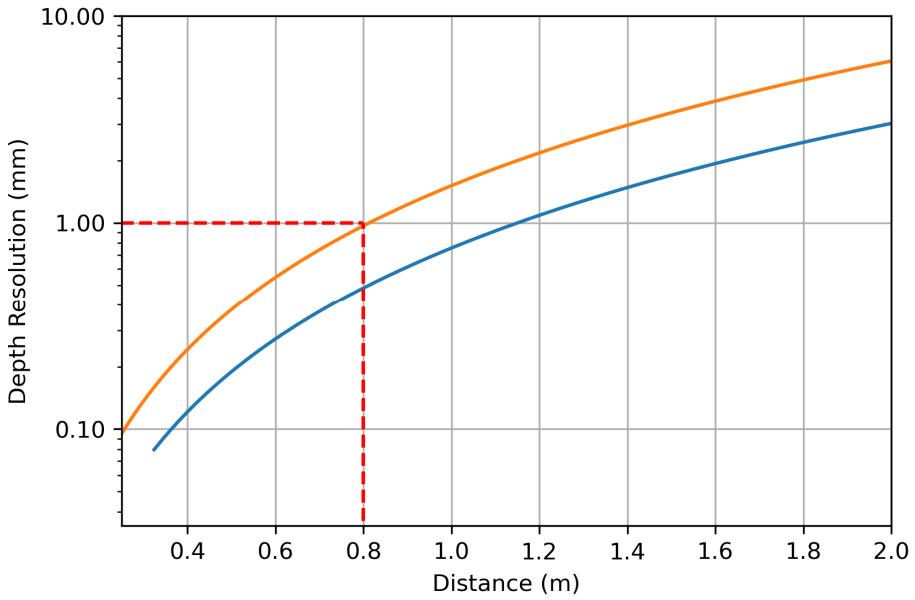

Figure 3: Depth resolution limit of the Structure Sensor 3 when used in full-resolution (blue) and in 2x2 binned (orange) modes. Log scale in the vertical axis. Red-dashed lined: 1mm resolution at 0.8m, binned mode.

2.1. Laser projector and reduced sensitivity to ambient light

The Structure Sensor 3 measures depth via assisted stereo [7]. This means that depth is measured primarily by performing triangulation between the disparity of the images seen by the left and right IR cameras, using a disparity search window to find corresponding points in each camera. A laser projector is provided to ensure that all scenes have at least a minimum of diversity in features, thus enabling the continuous detection of disparity. This has the added benefit of enabling the filtering of light from other light sources, thus preventing interference from external illuminators, such as room or sunlight.

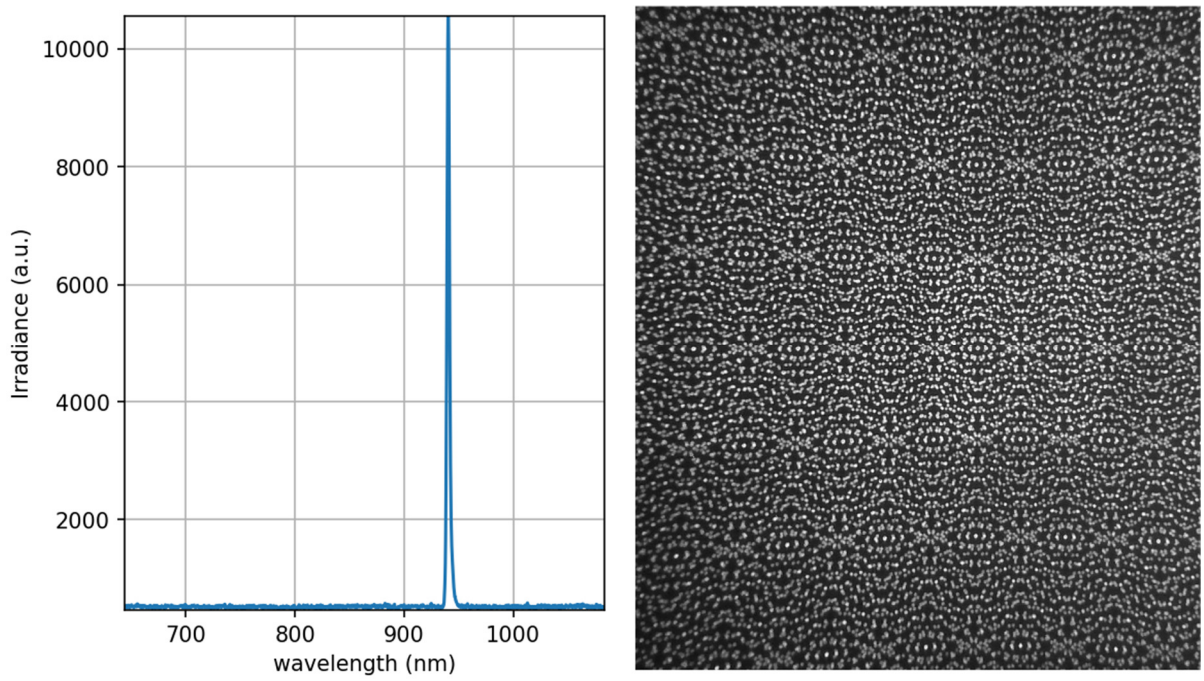

Figure 4: (Left) Typical spectrum of light emitted by a Structure Sensor 3 laser projector, showing a narrow spectrum, centered at 940.6nm, close to the nominal peak wavelength of 940nm. (Right) Sample projected dot pattern, zoomed into center 850x1000 pixels for clarity, showing a modulation ratio of 87.5%.

The proliferation of mobile sensing technologies unlocked the development of vertical cavity emission lasers (VCSELs) operating at the 940nm center wavelength [8]. This translates to affordable and compact laser sources capable of operating deeper in the near infrared. The CMOS cameras of the Structure Sensor 3 possess bandpass infrared filters with a passing band matched with the transmission spectra of its VCSEL light sources. By filtering out light away from the 940nm center wavelength, the sensor becomes largely immune to radiation from other light sources. Figure 4 (left) shows a typical spectrum of the light emitted by one of its laser projectors, showing a narrow peak centered at 940.6nm and a full-width-half-maximum bandwidth of 3nm. Figure 4 (right) shows a sample dot pattern projected by the sensor during operation, showing a relatively large modulation ratio of 87.5%.

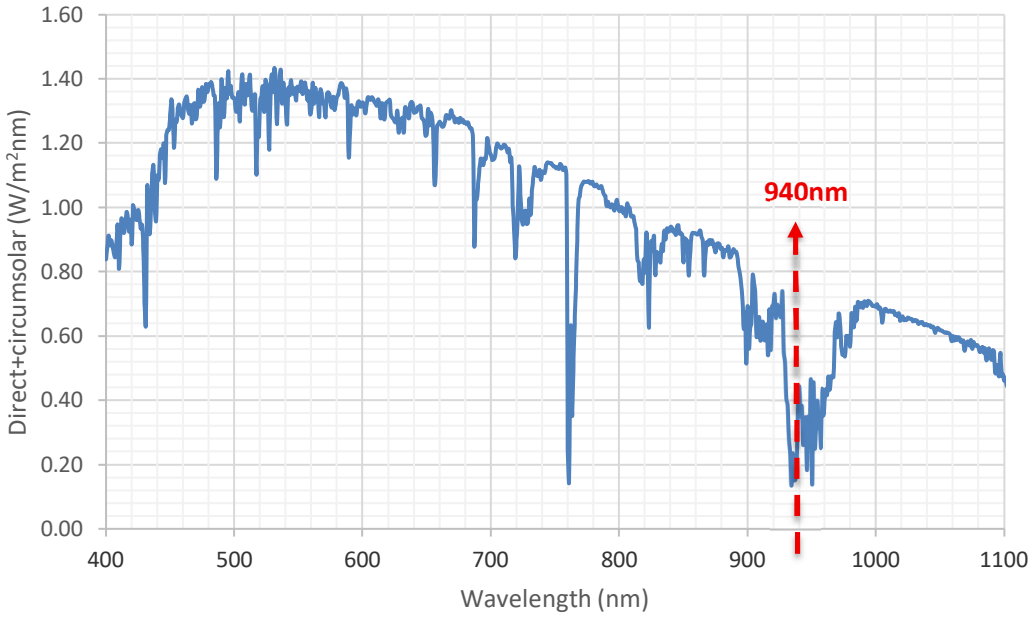

Figure 5: Solar spectrum on the surface of the Earth, highlighting the low light concentration at 940nm. Modified from [9].

Figure 5 shows the solar spectrum at the surface of the Earth, showing the low concentration of sunlight at 940nm. The reason being the presence of water vapor in the atmosphere: water presents a strong absorption peak at 940nm which removes much of solar radiation at that wavelength. Note that CMOS sensors have a responsivity that is limited within the wavelength range from 400nm to 1100nm. Hence, the solar spectrum in wavelength ranges below 400nm and above 1100nm are of little consequence, even if the bandpass filter does not block light very well beyond that wavelength range. Consequently, the Structure Sensor 3 is significantly less affected by sunlight than any of its predecessors.

2.2. Laser modes

The Structure Sensor 3 features two distinct laser modes: Normal and Fast, the main difference between the two being that in Fast mode the laser is illuminated 67% brighter. Note, however, that the maximum duty cycle of the laser is clamped at different values depending on the laser mode to prevent overheating and loss of optical conversion efficiency. This results in an effective maximum exposure time available to each laser mode, depending on the frame rate at which the sensor is streaming depth. The maximum exposure time available at each frame rate and each laser mode is summarized in Table 2.

Table 2: Maximum exposure times available at different frame rates and different laser modes.

Laser Mode	Frame rate (fps)	Max. exposure time (ms)
Normal	60	8.33
	30	16.67
	15	16.67
Fast	60	5.0
	30	10.0
	15	10.0
Off	30	33.33
	15	33.33

As is the case with all its predecessors, the Structure Sensor 3 also meets IEC 60825-1: 2014 laser safety standards and, hence, is also a certified Class 1 laser device, in all laser modes.

2.3. Battery capacity and discharge rate

How much electrical energy a battery can deliver to an electronic device is measured by its capacity, usually expressed in terms of milli Amperes of current available at a constant voltage for one hour of operation, abbreviated as mAh. The Structure Sensor 3 almost doubles the battery capacity compared to the Structure Sensor Pro. However, given its more powerful processor and memory access, at an increased clock rate (1GHz) this results in increased current consumption during streaming, resulting in a net increase in operational time of approximately 74%. During normal operation (Normal laser mode, 5ms exposure time) the Structure Sensor 3 is typically able to continuously scan (i.e., stream depth maps) for 2.5h when powered by the battery alone (i.e., when connected to an iOS device with an older Lightning connector). That time can be considerably longer when using the newer – and increasingly more common - USBC connector, the total time depending on the power delivery capabilities of the specific device.

The laser projector delivers concentrated bursts of light at time intervals exactly synchronized with the exposure time of the IR cameras. This not only requires specialized current drive circuitry, but also batteries that are capable of providing the required current continuously. Such bursts of light provide the benefit of “freezing time”, thus making the sensor less sensitive to motion blur, but the associated bursts in current can slowly degrade the battery capacity. Thus, by using batteries with twice the discharge rating of its predecessor, the battery of the Structure Sensor 3 is expected to last at least twice as long.

2.4. Binning mode

Sensor presets are an SDK concept that enables sensors to be easily and efficiently configured for different applications. In addition to supporting all the presets available in the Structure Sensor Mark II, the Structure Sensor 3 features new presets based on its unique ability to bin multiple pixels.

CMOS sensors operate by converting light (photons) into electrical signals (photoelectrons) which are then converted into digital signals (bit) using an analog-to-digital-converter [10]. Binning is a technique that enables the combination of the signals (bits) from neighboring pixels into one, larger *virtual* pixel. For example, in 2x2 binning the bits of every 4 neighboring pixels are combined (added) into one, resulting in a pixel with 4 times the light detection ability of a single pixel. Since noise is added non-coherently, this results in an increase in signal-to-noise ratio (SNR) of N / sqrt(N) = 2, or 3dB. The tradeoff being the loss in resolution. Nevertheless, for applications in which the resolution of the depth map is not as important, the improvement in SNR results in significant performance improvements. Furthermore, the increase in operating range directly translates into ease of operation.

2.5. Confidence threshold and aggregate window size

Depth from stereo starts with the determination of the disparity offset between the left and right IR cameras. This is a map of the disparity, measured in fractional pixels, at each point in the FOV of the IR cameras. Close objects result in larger disparity while objects at infinity result in zero disparity.

The disparity matching is followed by a cost aggregation step, performed over a support window of a predefined size. The larger the window, the larger the certainty of the match, but also the lower the transverse resolution of the resulting disparity map. That is because the window acts as a convolutional low-pass filter, its exact operation depending on the nature of the matching algorithm [11]. In addition to finding the disparity, the matching algorithm also returns a metric determining the quality of the match. This is a value provided in arbitrary units, ranging from zero to 15, where 15 (zero) determines the best (worst) match possible.

When using a mobile scanner, such as any sensor in the line of Structure Sensors, it makes sense to filter pixels in the depth maps as a function of their underlying confidence threshold. After all, not all pixels available on the map are equally valuable, and a user has the opportunity to visit missed spots in the patient body many times over the range of a few seconds, given the relatively high frame rate of the sensor. This suggests the definition of a parameter that we decided to call the *confidence threshold*: the minimum confidence value of pixels accepted by the depth map. Intuitively one should expect high values of the confidence threshold to result in depth maps that have lower *coverage* (i.e., the ratio of valid pixels over the total number of pixels) but *high quality* pixels, while low values should result in high coverage but *low quality* pixels. This begs the question: how does one measure the *quality* of a depth pixel?

Figure 6 shows the mean standard deviation of the disparity (blue) and the coverage (green) of the Structure Sensor 3 as a function of the confidence threshold when the sensor is facing a solid wall 80 cm away. The sensor is operating in binned mode (540x600 pixels), and 40 images are captured in each case. In the top plot the aggregation window size is set to 11x11, close to the default size (11x15), while in the bottom plot the aggregation window size is set to 7x7. Since the sensor is facing a solid wall, one should expect any variation in the disparity readings to be a measure of disparity error and, hence, to reduce as we filter away low quality depth pixels in favor of good quality ones. This is exactly what happens when the confidence threshold increases. Interestingly, in both cases we observe two distinct regions: a region where the mean standard deviation of the disparity drops linearly as a function of increasing confidence thresholds, followed by a flat region where additional increases in the confidence threshold result in almost no reduction in the mean standard deviation.

Two distinct regions are also observed in the coverage curve, but in opposite direction: that curve starts flat, where initial increases in the confidence threshold have a negligible effect in coverage, followed by a region where increases in the confidence threshold results in a rapid drop in coverage.

These distinct regions in the two curves provide us with a simple algorithm to optimize the confidence threshold: select the lowest value in the flat region of the two curves. In the case of the larger, 11x11 aggregation window (top plot) that results in a confidence threshold value of 7. In the case of the smaller 7x7 aggregation window (bottom plot) that results in a confidence threshold of 10.

Also note that in both cases the resulting mean standard deviation of disparity ($\overline{\sigma d_\iota}$) has a similar value (0.02 pixels). Inspection of experimental results show that the mean disparity ($\overline{d_\iota}$) is 58.375 pixels. Given that the mean depth $\bar{z}$ of a rectified stereo vision system is given by Equ. 1:

$$\bar{z} = \frac{B.f}{\overline{d_\iota}.p}$$

where B is the baseline distance, f is the focal length of the IR cameras (assumed identical) and p is

the pixel pitch. By calculating the differential of Equ. 1, we have that the mean differential depth is given by Equ. 2:

$$\overline{\sigma z} = \frac{\bar{z}.\overline{\sigma d_\iota}}{\bar{d}_\iota}$$

From which we get a mean differential depth of 0.27mm. Referring back to Figure 3, we can then interpret this result as follows: at a distance of 80cm, the Structure Sensor 3 in binning mode can resolve depth with a resolution of 1mm with a standard deviation of 0.27mm, thus showing that the sensor is more than capable of meeting the depth resolutions shown in that plot.

In addition, one can calculate the sensor SNR, in dB, using Equ. 3:

$$SNR = 20 \log_{10} \frac{\overline{d_\iota}}{\overline{\sigma d_\iota}}$$

Resulting in an SNR of 69.3dB, a high value, consistent with the low depth noise calculated using Equ. 2.

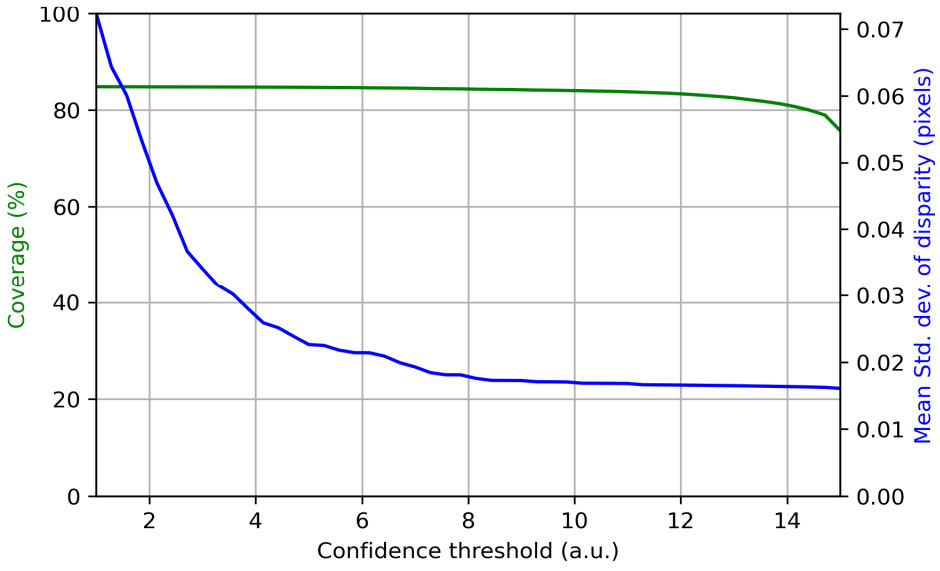

Figure 6: Structure Sensor 3 Mean Standard Deviation (blue) and % Coverage (green) as a function of the Confidence Threshold. The disparity search window is 11x11 (7x7) in the top (bottom) plot, showing that a confidence threshold of 7 (10) provides significant noise reduction with almost no loss in coverage.

Note that in the Structure SDK up to version 3.5 the confidence threshold has been limited within the range from zero to 7. This is to prevent developers from inadvertently selecting values that are too large, resulting in low coverage, which can cause tracking errors and, hence, poor 3D reconstruction. Starting with SDK 3.6, developers have the ability to adjust the aggregate window size, in which case it makes sense to also provide them with the ability to select a wider range of confidence thresholds, as described above.

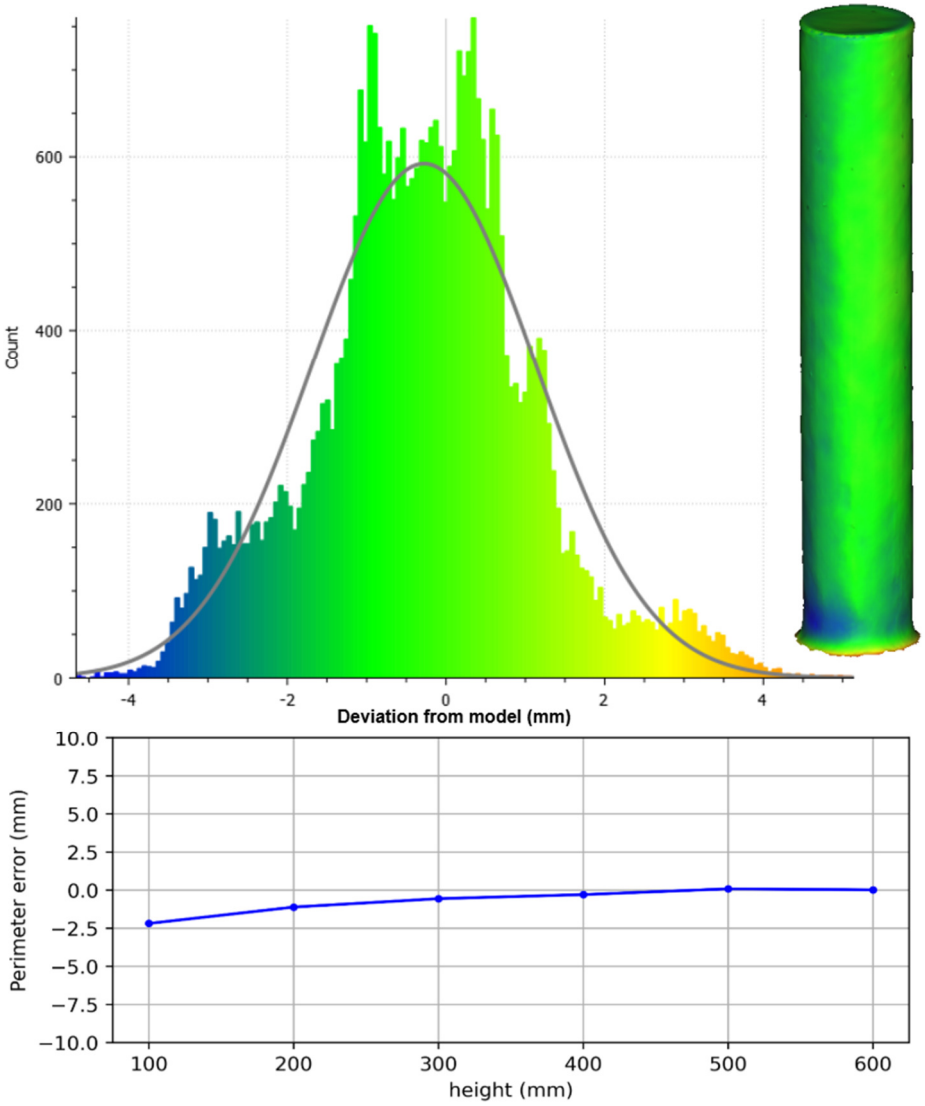

Figure 7: Histogram (top left) of a typical 3D reconstruction of a PVC pipe (top right) performed using a Structure Sensor 3, with a 7x7 aggregate window and confidence threshold of 10, showing a 0.29% relative RMS deviation in perimeter measurements from an ideal model (bottom).

2.6. Sample accuracy result

Figure 7 shows a sample scan of a PVC pipe with a length of 660.92mm and a mean diameter of 114.17mm. The scan was performed in a room with open windows, exposed to indirect Sunlight. For best results, the scan was performed using an aggregate window size of 7x7 and a confidence threshold of 10. Reconstruction was performed using the Structure SDK, without any post-processing steps. The resulting model was then compared to the mesh of a perfect cylinder of the same diameter and length, using interconnected points (ICP) to register the two meshes [12]. Then, the signed distances between the scanned and the reference mesh were calculated and plotted in the histogram shown in the top left of Fig. 7, resulting in a mean deviation of -0.27mm and a standard deviation of 1.4mm.

The plot in the bottom of Fig. 7 shows the perimeter error – the difference between the measured perimeter and the nominal value of 358.68mm – at multiple distances from the ground plane, starting from 10cm up to 60cm in 10cm increments, showing a close match between measured and nominal values, with a small deviation close to the ground plane. Calculating the absolute RMSE (root-square mean error) of the perimeters, we obtain a value of 1.04mm. Dividing that by the nominal perimeter value and representing it as a percentage, we get a relative error of 0.29%.

Note that several error sources contribute to the RMSE of a 3D reconstruction, both in the mapper used to integrate sensor observations, as well as in the tracker used to estimate poses, and those errors can accumulate during a scanning event. See [13] for more details on how those errors affect the accuracy of the 3D reconstruction, and on the use of pipes for characterizing the reconstruction of human limbs.

2.7. Sensor presets

Sensor presets are a convenient way for developers to optimize their software applications to specific use-cases, encapsulating several sensor-specific parameters into a short description that broadly describe their intended purpose. They were first introduced in the Structure SDK with the launch of the Structure Sensor Mark II, expanded with the launch of the Structure Sensor Pro, and now expanded once again with the launch of the Structure Sensor 3.

The Structure Sensor 3 includes all presets available on the Structure Sensor Pro, with the exception of the Hybrid preset. That preset has been mapped to the Simplified preset, its closest equivalent in the Structure Sensor 3. In addition, the new sensor benefits from several new presets, such as Simplified, Ultra Close, Outdoor, Long Range and Body Detailed. All presets available in the Structure Sensor 3 are summarized in Table 3.

Table 3: List of Structure Sensor 3 presets.

Preset name	Resolution (pixels)	Frame rate	Gain**	Laser Mode	Exposure (ms)**	Min. range	Max. range
Simplified (Default)	520x600*	60fps	2x	Normal	8.33	27cm	5m
Body	752x940	30fps	2x	Normal	10	36cm	102cm
Body Detailed	1040x1200	15fps	2x	Normal	10	36cm	102cm
Close	752x940	30fps	1x	Normal	5	32.5cm	75cm
Medium Range	752x940	30fps	4x	Normal	16.7	56cm	10m+
Long Range	752x940	30fps	4x	Fast	10	56cm	15m+
Outdoor	752x940	30fps	1x	Fast	10	56cm	12m+
Ultra Close	520x600*	60fps	1x	Normal	2.5	25cm	2m
Room Scanning	520x600*	60fps	4x	Fast	10	29cm	12m+
Dark Object	520x600*	60fps	8x	Fast	10	27cm	5m

* Binned presets

** Gains and exposure times associated with each preset are default values that can be overridden via SDK, depending on the application.

The Simplified mode extends to the Structure Sensor 3 the same *simplicity* available in the original Structure Sensor: a single preset that works well in most situations. When in doubt, start with that. So, it should be no surprise that this mode, with its wide range of operation, high frame rate and large SNR, was selected as the “Default” preset in the Structure Sensor 3.

Ultra Close combines binning to push the envelope of the range originally available in the Close mode. Note, however, that the loss of maximum range is significant if compared to the Simplified mode, making that preset preferable for most applications.

Long Range and Outdoor benefit from the new Fast laser mode, which was also extended to the pre-existing Room Scanning preset. In addition, Outdoor was created to showcase the superior operation of the Structure Sensor 3 outdoors, thanks to its new wavelength of operation.

Finally, the Body Detailed mode was added to enable the highest resolution possible in the Structure Sensor 3: 1040x1200. Note, however, that this resolution comes at the cost of halving the frame rate to 15fps. This means that the subject should be still during scanning, and the operator should move at a consistent and slow rate to prevent tracking loss. Note that motion blur is not a problem, since the exposure time (10ms) is sufficiently short.

2.8. IP53 compliance

The Structure Sensor 3 is the first in the Structure line of sensors to be IP (Ingress Protection) compliant. IP53 means that 1) the sensor is dust-protected. Meaning, while dust ingress is not entirely prevented, the amount of dust that can enter the enclosure is not sufficient to interfere with its operation. And 2) the sensor is protected against spraying water. This protection applies to water sprayed at any angle up to 60 degrees from the vertical. In practice it means the sensor can be cleaned and chemically sterilized without affecting its operation.

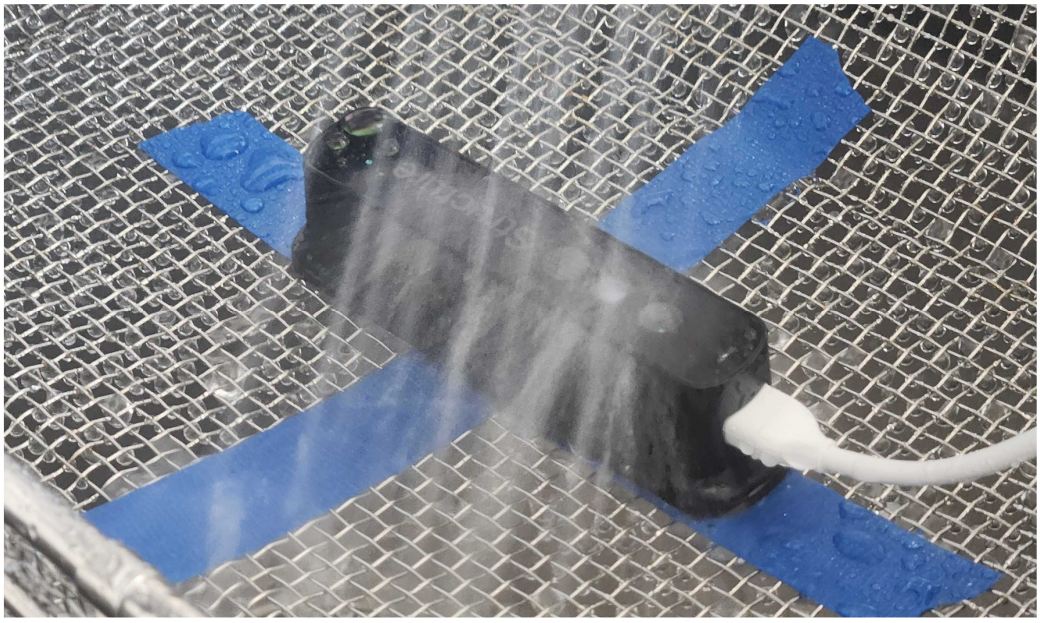

Figure 8: IP53 water spray testing of the Structure Sensor 3

Figure 8 shows a sample Structure Sensor 3 undergoing water spray testing while successfully passing IP53 testing.

3. The Structure SDK

The Structure SDK greatly simplifies the development of user applications for object reconstruction. It removes the need for developers to delve into the intricate details and complexities of the reconstruction process by providing them with a powerful framework that streamlines application creation.

3.1. Main features of the Structure SDK

The Structure SDK provides a comprehensive suite of tools specifically engineered for advanced 3D scanning, object tracking, and subsequent mesh processing. Its robust capabilities are designed to cater to a wide array of application requirements, from high-fidelity object digitization to large-scale environment mapping. Its core capabilities include (see Figure 9):

3.1.1. Streaming Management

This fundamental feature allows developers to efficiently process diverse data streams from a variety of sensors and mobile devices in real time. This includes, but is not limited to, depth frames (providing crucial spatial information), color (RGB) frames (for visual fidelity), and IMU (Inertial Measurement Unit) data, offering information on motion and orientation in sensors such as the Structure Sensor (Mark II), and from iOS hosts. This real-time processing capability is essential for dynamic environments and interactive applications.

3.1.2. Real-Time Object Reconstruction

At the heart of the SDK's power is its robust implementation of various Simultaneous Localization And Mapping (SLAM) algorithms. These complex algorithms, which simultaneously build a map of the environment and track the device's location within it, are exposed through an intuitive and easy-to-use API. This makes real-time object reconstruction one of the SDK's standout features, enabling the precise digital capture of physical objects and environments.

3.1.3. Mesh Texturing

Beyond simply reconstructing the geometry of an object, the SDK provides sophisticated algorithms for mesh colorization. This advanced functionality allows for the application of realistic color textures to the reconstructed 3D meshes. The result is not just a geometric representation, but a fully textured digital twin that accurately reflects the visual appearance of real-life objects, enhancing the realism and utility of the scanned data.

3.1.4. Anthropometrics

The automated detection of anatomic landmarks and their measurements, also known as *anthropometrics*, is a specialized feature of the Structure SDK that leverages advanced algorithms for automating the precise measurement of various body parameters, many of which can be useful in the research of diagnostic and personalized therapies of scanned subjects. This feature is not limited to meshes captured using the Structure SDK, but is available to any external mesh imported into the application. More details about the automated foot anatomic landmark detection and metrics, available starting in SDK 3.5, are available in [14].

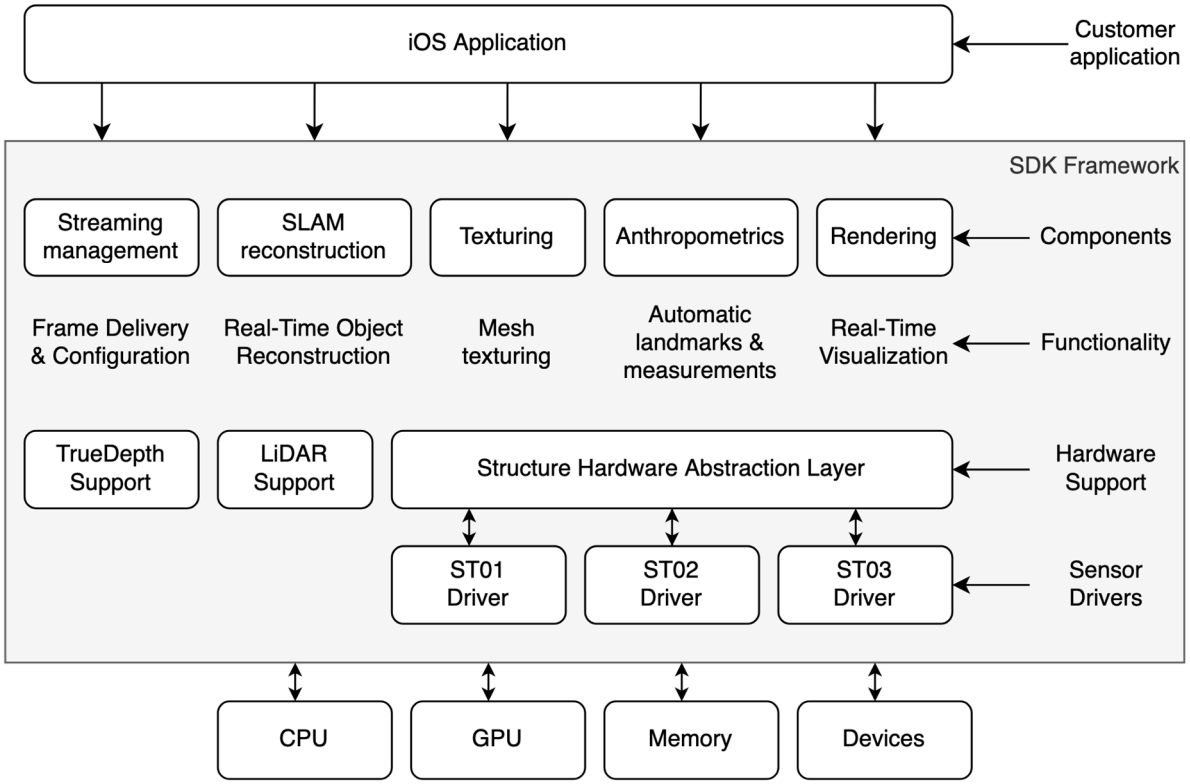

Figure 9: Structure SDK Architecture Diagram.

3.1.5. Visual Feedback

To facilitate an interactive and informative user experience, the SDK includes robust rendering capabilities. This enables real-time visualization of the object reconstruction process as it unfolds. Developers and users receive instant visual feedback and a live preview of the reconstruction, allowing for immediate assessment and adjustments, thereby optimizing the scanning workflow and ensuring desired outcomes.

3.2. Structure SDK advanced design

The SDK's design philosophy emphasizes flexibility in its deployment, supporting various integration paradigms to suit different development needs. Developers can leverage its core functionalities for custom application development, integrating specific modules as required, or utilize higher-level APIs for rapid prototyping and deployment of 3D scanning solutions.

3.2.1. Abstraction Layer for Multi-Sensor Integration

The proliferation of diverse sensor technologies within the domain of 3D data acquisition necessitates a robust and flexible software architecture to ensure seamless interoperability and ease of development. Our proposed abstraction layer serves precisely this critical role, facilitating comprehensive compatibility across all available sensors integrated within the SDK. This layer is designed to effectively manage and conceal the inherent heterogeneities among different sensor models.

By abstracting these low-level hardware distinctions, the abstraction layer presents a unified and consistent programming interface to developers. This uniformity not only simplifies application development but also enables effortless switching between different sensor types without requiring extensive code modifications or sensor-specific conditional logic. A testament to its efficacy, our existing iOS applications, initially developed a decade ago for the first-generation Structure Sensor (ST01), are fully compatible with the latest ST02 and ST03 sensor families, demonstrating sustained functionality and performance without necessitating any refactoring or special adaptations for newer hardware.

3.2.2. Supported Sensor Technologies

The SDK offers comprehensive support for a range of 3D sensing hardware, ensuring broad applicability across various mobile platforms and use cases. Table 4 provides a consolidated overview of the currently integrated and supported sensor technologies.

Table 4: Sensors currently supported by the Structure SDK.

Sensor Family / Technology	**Type of Sensing**	**Integration Methods / Notes**
Structure Sensor ST01	Structured Light	Legacy support, foundation hardware for initial SDK development.
Structure Sensor ST02	Assisted Stereo	Enhanced structured light sensor with improved range and accuracy.
Structure Sensor ST03	Assisted Stereo	Latest generation structured light sensor, offering superior performance, robustness and new features.
Apple LiDAR	Flash Time-of-Flight	Integrated via Apple's native frameworks: AVFoundation and ARKit. Primarily for room scanning.
Apple TrueDepth	Structured Light	Integrated via Apple's native frameworks: AVFoundation and ARKit. Primarily for small object, close range scanning.

3.2.3. Advantages of Utilizing the Structure Sensor and SDK

The adoption of the Structure Sensor in conjunction with its accompanying SDK offers several compelling advantages for developers and researchers engaged in 3D scanning and spatial computing applications.

Exceptional Processing Speed: Our foundational work in algorithm optimization dates back to the early generations of iPhone and iPad hardware. This long-standing commitment to efficiency has resulted in highly optimized processing routines. Consequently, even with the increasing complexity of modern mobile processors, the per-frame processing time for depth and color data remains remarkably low, typically requiring only a few milliseconds. This high processing speed facilitates real-time applications and enhances user experience.

Improved Reliability: A significant investment of effort has been dedicated to rigorous testing and the systematic elimination of potential errors and vulnerabilities within the SDK. Every new release undergoes a meticulous validation procedure, encompassing comprehensive testing across diverse hardware and environmental conditions, prior to its deployment to users. Furthermore, all our proprietary applications built upon the SDK are subjected to and successfully pass the stringent review processes mandated by Apple, underscoring their stability and adherence to platform standards.

Robustness Across Scenarios: The SDK demonstrates remarkable robustness, supporting a multitude of operational scenarios and accommodating a broad spectrum of sensor configurations. This adaptability ensures consistent performance and reliable data acquisition across varied environments and application demands.

Comprehensive Support and Partnership Opportunities: Beyond the technical capabilities of the SDK, we are committed to providing robust support and guidance throughout the application development lifecycle. For organizations that may lack an in-house development department, we facilitate connections with our trusted partners who possess extensive expertise in integrating our technology, ensuring that clients can successfully realize their project objectives.

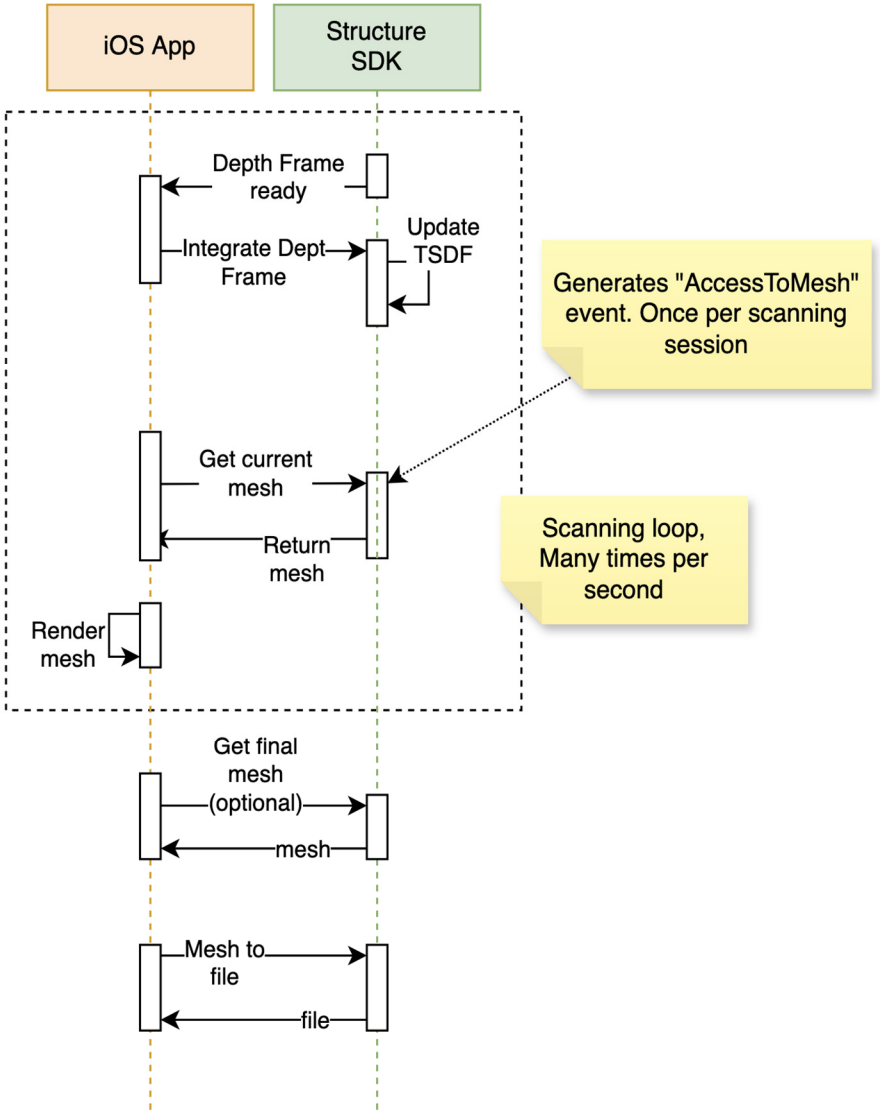

Figure 10: Structure SDK Continuous Mesh Access Flow Diagram.

3.3. Mesh access modes

For maximum flexibility, the Structure SDK offers two mesh access modes: the legacy *continuous access mesh* mode and the new and simplified *mesh extract* mode. Both are explained in detail next.

3.3.1. Continuous mesh access

With *continuous mesh access* the developer has constant access to the mesh data during the whole reconstruction process while the truncated signed differences function (TSDF) is continuously updated [2]. This enables advanced features like assisted AR applications. One of the key advantages of continuous mesh access, as illustrated in Figure 10, is the developer's uninterrupted ability to retrieve and manipulate mesh data throughout the entire reconstruction process. This constant stream of information opens up a wide array of possibilities for creating sophisticated and interactive AR applications.

For instance, with continuous mesh access, AR applications can offer advanced features such as real-time environmental understanding and dynamic object placement. Developers can leverage this consistent data flow to implement functionalities like intelligent occlusion, where virtual objects are realistically hidden behind physical ones, or precise spatial mapping, allowing for highly accurate virtual overlays on the real world. This capability is crucial for developing AR experiences that seamlessly blend digital content with the physical environment, leading to more immersive and practical applications across various industries, including design, education, and entertainment.

In healthcare, continuous mesh access has been successfully used to enable pre-operative planning [15] and surgical navigation [16], resulting in considerable reduction in time required in the operating room, reduced times when patients need to undergo general anesthesia, with associated improvements in patient safety and cost savings.

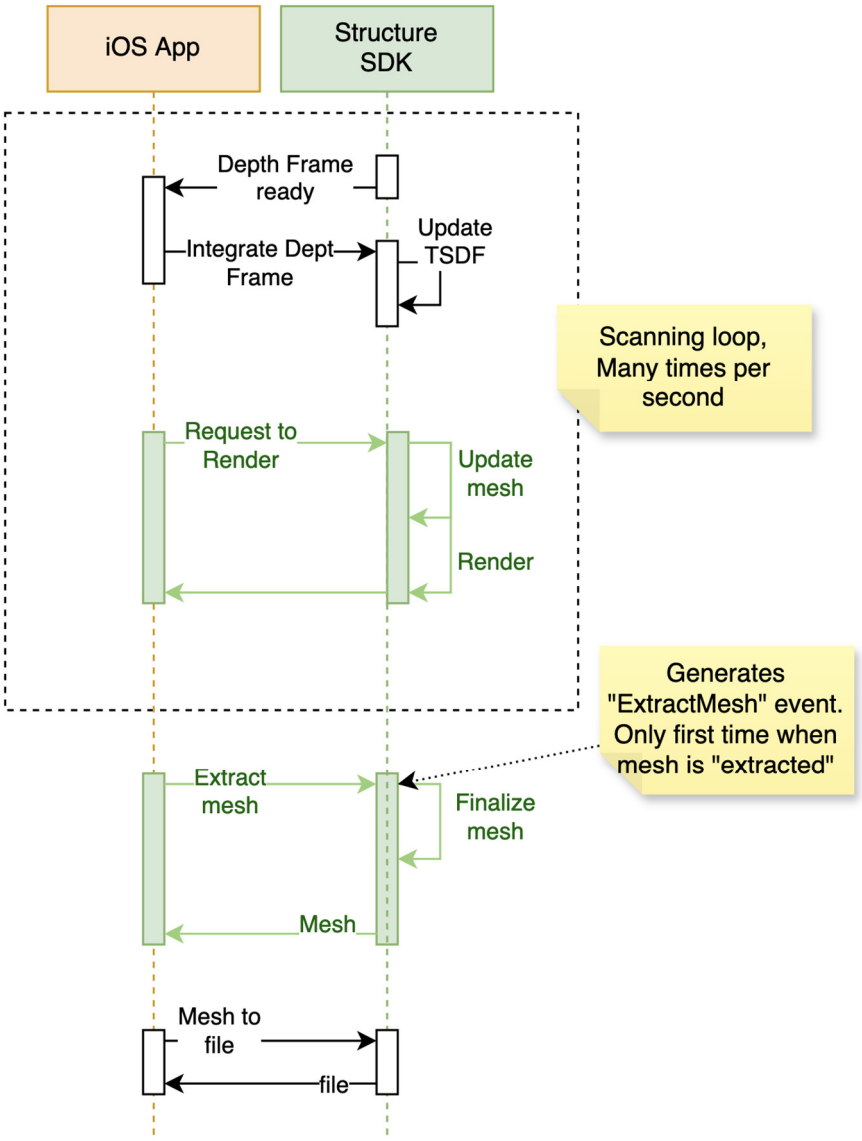

Figure 11: Structure SDK Mesh Extract Flow Diagram.

3.4. Mesh extract mode

The *mesh extract* mode, illustrated in Figure 11, prioritizes user accessibility and reduces technical complexity. Within this operational mode, the SDK assumes full responsibility for all visualization processes throughout the reconstruction phase. This integrated approach not only simplifies the user experience but also facilitates the collection of more precise and predictable usage metrics, offering valuable insights into application performance and user engagement.

Due to its streamlined nature and robust internal management of visualization, this mode proves highly suitable for the vast majority of applications currently in development or deployment. Consequently, it is strongly recommended as the default implementation for new projects, ensuring a smooth and efficient development cycle while maintaining high standards of data integrity and user satisfaction. Its intuitive design and automatic handling of intricate technical aspects make it an ideal choice for developers seeking to simplify their workflow and deliver reliable, user-friendly solutions.

4. Conclusion and Future Steps

The Structure Sensor 3 provides a significant leap in 3D scanning for anatomic scanning applications. Its improved robustness, increased immunity to sunlight, IP53 compliance and expanded battery life help meet the demands of many healthcare operations. New features exclusively available to the Structure Sensor 3, such as binning and confidence threshold, provide its users with the improved accuracy and flexibility required to meet a variety of human scanning applications, ranging from the detailed scanning of single body parts to the accurate scanning of the whole human body. Experimental results demonstrate maximum depth resolution of 0.08mm, typical noise variation of 0.27mm when 80cm away from the scanned object, and a typical 3D reconstruction RMSE of 0.29% when used with the Structure SDK to measure objects with dimensions similar to those of large human limbs (i.e., arms and legs).

In addition, the Structure SDK provides the accuracy and flexibility required for different healthcare applications. The traditional access-to-mesh method enables AR surgical applications, while the new mesh extract mode facilitates the accounting for body scanning applications – especially useful for scan-to-fabricate applications. Finally, the recent addition of foot anthropometrics demonstrates that the Structure SDK will continue to expand as a centralized platform for the capture and measurement of the human body for healthcare applications. In the future, users should expect the automated detection of landmarks and metrics to be extended to other parts of the human anatomy, starting with facial features.

4.1. Future Steps

The modularity and flexibility of the Structure SDK enables it to be ported to other operating systems than iOS, starting with Linux and Android, then expanding to Windows. Even though the complexity of the underlying 3D reconstruction algorithms makes this no trivial task, it has been facilitated by the continuous and gradual porting of the Structure SDK code base to C++, a versatile computer language that is efficiently compiled over multiple OS's without compromising execution speed even when multi-threaded. In addition, one should expect to see the ability to select multiple aggregate windows, given the benefits described in Section 2.5. Finally, one should expect to see the Structure SDK supporting more relevant sensors over time.

One should also expect the line of Structure Sensors to continue to grow, with potential improvements such as the use of USB3 communications for greater resolution, greater frame rates and lower latency – thus further improving the mobile scanning experience – and a variation in the baseline distances between the stereo cameras, thus enabling, for example, the scanning of smaller objects at closer distances or of larger environments at farther distances. One should also expect the use of continuously more powerful processors capable of handling the additional information bandwidth, and porting embedded edge-AI capabilities. Finally, one should keep in mind that the act of scanning a patient affords the clinician with the opportunity to capture more health data than simply the patient's exterior anatomy. As such, one should expect future sensors to be able to capture additional patient parameters, thus helping fulfill some of the promises of personalized healthcare, such as improved diagnostics and personalized therapy.

Acknowledgements

The authors sincerely thank Mehdi Partou, Megan Renny and Matthew D'Aguiar for performing several of the Structure Sensor 3 measurements required for validation. We also thank Jon Edwards, Anthony Swenson, Matthew Quick and Ji-ren Xu for their significant contribution to the sensor design and mass production.

References

[1] Ren, X., Fox, D., & Konolige, K. (2013, December). Change Their Perception: RGB-D Cameras for 3-D Modeling and Recognition. IEEE Robotics & Automation Magazine, December, 49–59.

[2] Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J., Kohli, P., Shotton, J., Hodges, S., & Fitzgibbon, A. (2011). KinectFusion: Real-time dense surface mapping and tracking. 2011 10th IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2011. https://doi.org/10.1109/ISMAR.2011.6092378

[3] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., & Fitzgibbon, A. (2011). KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. UIST'11 - Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology. https://doi.org/10.1145/2047196.2047270

[4] Popescu, C. R., & Lungu, A. (2014). Real-Time 3D Reconstruction Using a Kinect Sensor. Computer Science and Information Technology, 2(2), 95–99. https://doi.org/10.13189/csit.2014.020206

[5] Oranges, C. M., Madduri, S., Brantner, P., Msallem, B., Giordano, S., Benitez, B., Kalbermatten, D. F., Schaefer, D. J., & Thieringer, F. M. (2019). Three-dimensional Assessment of the Breast: Validation of a novel, simple and inexpensive scanning process. In Vivo, 33(3), 839–842. https://doi.org/10.21873/invivo.11548

[6] Redaelli, D. F., Gonizzi Barsanti, S., Fraschini, P., Biffi, E., & Colombo, G. (2018). Low-cost 3D devices and laser scanners comparison for the application in orthopedic centres. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42(2), 953–960. https://doi.org/10.5194/isprs-archives-XLII-2-953-2018

[7] Marr, D., & Poggio, T. (1976). Cooperative Computation of Stereo Disparity: A cooperative algorithm is derived for extracting disparity information from stereo image pairs. Science, 194(4262), 283–287.

[8] Cheng, C. H., Shen, C. C., Kao, H. Y., Hsieh, D. H., Wang, H. Y., Yeh, Y. W., Lu, Y. T., Chen, S. W. H., Tsai, C. T., Chi, Y. C., Kao, T. S., Wu, C. H., Kuo, H. C., Lee, P. T., & Lin, G. R. (2018). 850/940-nm VCSEL for optical communication and 3D sensing. In Opto-Electronic Advances (Vol. 1, Issue 3, pp. 1–11). Chinese Academy of Sciences. https://doi.org/10.29026/oea.2018.180005

[9] ASTM G-173-03, Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface (2020), accessed from https://www.nrel.gov/grid/solar-resource/spectra-am1.5, July 1, 2025.

[10] Skorka, O., & Joseph, D. (2014). CMOS digital pixel sensors: technology and applications. Proc.SPIE, 9060, 90600G. https://doi.org/10.1117/12.2044808

[11] Hamzah, R. A., & Ibrahim, H. (2016). Literature Survey on Stereo Vision Disparity Map Algorithms. Journal of Sensors, 2016(1), 8742920. https://doi.org/https://doi.org/10.1155/2016/8742920

[12] Rusinkiewicz, S., & Levoy, M. (2001). Efficient variants of the ICP algorithm. Proceedings of International Conference on 3-D Digital Imaging and Modeling, 3DIM. https://doi.org/10.1109/IM.2001.924423

[13] Silveira, P. (2021). *Human Limb Scanning with Improved Accuracy using the Structure Sensor Pro™*. https://s3.amazonaws.com/io.structure.assets/White+Paper+-+Human+Limb+Scanning+with+Improved+Accuracy_v1.0+.pdf

[14] Mulligan, J., Miftahul Hoque, K., Tokar, A., Gladyshev, D., & Silveira, P. E. X. (2025). Foot Landmark Detection with Structure. Accepted for publication In N. D'Apuzzo (Ed.), *Proceedings of 3DBODY.TECH - 16th International Conference and Exhibition on 3D Body Scanning and Processing Technologies*. Lugano, Switzerland, October 2025. https://doi.org/10.15221/25.18

[15] Lee, M., Woo, J., Se, ·, Peak, H., Kim, H. G., Woo, ·, Lim, S., Chung, J., Jee, ·, Lee, E., Jeoung, ·, Kim, H., Park, S., Kim, J. M., Jun, ·, & Lee, W. (2024). An exploratory clinical trial of preoperative non-invasive localization before breast-conserving surgery using augmented reality technology. 206, 31–44. https://doi.org/10.1007/s10549-024-07272-3

[16] Choi, S., Koo, K., Roh, T. H., & Park, C.-U. (2025). Precision Needle Localization Using Electromagnetic Tracking and AR Visualization in Ventricular Interventions. Accepted for publication in 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).