**Proceedings of the ASME 2023 18th International
Manufacturing Science and Engineering Conference
MSEC2023
June 12-16, 2023, New Brunswick, NJ, USA**

MSEC2023-104380

SINGLE SHOT 3D SHAPE MEASUREMENT OF NON-VOLATILE DATA STORAGE DEVICES

Badrinath Balasubramaniam
Department of Mechanical Engineering
Iowa State University
Ames, Iowa 50011

Beiwen Li[*]
Department of Mechanical Engineering
Iowa State University
Ames, Iowa 50011
Email: beiwen@iastate.edu

ABSTRACT

The value of electronic waste at present is estimated to increase rapidly year after year, and with rapid advances in electronics, shows no signs of slowing down. Storage devices such as SATA Hard Disks and Solid State Devices are electronic devices with high value recyclable raw materials which often goes unrecovered. Most of the e-waste currently generated, including HDDs, is either managed by the informal recycling sector, or is improperly landfilled with the municipal solid waste, primarily due to insufficient recovery infrastructure and labor shortage in the recycling industry. This emphasizes the importance of developing modern advanced recycling technologies such as robotic disassembly. Performing smooth robotic disassembly operations of precision electronics necessitates fast and accurate geometric 3D profiling to provide a quick and precise location of key components. Fringe Projection Profilometry (FPP), as a variation of the well-known structured light technology, provides both the high speed and high accuracy needed to accomplish this. However, Using FPP for disassembly of high-precision electronics such as hard disks can be especially challenging, given that the hard disk platter is almost completely reflective. Furthermore, the metallic nature of its various components make it difficult to render an accurate 3D reconstruction. To address this challenge, We have developed a single-shot approach to predict the 3D point cloud of these devices using a combination of computer graphics, fringe projection, and deep learning. We calibrate a physical FPP-based 3D shape measurement system and set up its digital twin using computer graphics. We capture HDD and SSD CAD models at various orientations to generate virtual training datasets consisting of fringe images and their point cloud reconstructions. This is used to train the U-NET which is then found efficient to predict the depth of the parts to a high accuracy with only a single shot fringe image. This proposed technology has the potential to serve as a valuable fast 3D vision tool for robotic remanufacturing and is a stepping stone for building a completely automated assembly system.

1 INTRODUCTION

The amount of electronic waste currently produced is worth \$57 billion, of which Europe is the leading collector. Only a small fraction of this (20%) is recycled and recovered sustainably, which can offset tonnes of CO_2. E-waste is described as all waste from discarded electrical and electronic equipment and can include devices ranging from large household appliances such as dishwashers and dryers to consumer electronics such as laptops, tablets and mobile phones. Global E-waste has only been increasing rapidly year after year. The E-waste generated in 2019 was 21% more than the E-waste generated in 2015. More than 80% of this was either dumped illegally or burnt openly. Recycling and re-using E-waste properly can lead to the recovery of a wide range of valuable raw materials and the fact that less than 20% of this is recovered sustainably poses a serious environmental threat. E-waste recovery and collection can be designated as part of the formal or informal sector. The process of formal recy-

---

[*] Address all correspondence to this author.

cling is regulated and involves specially built facilities equipped with highly specialized machinery that allows salvageable materials to be extracted safely. Informal waste recycling is unofficial and unregulated. Moreover, the majority of e-waste recycling is done by the informal sector. The processes and practices used to recycle and reuse e-waste by the informal recycling sector, results in hazardous chemicals being released to the environment. This, along with the sheer volume of E-waste being dumped in landfill every year, poses a serious threat to the environment and the safety of workers in the recycling sector [1, 2].

Hard disk drives (HDDs) and Solid State Drives (SSDs) contain a lot of valuable metals such as Gold, Silver, Aluminum, Palladium, and Neodymium and that can be recovered through recycling. However, a big portion of these are still improperly landfilled with the rest of municipal solid waste. One of the main disincentives to their recovery is the lack of recovery infrastructure. Another problem plaguing the recycling of HDDs is incomplete recycling. Typically, the recovery of HDDs involves simply taking apart the printed circuit boards and recovering non ferrous metals from it such as Gold, Silver, Palladium. Neodymium, a rare earth metal used in HDD magnets, often goes overlooked. It is not in abundant supply yet its recovery rate is less than 1%. This is an example of material leakage, wherein material from a component is not recovered for re-use. This could cause material shortage and supply chain issues in the future. The reason for material leakage in HDDs is because hard disks are difficult to disassemble by design. This, combined with the labor shortage in the recycling industry, the insufficient recycling infrastructure and the environmental threat caused by dumping the bulk of unrecovered e-waste into landfill highlights the need for better infrastructure in the form of automation and robotic disassembly of these components [3].

Due to the high dexterity as required by disassembling of HDDs and SSDs, it is crucial to develop computer vision tools to provide visual input to a disassembly robot such that the robot can accurately ”see” the parts being disassembled. To achieve this, it is crucial to conduct fast and accurate 3D geometric profiling such that the key components (e.g., screws, platter, flange, etc.) can be precisely located with minimum latency induced. Taking this into account, researchers have tried to embrace 3D sensors for assisting robot assembly operations. Niedermayr et al. [4] evaluated the performance of the Azure Kinect sensor, a time of flight based sensor, to assist assembly operations. It was found that this sensor had limitations, which included underexposure or overexposure due to object reflectivity and multi-path errors due to transmittance of the material. In addition to these drawbacks, time of flight sensors in general are only accurate to the millimeter or centimeter [5]. While this is still acceptable for assembly assistance tasks, the limitations render it unsuitable for precision disassembly tasks, especially for complex electronic components like HDDs and SSDs. Ogun et al. [6] focused on flexible automated assembly of clearance fit machine components using 3D vision and the recognition and estimation of the poses of the components are achieved by matching the CAD models of the parts with the acquired point cloud data of the scene. They used a shaft and three rings for the assembly task, wherein the rings were placed on the tapered section of the shaft by a vision guided robotic arm with a laser scanner mounted on it. However, the limitation of this method is the fact that the laser scanner performs line scanning sequentially and hence cannot be used in high speed assembly tasks. Yildiz et al. [7] developed a deep learning based visual detection scheme for hard drive disassembly. Their setup consists of a tilting table that holds the device at 0°or 45°. Two cameras, one monocular camera which acquires RGB images at 3.5 FPS and another RGB-D stereo camera with depth sensing accuracy of 0.06 cm, were positioned 45°to each other such that they could capture top down images of the hard disk. Their pipeline takes in an RGB image and a point cloud. The RGB images are fed to the part detection and screw detection modules which uses deep neural networks to segment parts and detect screws. Gap detection relies on the point cloud acquired from the RGB-D stereo camera. The limitation of this method is that the stereo camera can provide an inaccurate point cloud when it encounters shiny and reflective parts. Additionally, the depth accuracy of the point cloud can be improved upon with structured light techniques which would be more suitable as a vision system for high precision disassembly tasks.

In the domain of 3D computer vision, compared to other 3D vision technologies such as time-of-flight, laser scanning or binocular stereo vision, structured light technology has the merit of simultaneous whole-area scanning with high accuracy (e.g., 0.1 mm or higher), and thus is by far the most suitable technology that can be used by a disassembly robot. Fringe Projection Profilometry (FPP) is a variation of the structured light method that involves projecting and capturing continuous sinusoidal fringe patterns and can provide the most accurate 3D reconstruction results compared to other discrete structured light pattern codification technologies. However, an important drawback of FPP is that it can be inaccurate when the object to be scanned has highly specular and reflective components. This is because of the loss of information due to saturation of the fringe image in these parts. This holds true in the case of Hard Disks. The platter of the Hard Disk, which is machined to a high precision, is perfectly reflective which makes it impossible to perform traditional FPP-based 3D shape measurements. To measure such a shape with FPP, special surface treatment such as spray paintings are needed before any patterns are projected onto the object surfaces [8]. Moreover, typically multiple fringe patterns needs to be properly projected by the video projector and captured by the camera, which makes the measurements speed compromised. In this paper, we tackle these limitations of FPP by using a deep learning and computer graphics based single shot approach to reconstruct the 3D shape of HDD and SSD components instead. This method has the advantage of predicting the depth map of the device with a single

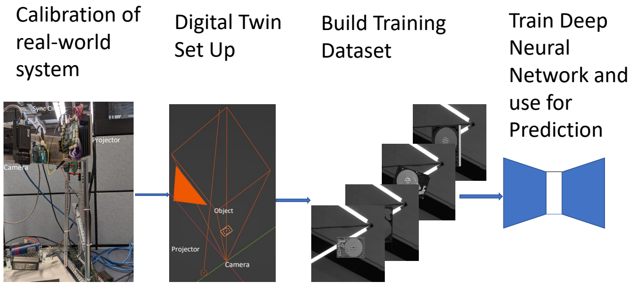

FIGURE 1: SCHEMATIC OF FRAMEWORK

image thereby making it a fast 3D vision tool that can be used by a disassembly robot. Moreover, our approach is further able to predict the geometric profile of the conventionally hard-to-measure reflective platter surfaces along with the outlines of the key disassembly components (e.g., screws) with reasonable accuracy. A brief overview of our approach, as outlined in Fig. 1, is given here:

1. Calibrate the real-world FPP system and compute its intrinsic and extrinsic matrices for both the camera and the projector
2. Determine the location and orientation of the camera and projector based on its extrinsic matrices
3. Set up the virtual camera and projector system in the CG-based platform *Blender* [9] based on the location and calibration parameters computed
4. Build a training data set consisting of Hard Disk and SSD models using virtual scanning in *Blender*
5. Train deep neural network which takes in a single fringe image as input and outputs depth map of the object being scanned
6. Utilize the trained deep neural network to predict the geometric profile of the device with a single-shot fringe image taken from the real-world system

2 CALIBRATION

Our physical set up of the FPP system consists of a complimentary-metaloxide-semiconductor (CMOS) camera (model: FLIR Grasshopper3 GS3-U3-23S6M-C) for image acquisition and a DLP LightCrafter 4500 projector for projecting fringe patterns. The camera resolution was set to 544 x 514 pixels and the projector was set to 912 x 1140 pixels. The calibration process was done in accordance with the procedure outlined by Li et al. [10].

We utilize a pinhole model for both the camera and projector. Figure 2 is an illustration of a typical FPP system. The camera is described by the extrinsic and intrinsic matrices. The extrinsic matrix describes the position of the camera in world coordinates, namely its x,y,z coordinates and its rotations R_x, R_y, R_z about the three axes. The intrinsic matrix contains parameters such as the focal lengths in the principal directions of the image plane, the principle points, and the skew factor. The projector is described by similar parameters.

A point on the world coordinate will be projected onto the imaging plane at location (u,v) where s is a scaling factor. This model can be further rewritten as

$$s^c I^c = A^c [R^c, t^c] X^w, \tag{1}$$

where $I^c = [u^c, v^c, 1]^T$ is the coordinates of the image point in the image coordinate system, $X^w = [x^w, y^w, z^w, 1]^T$ is the object point in the world coordinate system, and R_c and t_c are rotation and translation matrices respectively between the world coordinate system and the camera coordinate system. A^c, the intrinsic parameter matrix of the camera, is

$$A^c = \begin{bmatrix} \alpha^c & \gamma^c & u_0^c \\ 0 & \beta^c & v_0^c \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where (u_0^c, v_0^c) are the coordinates where the optical axis intersects the imaging sensor plane, α^c and β^c are focal lengths along the principal axes of the image plane and γ^c is the skewness parameter.

We align the world coordinates with the camera coordinates. Thus the rotation matrix R_c of the camera is equivalent to the identity matrix and the translation vector is set to the origin.

$$R^c = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; t^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3}$$

The model for the projector can also be written as:

$$s^p I^p = A^p [R^p, t^p] X^w, \tag{4}$$

where $I^p = [u^p, v^p, 1]^T$ are the corresponding coordinates of the image point in the projector coordinate system, and s^p is a scale factor. R_p and t_p are rotation and translation matrices respectively between the world coordinate system and the projector coordinate system. Our goal when calibrating the system is to estimate the intrinsic and extrinsic projection matrices of the camera and the projector. The intrinsic matrices of the camera can

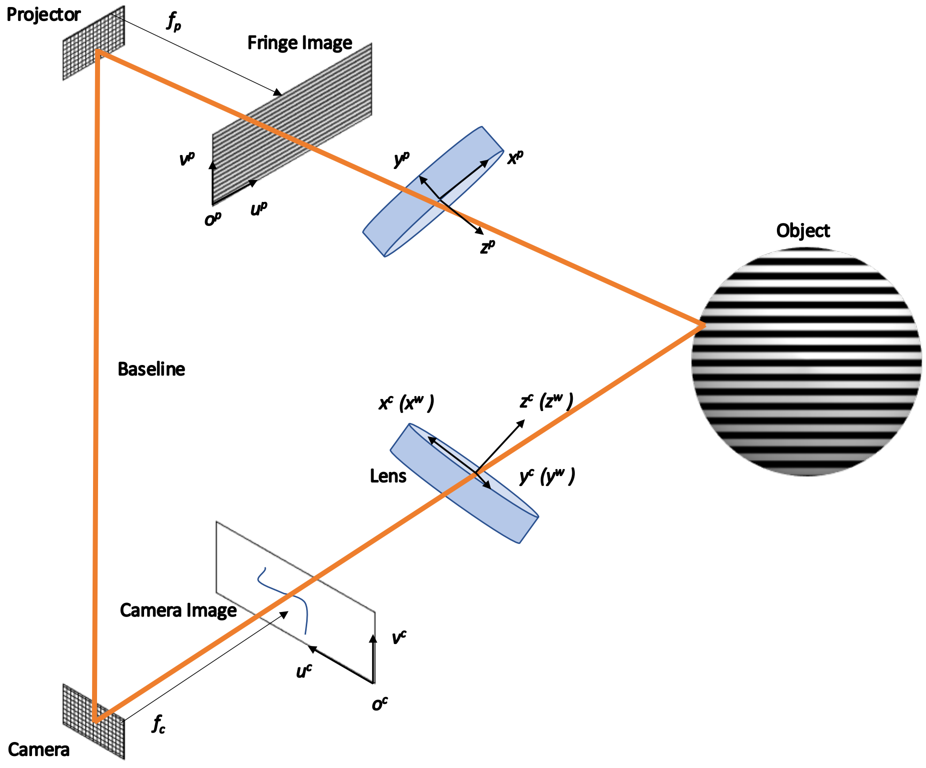

FIGURE 2: SCHEMATIC OF FPP SYSTEM

be calibrated by the method suggested by Li et al. [10] using a black/white circle board as a target, as shown in Fig. 3. In the figure, the diameter of each white circle is 10 mm and the distance between the circle centers is 20 mm, with the pattern contained in a rectangle spanning 100 mm × 90 mm. Now, the square checkerboard is a valid option that has been extensively used for calibration. However, the projector image created in this case is a discrete grid which involves quantization, leading to non-negligible error. An additional problem with the checkerboard is that the phase around the corner may be distorted. It has been demonstrated that the phase is prone to sharp changes from one square to another, which could introduce a mapping error for the corresponding projector pixel determination since the vertices of the squares captured by the camera is not always exact. To circumvent these problems, Li et al. [10] proposed to use the white circles on black background and further only circle centers are mapped to the projector image without generating the entire projector image. This avoids the projector image discretization error. The use of white circles avoids color changes at the circle centers , resulting in a smooth phase. This is the reason we pick this calibration method.

A total of 18 poses of the calibration board are captured and OpenCV's camera calibration toolbox is used to determine the circle centers and the intrinsic matrix A^c of the camera. A total of 52 images are captured per pose. The projector circle centers are then determined by using the horizontal and vertical phase maps and mapping the circle centers using OpenCV's camera calibration toolbox. The projector intrinsic matrix is then calibrated using the OpenCV camera calibration toolbox with the projector circle centers as input. The extrinsic parameters that define the transformation from the camera coordinate system to the projector coordinate system is computed using the OpenCV stereo calibration toolbox. The intrinsic matrix for the camera was calculated as:

$$A^c = \begin{bmatrix} 1526.58 & 0 & 252.65 \\ 0 & 1528.16 & 274.49 \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

The intrinsic matrix for the projector was calculated as:

$$A^p = \begin{bmatrix} 1100.67 & 0 & 439.45 \\ 0 & 2212.52 & 1188.38 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

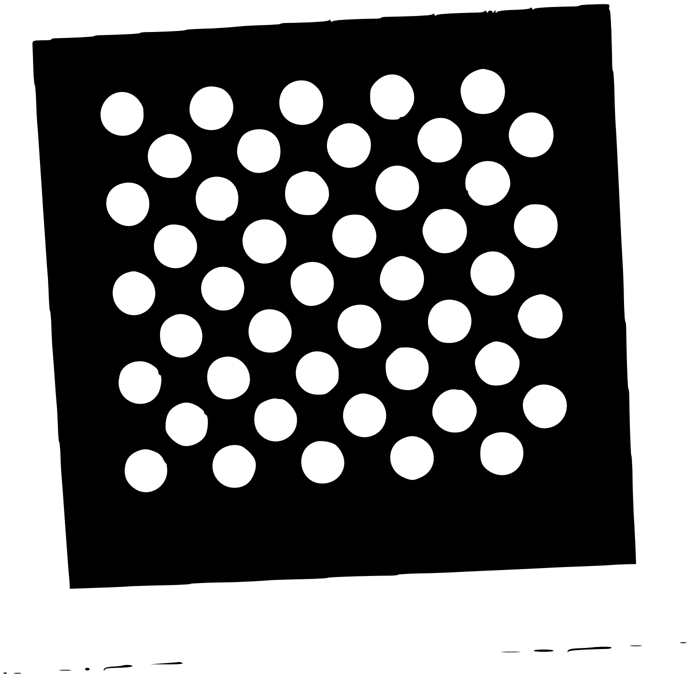

FIGURE 3: 5 X 9 CIRCLE BOARD USED FOR CALIBRATION

The rotation matrix for the projector was calculated as:

$$R^p = \begin{bmatrix} -0.999 & 0.003 & 0.004 \\ -0.0032 & -0.999 & -.00032 \\ 0.004 & -0.0032 & 0.999 \end{bmatrix} \tag{7}$$

The translation matrix for the projector was calculated as:

$$t^p = \begin{bmatrix} 0.662 \\ -142.058 \\ -43.311 \end{bmatrix} \tag{8}$$

It is also important to discuss the reprojection error associated with this calibration. In our case, the projector error was calculated to be 0.2252 pixels and the stereo error was calculated to be 0.2119 pixels. This re projection error is caused by the uncertainty of finding circle centers. The calibration process involves reorienting and repositioning the calibration target such that all the poses cover the entire field of view. When the angle between the calibration target plane and camera sensor plane is larger, the camera imaging pixels are no longer square or small, and this makes it difficult to locate circle centers accurately.

3 Mapping digital twin FPP system in Blender

Zheng et al. [11] developed a virtual training pipeline based on the digital twin of the structured light system using the open-source software Blender. The digital twin system seeks to replicate the structured light system in a virtual environment. The digital twin system requires the following parameters as input:

1. The location and orientation of the camera
2. The intrinsic parameters of the camera
3. The location and orientation of the projector
4. The intrinsic parameters of the projector

The world coordinate system is aligned and oriented with the camera coordinate system. Thus the location of the camera is at $(0,0,0)$. The intrinsic parameters of the camera are calculated from the calibration of the physical system. The location and orientation of the projector still needs to be calculated. We know that the extrinsic parameter matrix transforms the camera and projector coordinate systems to the world coordinate system. Since the camera coordinate system is aligned with the world coordinate system, we obtain:

$$0 = R^p P + t^p, \tag{9}$$

where P is the location of the projector in world coordinates, R^p is the rotation matrix of the projector with respect to the world coordinates, and t^p is the translation matrix of the projector.

$$P = -R^{p^{-1}} t^p. \tag{10}$$

The rotation angles of the projector about the x,y,z axes can be computed from the rotation matrix R_p. The intrinsic parameters of the projector are input from the computation during calibration.

4 FRINGE IMAGE GENERATION

We collect nine SATA HDD CAD models and three NVMe SSD CAD models from the internet. These models did not contain material property information such as metallicity and roughness which determine how light interacts with the object. These parameters were manually input to the textures of these models and were modeled to closely approximate that of a real-world component. Each of the CAD models is rotated about each of the x,y,z axes leading to a total number of 396 different orientations combined of all of these models. Three high frequency fringe images are captured per orientation for the three-step enhanced two frequency phase shifting method [12], with a total of 1188 high frequency fringe images captured this way. This comprises our large input dataset that we use for network training. For each of the 396 orientations captured above, we pick one of the high frequency images, resulting in one high frequency image per orientation with a total of 396 images. This comprises our small input dataset that we use for network training. Additionally, we generate a background mask for each image. However, performing FPP on the perfectly reflective Hard Disk Platter is still a problem due to the loss of fringe information in the platter region. To combat this, we collect two sets of images. One set of images consists of fringe patterns projected on the devices with perfectly reflective platters as in Fig. 4. Another set of images consists of fringe patterns projected on the devices with the roughness of the hard disk platters artificially increased, which results in fringe patterns being visible in the platter region as in Fig. 5. 3D reconstruction is performed on the latter set of images to retrieve accurate geometric profiles of the devices used as ground truth for the deep neural network in the following steps. The former set of images is used as input to the neural network.

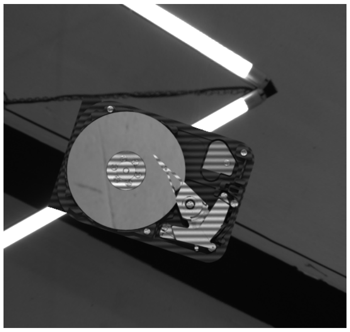

FIGURE 4: FRINGE IMAGE OF SAMPLE HDD WITH PERFECTLY REFLECTIVE PLATTER

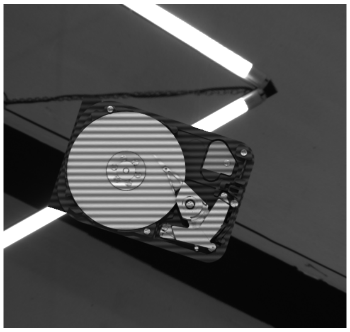

FIGURE 5: FRINGE IMAGE OF SAMPLE HDD WITH ARTIFICIALLY ROUGH PLATTER

It is important to note that suppressing the specular reflection of the highly reflective area in the disk would indeed be a challenge if instead of using this a we were to simply perform traditional FPP on the disk and the associated reconstruction. However, we are circumventing that challenge by using a neural network trained on the hard disk images with the reflective platter. The neural network simply handles the prediction of the disk geometry and is able to predict the associated 3D geometry from the input fringe image of the hard disk, despite the platter being highly reflective. We shall demonstrate this in the subsequent sections.

5 3D RECONSTRUCTION OF VIRTUAL DATA

5.1 Principles of FPP

The typical setup for FPP includes a camera, projector and an object in the scene. Fringe patterns projected by the projector onto the object are captured by the camera. These fringe patterns are distorted based on the geometry of the object. The geometric profiles of the objects can be retrieved through the analysis of the carrier phase information, which can be computed by a phase shifting algorithm. It is important to note that this technique lends itself to a depth uncertainty between a range of sub 0.01 mm to a little over 0.1 mm depending on the size of the device being imaged [13]. Additionally, because our set-up uses a digital projector, the computer generated fringe patterns projected have high accuracy and high repeatability [14]. Furthermore, several physical set ups involving fringe projection profilometry using digital projection have demonstrated high measurement precision [15, 16].

The three step phase shifting algorithm, which involves using three steps each at a phase shift of $2\pi/3$, utilizes the minimum number of steps needed to solve for the phase uniquely, and is thus useful for high-speed 3D imaging applications. The three phase shifted fringe images, $I_1(x,y)$, $I_2(x,y)$, and $I_3(x,y)$, can be mathematically described as

$$I_1(x,y) = I'(x,y) + I''cos\left[\phi(x,y) - \frac{2\pi}{3}\right], \tag{11}$$

$$I_2(x,y) = I'(x,y) + I''cos[\phi(x,y)], \tag{12}$$

$$I_3(x,y) = I'(x,y) + I''cos\left[\phi(x,y) + \frac{2\pi}{3}\right], \tag{13}$$

where $I'(x,y)$ is the average intensity, $I''(x,y)$ is the fringe or intensity modulation, and $\phi(x,y)$ is the wrapped phase.

$$\phi(x,y) = tan^{-1}\left[\frac{\sqrt(3)(I_1 - I_3)}{2I_2 - I_1 - I_3}\right] \tag{14}$$

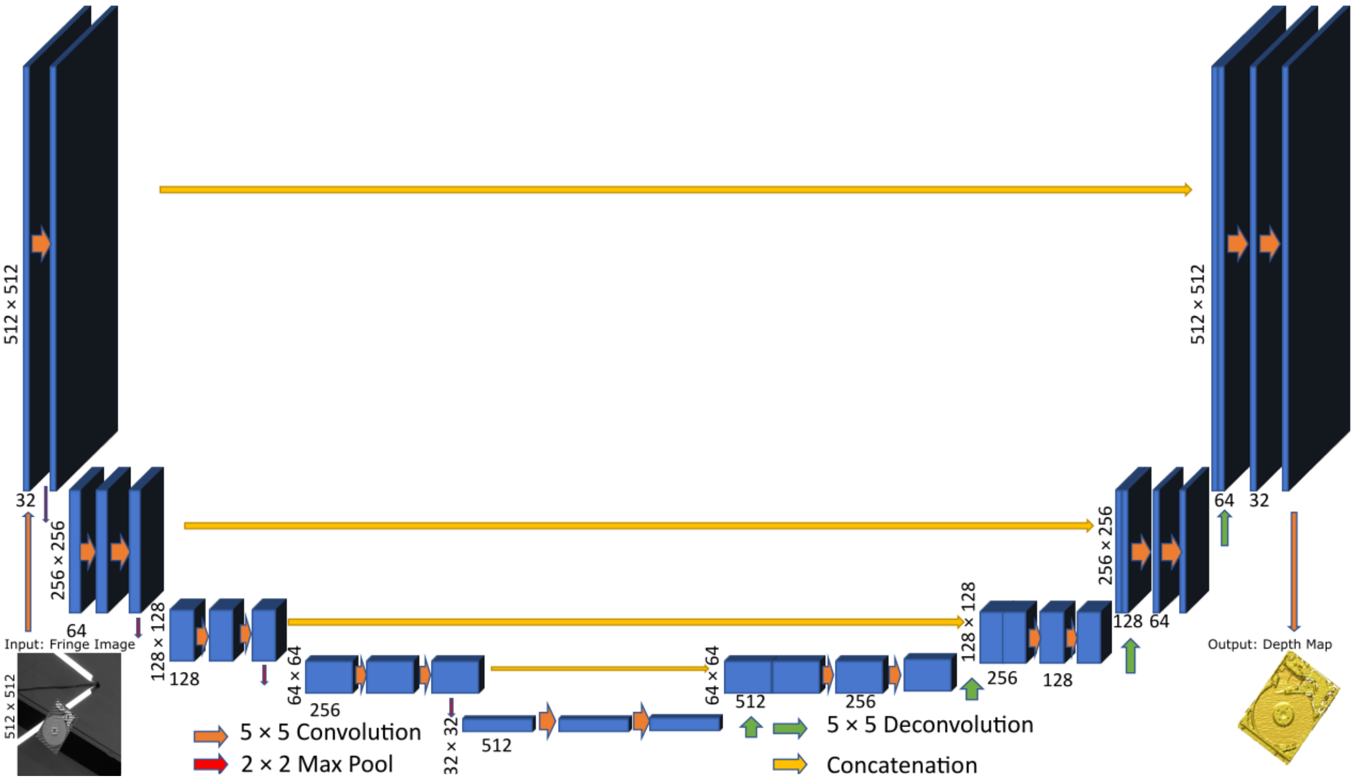

FIGURE 6: U-NET ARCHITECTURE

Simultaneously solving these equations gives us the wrapped phase map. This wrapped phase map is a relative phase map that varies from $[-\pi, \pi]$ and needs to be unwrapped to obtain a continuous phase map [17].

$$\phi = \Phi \quad (mod\ 2\pi) \tag{15}$$

where Φ is the continuous phase map.

For each point (u^c, v^c) on the camera image plane, we can compute its absolute phase, use that to identify the corresponding vertical stripe u^p in the projector plane. Since we have already calibrated the system and computed its intrinsic and extrinsic matrices, from equations 1 and 4, we will need to solve the following linear equations:

$$f_1(x^w, y^w, z^w, u^c) = 0 \tag{16}$$

$$f_2(x^w, y^w, z^w, v^c) = 0 \tag{17}$$

$$f_3(x^w, y^w, z^w, u^p) = 0 \tag{18}$$

The above three equations enable us to solve for x^w, y^w, z^w which represent the world coordinates of the point on the object.

5.2 Generation of the Ground Truth 3D Data

In our case, it is necessary to perform 3D reconstruction to retrieve a depth map for each of the device orientations that we can use to train the deep neural network in the following step. However, performing FPP on the perfectly reflective Hard Disk platter is still a challenge due to the loss of fringe pattern information in the platter region (as shown in Fig. 4). To combat this, we collect an additional set of fringe images of the hard disks in the data set wherein we artificially increase the roughness of each hard disk platter, making it non reflective and ensuring that the fringe information in that area is not lost. We use the enhanced two-frequency three step phase shifting method [12] to compute the depth map for each orientation. This allows us to perform FPP accurately on the entire device and retrieve accurate geometric profiles of these virtual devices that we use as ground truth use for training the deep network. The input to the neural network is the fringe image of the hard disk with the smooth platter to mimic the scenarios in the real world.

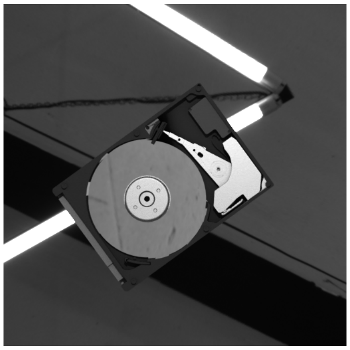

FIGURE 7: PHOTOGRAPH IMAGE OF SAMPLE HDD

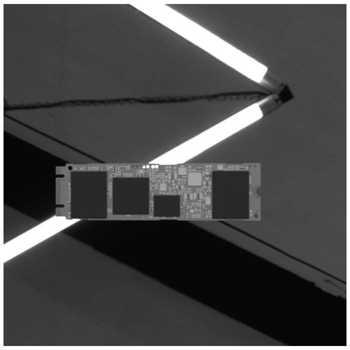

FIGURE 8: PHOTOGRAPH IMAGE OF SAMPLE SSD

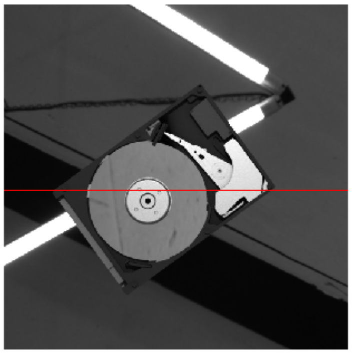

FIGURE 9: CROSS SECTION OF HDD

6 TRAINING THE SINGLE-SHOT DEEP NEURAL NETWORK

We train the network using the U-NET adopted by Zheng et al. [11]. Its architecture is depicted in Fig. 6. The network consists of an encoder and a decoder with residual connections. We use 5×5 kernels and 2×2 max pooling layers with a stride of 2 in both the encoder and decoder block. The convolutional layers in the encoder block down-sample the image, extracting vital feature information from it, and the deconvolutional layers in the decoder block up-sample the image. We use a Rectified Linear Unit (ReLU) activation function in each convolutional layer. We use the RMSE loss function and the RMSProp Optimizer. To ensure that the network learns useful information about the object and not the background, the background pixels were not used in the loss computation. To prevent overfitting, we use drop out. The network takes in a 512×512 Fringe Image as input and predicts the depth map as output. The network is trained on two different data sets. One data set consists of using just one fringe image per orientation. As a result, this data set consists of a total of 396 images. The other dataset consists of six fringe images per orientation. Each of these is used in the neural network as input with the ground truth being the same point cloud for all six. Thus, this dataset consists of a total of 1188 images. We used the same 70%-15%-15% train, validation and test split for both cases. We trained the network for a total of 17,100 epochs and 2750 epochs with the smaller and larger data-set respectively on one NVIDIA Quadro RTX 5000 GPU. We used a batch size of 2 for both datasets and used the pre-trained model from the smaller data-set to accelerate training for the larger data set.

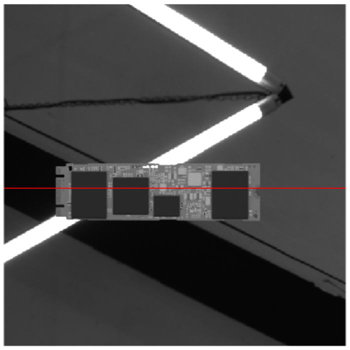

FIGURE 10: CROSS SECTION OF SSD

7 RESULTS

Table 1 shows the average training, validation and test RMSE losses by data set achieved by the network after training. We can see that the network is able to achieve an appreciable RMSE loss with both the data-sets. The training loss for the larger dataset is 2.52mm and the validation and test losses are 0.85mm and 0.914mm respectively. The training loss observed in the smaller dataset is 2.86mm, with the validation and test losses being 0.914mm and 0.998mm. The training loss is significantly higher because of the use of dropout on every single layer

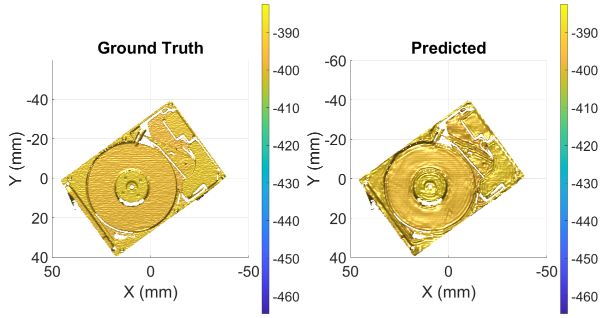

FIGURE 11: DEPTH MAP OF HDD- RMSE ERROR 0.8314MM

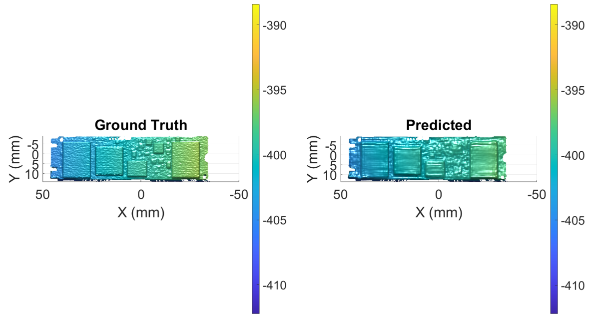

FIGURE 12: DEPTH MAP OF SSD- RMSE ERROR 0.9028MM

of the network during training. During testing and validation, dropout is turned off and due to this, all of the network's nodes are used for prediction, resulting in a lower loss value.

Figure 11 and Fig. 12 show the depth map of a sample HDD and SSD from the validation set (shown in Figure 7 and Figure 8) taken from the larger data set. We observe that the RMSE error is 0.8314mm and 0.9028mm for these individual cases respectively. In the case of the hard disk, we see that the network is able to make reasonably accurate predictions on the platter. Furthermore, faint outlines of the screws in the original prediction are visible in this picture. In the case of the SSD, the bumps and other miscellaneous features are predicted accurately by the network. These features are important for subsequent robotic disassembly that may be applied.

Figure 15 and Fig. 16 show the depth map of the same HDD and SSD from the validation set taken from the smaller data set. We observe that the RMSE error is 0.8059 mm and 1.1606 mm for these individual cases respectively. In the case of the hard disk, we see that the network is **also** able to make reasonably accurate predictions on the platter and additionally we see faint outlines of the screws. The geometric profile of the SSD is also predicted accurately by the network. These results show that even with reduced size of training dataset, the network is still capable of predicting depth maps with reasonable quality.

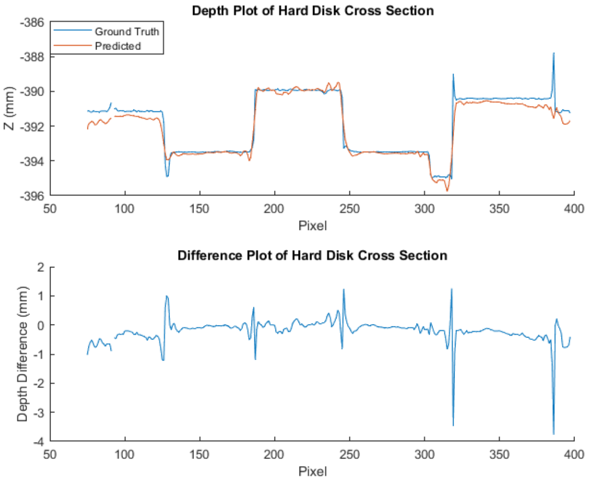

FIGURE 13: CROSS SECTION AND DIFFERENCE PLOT OF HDD

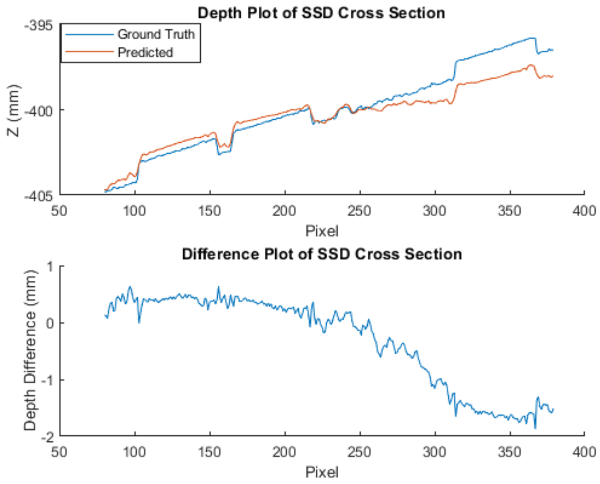

FIGURE 14: CROSS SECTION AND DIFFERENCE PLOT OF SSD

To help further demonstrate the comparison between the ground truth and the neural network's prediction of the geometric profile, we have attached depth plots of cross section of each of the profiles as well as the difference plots of each of both the

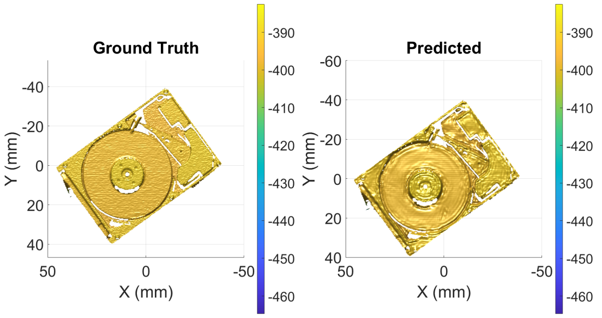

FIGURE 15: DEPTH MAP OF HDD- RMSE ERROR 0.8059MM

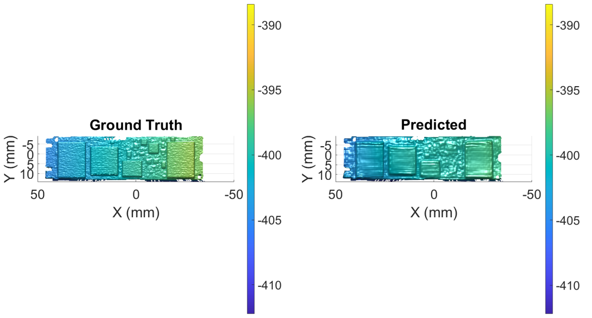

FIGURE 16: DEPTH MAP OF SSD- RMSE ERROR 1.1606MM

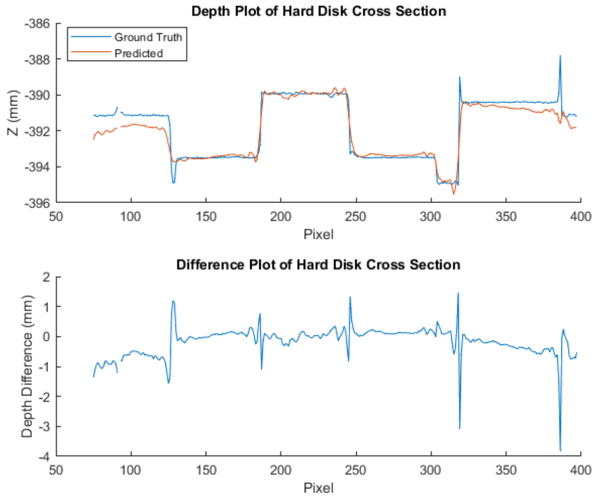

FIGURE 17: CROSS SECTION AND DIFFERENCE PLOT OF HDD

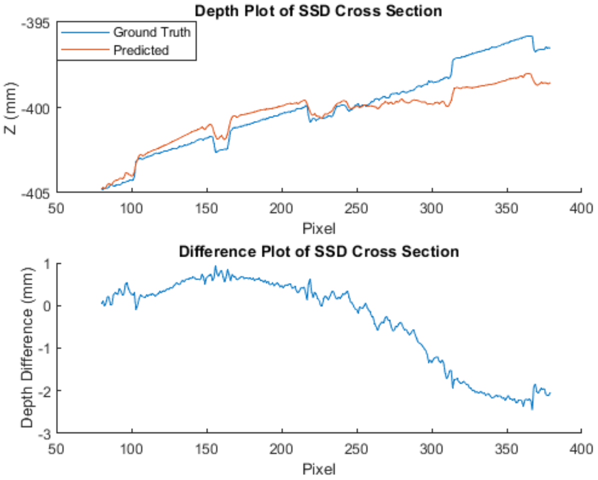

FIGURE 18: CROSS SECTION AND DIFFERENCE PLOT OF SSD

hard disk and SSD demonstrated in Fig. 7 and Fig. 8. Fig. 9 and Fig. 10 show the cross sections selected for this comparison. Fig. 13 and Fig. 14 show the depth and difference plots for the hard disk and the SSD respectively for the larger data set. Fig. 17 and Fig. 18 show the depth and difference plots for the hard disk and the SSD respectively for the smaller data set. In the case of the hard disk, we observe from these plots that for the majority of the cross section, the accuracy of the neural network is within a millimeter of the ground truth. We notice that the areas of maximum error occur when there is an abrupt change of depth in the cross section. In the case of the SSD, we notice that the accuracy of the network dips below the sub millimeter range as the cross section approaches the final raised portion along the X axis, corresponding to approximately the last fifth of the cross section.

It is important to note that these results are valid when the U-Net model is tested on the FPP system that it was calibrated with. As shown previously by Zheng et al. [11], testing the model on data acquired from an FPP system calibrated differently may lead to erroneous results.

From these figures, we observe that in both cases, the hard disk platter, on which it is impossible to perform traditional FPP based 3D reconstruction, is predicted with reasonable accuracy. Faint outlines of the locations of screws in both cases are visible as well, although the model does a marginally better job at overall accuracy with the larger data set.

We would like to clarify that the hard disk platter has a surface roughness of few tens of nanometers, and this method will

TABLE 1: AVERAGE RMSE LOSS

Data-set	Avg. Train Loss	Avg. Validation Loss	Avg. Test Loss
1188 Images	2.462mm	0.847mm	0.914mm
396 Images	2.886mm	0.914mm	0.998mm

not be able to capture it to that level of detail. However, it can still predict the overall geometric profile of the hard disk to an aggregate sub millimeter RMSE accuracy, as shown in the figures and in Table 1, which can be valuable for robotic disassembly.

8 CONCLUSION AND FUTURE WORK

We have shown in this work the efficacy of our proposed deep learning + computer graphics method in conducting single-shot 3D shape measurement for conventionally hard-to-measure electronic devices such as HDDs and SSDs. We map a calibrated FPP system using a virtual computer graphics-based platform Blender and generate a completely virtual dataset consisting of virtually 3D scanned CAD models of HDDs and SSDs imported from the internet. We have used this training data to build the neural network that can predict the geometric profile of these non volatile storage devices using a single shot. Compared to traditional FPP methods that require multiple fringe images for 3D reconstruction, our proposed method requires only a single shot image as input and is therefore faster than conventional FPP methods. In the meantime, the proposed method also addresses the limitation of conventional FPP when measuring highly reflective surfaces such as the disk platter. It is able to predict the geometric profile of the virtual hard disk with reasonable accuracy, including the profile of the platter. Given the ability to map high quality geometric details of HDDs and SSDs, the proposed method can be a valuable 3D machine vision tool for future applications such as robotic disassembly.

It is worth mentioning that in this work, all tests were conducted using virtual CAD models of HDDs and SSDs. Future work will be conducted to test the effectiveness of the proposed deep learning model on real world HDDs and SSDs such that it can be used for industrial sensing and measurements.

ACKNOWLEDGMENT

This research was funded by National Science Foundation Directorate for Engineering (CMMI-2132773). The views expressed here are those of the authors and are not necessarily those of the NSF.

REFERENCES

[1] Shahabuddin, M., Uddin, M. N., Chowdhury, J. I., Ahmed, S. F., Uddin, M. N., Mofijur, M., and Uddin, M. A., 2022. “A review of the recent development, challenges, and opportunities of electronic waste (e-waste)”. *International Journal of Environmental Science and Technology*, May.

[2] Perkins, D. N., Brune Drisse, M.-N., Nxele, T., and Sly, P. D., 2014. “E-Waste: A Global Hazard”. *Annals of Global Health,* ***80***(4), July, pp. 286–295.

[3] Sabbaghi, M., Cade, W., Olson, W., and Behdad, S., 2019. “The Global Flow of Hard Disk Drives: Quantifying the Concept of Value Leakage in E-waste Recovery Systems”. *Journal of Industrial Ecology,* ***23***(3), June, pp. 560–573.

[4] Niedermayr, D., and Wolfartsberger, J., 2022. “Analyzing the potential of a time-of-flight depth sensor for assembly assistance”. *Procedia Computer Science,* ***200***, Jan., pp. 648–659.

[5] Tychola, K. A., Tsimperidis, I., and Papakostas, G. A., 2022. “On 3D Reconstruction Using RGB-D Cameras”. *Digital,* ***2***(3), Sept., pp. 401–421. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[6] Ogun, P. S., Usman, Z., Dharmaraj, K., and Jackson, M. R., 2015. “3D vision assisted flexible robotic assembly of machine components”. In Eighth International Conference on Machine Vision (ICMV 2015), Vol. 9875, SPIE, pp. 342–348.

[7] Yildiz, E., Brinker, T., Renaudo, E., Hollenstein, J., Haller-Seeber, S. M., Piater, J., and Wörgötter, F., 2020. *A Visual Intelligence Scheme for Hard Drive Disassembly in Automated Recycling Routines*. Nov.

[8] Suresh, V., Liu, W., Zheng, M., and Li, B., 2021. “High-resolution structured light 3D vision for fine-scale characterization to assist robotic assembly”. In Dimensional Optical Metrology and Inspection for Practical Applications X, Vol. 11732, SPIE, p. 1173203.

[9] Community, B. O., 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam.

[10] Li, B., Karpinsky, N., and Zhang, S., 2014. “Novel calibration method for structured-light system with an out-of-focus projector”. *Applied Optics,* ***53***(16), June, pp. 3415–3426.

[11] Zheng, Y., Zheng, Y., Wang, S., Wang, S., Li, Q., and Li, B., 2020. “Fringe projection profilometry by conducting deep learning from its digital twin”. *Optics Express,* ***28***(24), Nov., pp. 36568–36583. Publisher: Optica Publishing Group.

[12] Hyun, J.-S., and Zhang, S., 2016. “Enhanced two-

frequency phase-shifting method". *Applied Optics,* ***55***(16), June, pp. 4395–4401. Publisher: Optica Publishing Group.

[13] Marrugo, A. G., Gao, F., and Zhang, S., 2020. "State-of-the-art active optical techniques for three-dimensional surface metrology: a review". *JOSA A,* ***37***(9), pp. B60–B77.

[14] Notni, G. H., and Notni, G., 2003. "Digital fringe projection in 3D shape measurement: an error analysis". In Optical Measurement Systems for Industrial Inspection III, W. Osten, M. Kujawinska, and K. Creath, eds., Vol. 5144, International Society for Optics and Photonics, SPIE, pp. 372 – 380.

[15] Zhong, K., Li, Z., Zhou, X., Li, Y., Shi, Y., and Wang, C., 2015. "Enhanced phase measurement profilometry for industrial 3d inspection automation". *The International Journal of Advanced Manufacturing Technology,* ***76***, pp. 1563–1574.

[16] He, W., Zhong, K., Li, Z., Meng, X., Cheng, X., Liu, X., and Shi, Y., 2018. "Accurate calibration method for blade 3d shape metrology system integrated by fringe projection profilometry and conoscopic holography". *Optics and Lasers in Engineering,* ***110***, pp. 253–261.

[17] Zhang, S., 2016. *High-Speed 3D Imaging with Digital Fringe Projection Techniques*, 1st edition ed. CRC Press, Boca Raton, Mar.