Eurographics Workshop on 3D Object Retrieval (2022)
S. Berretti, T. Thehoaris, M. Daoudi, C. Ferrari, and R. C. Veltkamp (Editors)

Short Paper

Single Shot Phase Shift 3D Scanning with Convolutional Neural Network and Synthetic Fractals

K. Li[†1,2], M.Spehr[1‡], D.Höhne [1 §], C.Bräuer-Burchardt [1], A.Tünnermann [1,2], P.Kühmstedt [1 ¶]

[1] Fraunhofer Institute for Applied Optics and Precision Engineering, Germany
[2] Abbe School of Photonics, Friedrich Schiller University Jena, Germany

Abstract

The phase shift algorithm is an important 3D shape reconstruction method in industrial quality inspection and reverse engineering. To retrieve dense and accurate point clouds, the conventional phase shift methods require at least three fringe projection patterns, limiting its application to statics or semi-statics scenes only. In this paper, we propose a novel and low-cost single-shot phase shift 3D reconstruction framework using convolution neural networks (CNN) trained on 3D synthetic fractals. We first design and optimize a novel projection pattern that compresses the phase period orders and the ambiguous phase information into a single image. Then, we train two different CNNs to predict the ambiguous phase information and the period orders separately. The CNNs were trained on randomly generated 3D shapes whose geometric complexity is modeled by recursive shape generation algorithms which can create an unlimited amount of random 3D shapes on the fly. Initial results demonstrate that our method can produce high-quality point clouds from just a pair of 2D images, thus improving the temporal resolution of a phase-shift 3D scanner to the highest possible. As we also include different real-world lighting and textural conditions in the training data set, experiments also demonstrate that our CNN models which were trained on random synthetic fractals only can perform equally well in the real world.

CCS Concepts

• ***Computing methodologies*** → ***3D imaging; Neural networks;*** *Modeling methodologies;*

1. Introduction

A rapid and accurate 3D shape retrieval method is crucial in computer vision, robotics, and reverse engineering applications. Although 3D scanners such as a time of flight (TOF) camera [CSC*10, Zha12] can acquire the 3D point cloud of an object at a high framerate, the quality of these measurement results is generally not sufficient for industrial applications where sub-millimeter point cloud accuracy is required. For many tasks such as 3D surface quality control, a structured light (SL) 3D scanner is the preferred 3D shape retrieval method [IOF]. By illuminating the measurement object with coded projection patterns for stereo-matching, an SL 3D scanner is a widely adopted alternative to ToF sensors [Gen11]. However, one major limitation of the conventional SL 3D scanners is that the measurement is limited to static or semi-static scenes only. Since multiple patterns have to be sequentially projected onto the measurement object, the required processing time for a single scan result in SL 3D scanners having low temporal resolution and framerate [PMS10]. Nowadays, how to develop an effective SL 3D scanner with only a single projection pattern without significantly degrading the spatial measurement accuracy is still an active and challenging field of research.

In recent years, the success of deep convolution neural networks (CNN) [JAFF16, LIMK18] in solving inverse imaging problems has motivated researchers to investigate using machine learning approaches to reduce the number of projection patterns needed in SL 3D scanning. Many of the past research focuses on improving the conventional phase shift (PS) algorithm [FCG*19,dJD19,NWW20, BLL19], one of the most popular SL 3D scanning algorithms for accurate point cloud retrieval. The PS algorithm can achieve high spatial resolution by projecting a series of sinusoidal fringe patterns, with the increase in the number of fringe patterns resulting in higher spatial resolution and lower measurement uncertainty. In order to perform 3D measurement of a scene that includes discontinuous objects, the PS algorithm is also commonly applied with additional series of gray code patterns encoding the period order of the fringes. Such a conventional PS approach could easily result in 15-20 projection patterns for a single 3D scan. However, previous research has already demonstrated that a CNN is able to infer

[†] Now: Deutsches Elektronen-Synchrotron; ke.li1@desy.de

[‡] Now: Chair of Web Engineering, Applied Computer Science Department, University of Applied Sciences Erfurt, Germany; marcel.spehr@fh-erfurt.de

[§] Now: ZEISS Group Jena, Germany; Daniel.Hoehne@zeiss.com

[¶] peter.kuehmstedt@iof.fraunhofer.de

DOI: 10.2312/3dor.20221179

the ambiguous phase information from just a single fringe projection pattern [FCG*19, NW21]. In addition, by coding each fringe period with a unique texton pattern [BLL19], a CNN is also able to extract the period order information from one single projection robustly. In summary, using deep learning methods in improving the phase shift algorithm provides a promising opportunity in the search for a robust single shot SL 3D reconstruction framework.

In this paper, we extend the previous work in using CNN for fringe pattern and gray code pattern reduction and develop a novel single-shot 3D scanning framework that could robustly measure scenes with discontinuous objects and improve the temporal resolution of SL 3D scanners to the highest possible. In our framework, we first propose a novel fringe projection pattern that is optimized to compactly encode both the period order information and the ambiguous phase information. Then, we develop two types of CNNs to classify the period order of the fringe pattern and extract the ambiguous phase features of the fringe pattern. We demonstrate that our CNN models are able to accurately predict both the ambiguous phase information and the period order information, therefore, providing smooth and accurate global phase information for stereoscopic disparity search. In addition, we present a novel and low-cost 3D synthetic data generation approach through which we can generate an unlimited amount of 3D shapes on the fly. We use a recursive shape generation approach to control the level of complexity of the 3D shapes. In the initial experiments, we demonstrate that our single-shot CNN framework trained with only random synthetic fractal shapes can accurately reconstruct 3D point clouds in both real-world and computer-simulated environments. Our framework at the moment can already achieve 97% point cloud completeness and 89% point cloud correctness in the real-world evaluation dataset.

2. Related Work

2.1. 3D Scanning

Based on hardware components, 3D sensors can be classified as passive 3D sensors [BRR11, YWhZ*18] or active 3D sensors [Zha12, LPC*00, RCM*01]. Passive 3D sensors use binocular stereo vision and reconstruct the 3D scene with a correspondence search [BSGF10] algorithm. Although a passive 3D sensor has low hardware costs and a high frame rate, it requires expensive parameter search [YWhZ*18] and has low 3D reconstruction accuracy. Moreover, it could not be used to measure texture-less objects. Commonly known active 3D sensors are ToF sensors [LPC*00] and SL sensors [RCM*01]. The ToF sensors calculate the distance between the camera and the object by measuring the time it takes the projected infrared light to travel from the camera, bounce off the object's surface, and return to the sensor. Such a sensor could provide medium 3D reconstruction accuracy, however, at a low resolution [Zha12]. An SL 3D sensor is composed of a camera system and a projector projecting structured light onto the measurement object [Gen11, RCM*01]. SL sensor is most commonly used in industrial applications [IOF, HDL*18] as it could produce high accuracy point cloud at sub-millimeter level.

2.2. Structured-Light 3D Scanners

Based on the number of projection patterns, active stereo vision can also be classified as single-shot [LNS16, WZS20] or multi-shot [Zha16, WGZ19]. Most of the conventional single-shot SL 3D scanners project a statistical pattern to encode the spatial information of the 3D scene, requiring expensive computational parameter searches. Moreover, such coded projection patterns can not be used to measure colored or objects with texture information [LNS16]. Multi-shot SL scanners, on the other hand, are more robust in terms of computational time and measuring textural scenes. However, because of the low temporal resolution, multi-shot SL scanners are not suitable to measure non-static scenes. Extensive research has investigated improving the multi-shot SL algorithms using fewer projection patterns [HHJC99, PMS10]. However, mathematically, it is not possible to reduce the projection patterns of the conventional PS algorithms to only a single pattern, as it requires at least three sinusoidal patterns to perform phase retrieval and multiple gray code patterns for the temporal phase unwrapping (TPU) process.

2.3. CNN for Single-Shot Structured-Light 3D Scanning

With the rise of CNN in image transformation problems [JAFF16], it has been demonstrated that CNN can directly calculate the ambiguous phase information using a single sinusoidal image [FCG*19]. However, Feng et al.'s method can only predict the ambiguous phase information, therefore, can not be used to measure scenes with discontinuous objects. Budianto et al. demonstrates that instead of using a sequence of gray code patterns, the period order of the fringe pattern can be classified by a neural network [BLL19]. However, Budianto's method can not be used to predict the fringe pattern at the same time. It has also been demonstrated by Jeught et al. that a single sinusoidal pattern can also directly predict general depth information [dJD19], however, missing essential high-frequency features in the prediction. In this paper, we extend the methods of Feng et al. [FCG*19] and Budianto et al. [BLL19], and develop a contrast-enhanced projection pattern with optimized texton locations, encoding all essential information needed for PS 3D reconstruction in one image. Hence, to our best knowledge, our framework is the first PS framework that can predict both the ambiguous phase and the period order of the fringe pattern from only one projection pattern.

3. Method

3.1. The Phase-Shift Algorithm

3D reconstruction using the PS algorithm has three steps: ambiguous phase estimation, temporal phase unwrapping, and correspondence search. Figure 1 illustrates the phase-shift pipeline using 8 sinusoidal patterns and 7 gray code images. Each sinusoidal pattern is $\frac{\pi}{4}$ shifted. And the gray code images encode each of the 10 half periods of the sinusoidal pattern.

Given $N(N > 3)$ number of phase-shift steps, ambiguous phase $\phi_1(x,y)$ and $\phi_2(x,y)$, can be mathematically expressed as followed [RR93]:

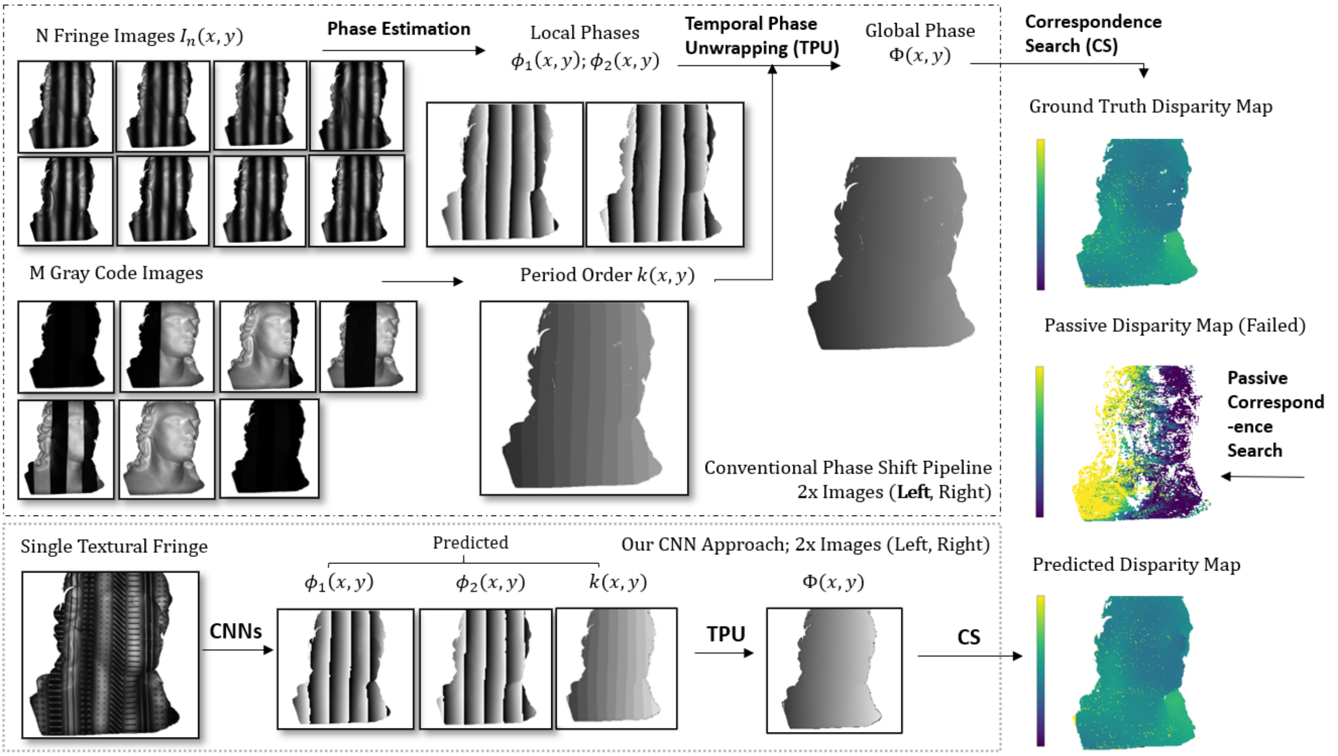
N Fringe Images $I_n(x,y)$
Phase Estimation
Local Phases $\phi_1(x,y); \phi_2(x,y)$
Temporal Phase Unwrapping (TPU)
Global Phase $\Phi(x,y)$
Correspondence Search (CS)
Ground Truth Disparity Map
M Gray Code Images
Period Order $k(x,y)$
Passive Disparity Map (Failed)
Passive Correspond -ence Search
Conventional Phase Shift Pipeline 2x Images (Left, Right)
Single Textural Fringe
Predicted
Our CNN Approach; 2x Images (Left, Right)
$\phi_1(x,y)$
$\phi_2(x,y)$
$k(x,y)$
$\Phi(x,y)$
CNNs
TPU
CS
Predicted Disparity Map

Figure 1: *Top framework: the* ***Conventional Phase-shift Approach****. The conventional phase-shift approach algorithm requires multiple projection sequences for one 3D scan, resulting in low temporal resolution. Bottom framework:* ***Our CNN Single-Shot Approach****. The approach reduces all the required projection sequences to one, making accurate predictions of the intermediates images in the conventional PS pipelines.*

$$\phi_1(x,y) = \arctan \frac{\sum_{n=1}^{N} I_n(x,y)\sin(\frac{2\pi n}{N})}{\sum_{n=1}^{N} I_n(x,y)\cos(\frac{2\pi n}{N})} \tag{1}$$

$$\phi_2(x,y) = -\arctan \frac{\sum_{n=1}^{N} I_n(x,y)\sin(\frac{2\pi n}{N})}{\sum_{n=1}^{N} I_n(x,y)\cos(\frac{2\pi n}{N})} \tag{2}$$

where $I_n(x,y)$ are the intensity of the projected PS images. We refer to the ambiguous phase $\phi_1(x,y)$ and $\phi_2(x,y)$ also as **local phase maps**.

Using the gray code images, we could calculate the period order information $k(x,y)$, from which in the combination with the local phase maps, could perform TPU and yield the **global phase map** $\Phi(x,y)$:

$$\Phi = \begin{cases} \phi_1 + [k - mod(k,2)]\pi, & mod(k,2) = 0 \\ \phi_2 + [k + mod(k,2) - 1]\pi, & mod(k,2) = 1 \end{cases} \tag{3}$$

Notice that in Equation 3, we use both of the local phase maps to avoid the phase jump artifacts [WL12]. For detailed mathematics derivation of the algorithm, we recommend the literature *Interferogram Analysis, Digital Fringe Pattern Measurement Techniques* [RR93].

Lastly, after image rectification, [ZDFL95], we can find the correspondence points between the left and right camera based on each epipolar line from the global phase maps $\Phi(x,y)$. Depth information z can be retrieved following Equation 4, where L is the camera baseline length, f is the camera focal length, u_1 is the pixel location of the global phase map for the left camera, u_2 is the pixel location of the global phase map for the right camera, and D is the disparity.

$$z = \frac{Lf}{u2 - u1} = \frac{Lf}{D}. \tag{4}$$

3.2. Framework Overview

Figure 2 presents an overview of our single-shot SL 3D scanning framework. To capture a real-world evaluation dataset, we use a forensic SL 3D scanner that is composed of two Basler ace acA2040-90um cameras with a resolution of 2048 × 1024 pixels and an Optoma ML750e Beamer projector unit with a resolution of 1280 × 800 pixels [IOF]. Although the PS algorithm is also applicable to monocular setup, our framework is based on a stereo setup as a stereo setup can expand the viewing angle of the scene, mitigating occlusion when scanning high reflectance surfaces. In addition,

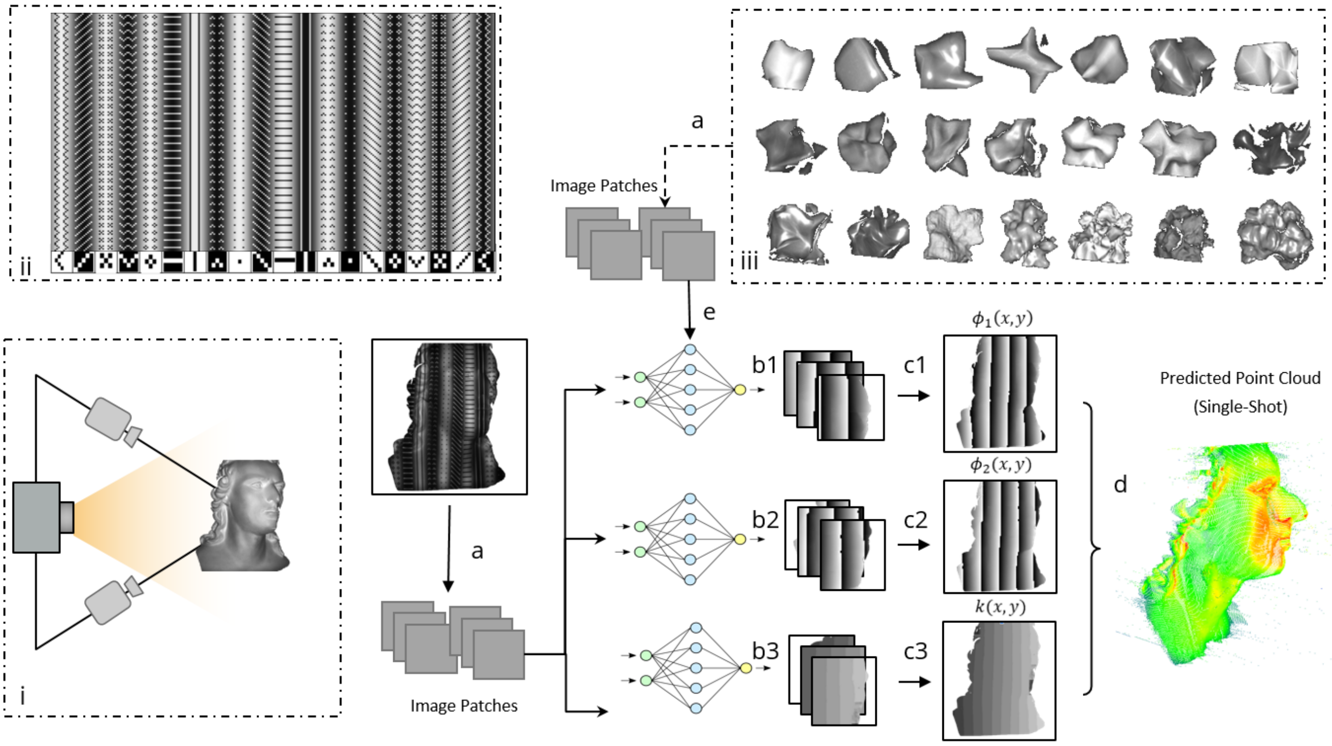

Figure 2: *Overview of Our Single Shot PS 3D Scanning Framework. (i) illustrate the stereo vision setup, (ii) show our compressed PS single projection pattern, and (iii) display examples of our training data: random 3D fractals with various geometrical complexities and surface properties. Process (a) divides full-scale images into* 64×64 *pixel patches. Process (b1, b2, b3) predicts corresponding learned labels. Process (c1, c2, c3) combines image patches to form full-scale images. Process (d) reconstructs 3D point clouds from stereo matching.*

as a stereo setup is invariant to the position and rotation of the projector, calibration of the projector is not required. Nonetheless, our framework could also be easily adapted to an SL system with one or multiple cameras.

After performing extrinsic and intrinsic calibration for the stereo system, we export the camera matrix and the estimated 3D poses of the cameras to Blender, a popular software for 3D simulation and modeling, to create a digital twin of the forensic 3D scanning system [IOF]. 3D fractals are randomly generated as measurement targets to create training data. The rendered images are divided into 64×64 pixel patches of training features and labels such that the CNN can extract features with accurate spatial details. Three separate neural networks are used to predict the three different intermediate images in the PS 3D reconstruction pipeline as shown in Figure 1. Finally, the predicted local phases and period order image patches are stitched back together. Correspondence search is performed to calculate the disparity between the stereo global phase maps. Final point clouds can be reconstructed based on the camera focal length and the disparity map.

3.3. The Single-Shot Projection Pattern

Figure 2 (ii) illustrates our single shot projection pattern. The projection pattern contains 10 sinusoidal periods. The half period order, indicated by the index in sub-figure (ii) of Figure 2 is each encoded by a unique pattern. We adapted and optimized the patterns first proposed by Budianto et al. [BLL19]. The ordering of the texton pattern is optimized through repetitive experiments such that the statistical similarity of the pattern is minimized for each neighboring pattern to avoid classification confusion. As Figure 1 illustrates, one CNN is trained to classify the period order $k(x,y)$ based on the unique texton pattern for each half period, and two other CNNs are trained to estimate the local phase $\phi_1(x,y)$, and $\phi_2(x,y)$.

3.4. Training Data Generation

Fractals are known as geometries or signals that resemble self-similarities. In computer graphics, statistical fractals are widely used to model complex natural geometries such as coastlines, clouds, and mountains [The90]. With the capability to model complex geometrical properties recursively based on an initial condition and simple rules, computer-generated fractals are becoming popular among computer vision researchers for automatic training data generation [KOM*20]. To generate random statistical fractals in 3D, we use the subdivide function in Blender [Com18]. The subdivide function connects meshes of 3D objects to random value generators, making it possible to randomly change the surface typology of an object recursively. By adjusting the strength of the

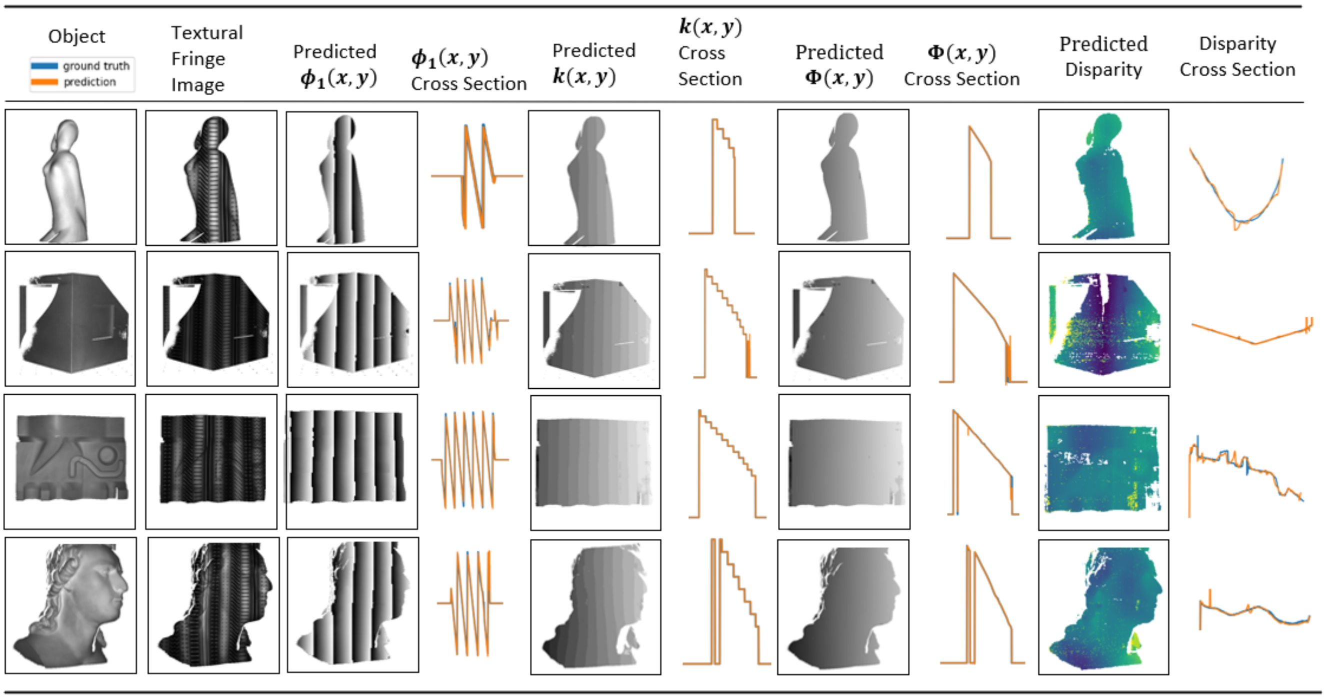

Figure 3: ***Real-World Raw Prediction Results of Intermediate Images in phase-shift Reconstruction Pipeline*** *The CNNs predict $\phi_1(x,y)$, $\phi_2(x,y)$, and $k(x,y)$ for both left and right cameras. From these intermediate results, we obtain the global phase map $\Phi(x,y)$ for correspondence search. Evaluation target from top to bottom: a female sculpture, an occluded object [TMD09], an industrial part [HKTN17], and a sculpture head.*

random movement of subdivided edges, the number of cuts (the recursion step), and the smoothness of the mesh surface to add random surface curvature, we can generate an infinite amount of unique 3D objects with a wide range of geometrical complexity.

We simulate real-world material properties by randomly attaching different functions that model the real-world optical properties of non-reflective dielectric surfaces to the random fractal geometries. The real-world materials are modeled with the physics-based bidirectional scattering distribution function (BSDF) shaders with different control parameters such as the index of refraction, specularity, sub-surface scattering, and transmission. In addition, common real-world textures such as dirt, scratch, and wall paints are randomly attached to the surfaces to improve the network's robustness when performing 3D scans in the real world.

For proof-of-concept experiments, we generate training data at 400 × 250 pixel, which allows rapid data generation of over 1000 simulated 3D scans of unique random fractals in one day on a computer with 64 GB RAM. The actual training data is 64 × 64 pixel patches that were randomly extracted from each full 400 × 250 pixels image. Figure 2 illustrates some examples of the image patches, showing that the 64 × 64 pixel patches window size is optimal to cover around two periods of the phase feature. This allows the CNNs to perform accurate feature extraction and classification. We generated over 5000 simulated 3D PS scans of random fractals with a resolution of 400 × 250 and selected 24 patches from each scan randomly. Our final training dataset consists of 120,000 64 × 64pixels image patches. Generation of such dataset takes approximately 5 days on a 64 GB RAM CPU.

3.5. Neural Network and Training Details

We used the CNN architecture of the perceptual style transfer network proposed by [JAFF16]. The implementation of our CNN is adopted from an open source github repository [Lee19]. We adjusted the input image size and output image size to 64 × 64 pixels to accommodate the actual training patch size. The CNN consists of one reflection padding layer, three convolution layers, five residual convolution layers [HZRS15], and three de-convolution layers. We used ReLU (Rectified Linear Unit) as an activation function for the convolution layer, and tanh for the de-convolution layer. All CNN models were trained using the Adam optimizer with the default learning rate of 0.001.

The CNN for predicting the period order $k(x,y)$ uses a pixel-wise mean squared error loss function. The CNNs for predicting the local phases $\phi_1(x,y)$ and $\phi_2(x,y)$ use a cyclic loss function to calculate the mean squared error of the difference of the cosine and the sine values of the predicted images $\hat{y}$ and ground truth image y, as shown in Equation 5. In addition, a binary segmentation map is created with OpenCV GrabCut [Bra00] to help the CNN distinguish the measurement target from the background. The multiplication

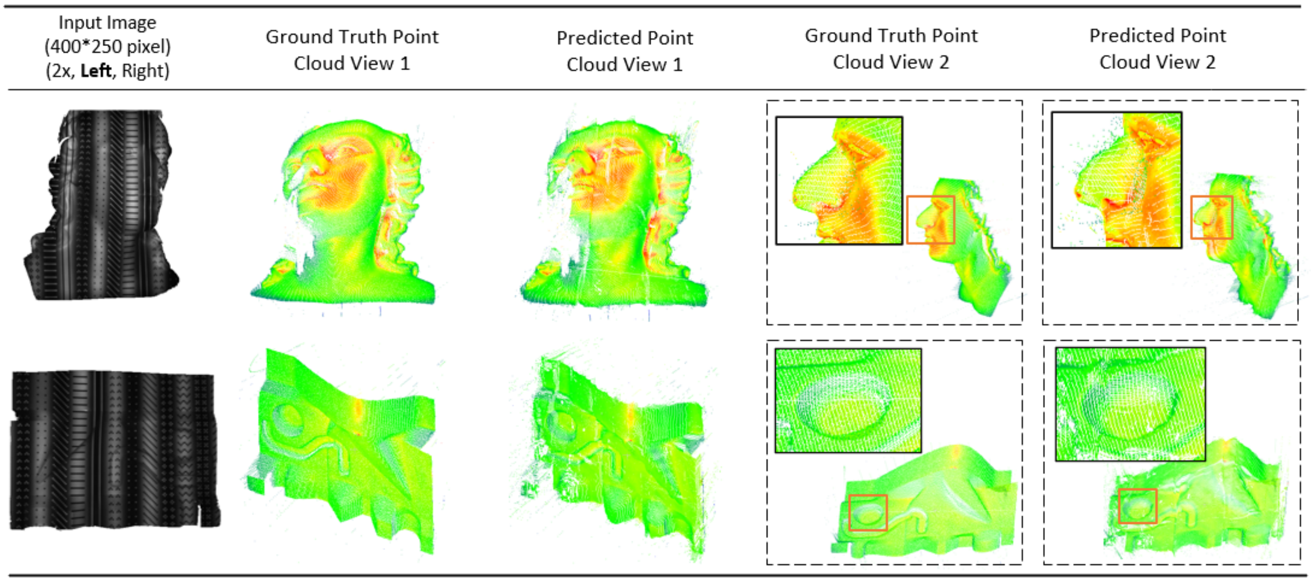
Input Image
(400*250 pixel)
(2x, Left, Right)
Ground Truth Point Cloud View 1
Predicted Point Cloud View 1
Ground Truth Point Cloud View 2
Predicted Point Cloud View 2

Figure 4: ***Real-World Raw Reconstructed Point Clouds Comparison*** *Reconstructed point clouds and ground truth point clouds of a Schiller bust (top) and an industrial part [HKTN17] from different viewing angles from the same 3D scan are presented.*

of the error with the binary segmentation mask (y_b) helps the loss function to focus on regions within the measurement target.

$$\mathcal{L}_c = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\cos\hat{y}_i - \cos y_i\right)^2 + \left(\sin\hat{y}_i - \sin y_i\right)^2\right] \times y_{b_i} \tag{5}$$

We trained three models to predict $\phi_1(x,y)$, $\phi_2(x,y)$, and $k(x,y)$ separately on an NIVIDIA RTX 6000 graphic cards. The CNNs are trained on synthetic fractals only. Each model was trained for 3000 epochs, which takes 4 days for each training.

3.6. Evaluation Dataset

We evaluate our models both qualitatively and quantitatively in real-world 3D scanning settings. Our evaluation dataset includes measurement targets that the CNNs have never seen before, such as sculptures, white-painted industrial parts, and objects for view planning that are hard to measure due to occlusion and sharp angle [TMD09]. We also compare the performance of our models when predicting synthetic data and real-world data.

The ground truth real-world evaluation dataset was captured using an industrial SL 3D scanner [IOF] with measurement uncertainty between $20-100$ μm, 2048×1280 pixels, and a measurement field of 325×200 mm^2. This is the same 3D scanner with the same configuration and calibration matrix which we used to generate the simulated 3D scanning environment in Blender. The ground truth PS scans use 8 project patterns of 10 periods of sinusoidal fringes and 7 gray code projection patterns, as illustrated in Figure 1. This ensures that the ground truth PS methods also have 10 period of fringes as the proposed single-shot projection pattern. To have a comparable point cloud quality, the real-world images were scaled down from 2048×1280 to 400×250 using OpenCV [Bra00].

Object	Female Bust (over-exposed)	Industrial Part [TMD09]	Schiller Bust
$C\uparrow$	101.66%	98.13%	99.20%
$R\uparrow$	–	91.4%	90.0%

Table 1: ***Quantitative measurement of the completeness and the correctness of the evaluation objects shown in Figure 3.***

3.7. Evaluation Metrics

We present two evaluation metrics, point cloud **completeness (C)** and point cloud **correctness (R)** to perform quantitative measurement of the quality of the single-shot CNN predicted point cloud against its' corresponding ground truth point cloud calculated with the conventional PS method.

The **completeness** $C = \frac{N_{cnn}}{N_{gt}} \times 100\%$ is calculated by the ratio of the number of valid points in the predicted disparity map (N_{cnn}) and the number of valid points in the ground truth disparity map (N_{gt}).

The **correctness** $R = \frac{R_{cnn}}{R_{gt}} \times 100\%$ is calculated by the ratio of the number of correct points (difference in disparity map is less than 1) in the prediction disparity map (R_{cnn}) to the number of valid points in the ground truth disparity map (R_{gt}).

Method	$C_b \uparrow$	$R_b \uparrow$	$C_{rw} \uparrow$	$R_{rw} \uparrow$
passive stereo	32%	3.45%	40.5%	3.2%
Ours	97.5%	90.7%	97.5%	89.5%

Table 2: ***Quantitative Results of Rendered and Real-World Dataset.*** *C_b measures the point cloud completeness of our rendered dataset. R_b measures the point cloud correctness using our rendered dataset. C_{rw} measures the point cloud completeness of the dataset with real-world 3D scans. R_{rw} measures the point cloud correctness using a real-world dataset.*

4. Results

Figure 3 shows the raw results of $\phi_1(x,y)$, $\phi_2(x,y)$, and $k(x,y)$ calculated with the conventional PS reconstruction pipelines and the raw results predicted by the CNNs. The cross-section lines randomly selected from a row of the predicted image show that only minor prediction errors occurred at phase jump areas for the local phase maps and boundary areas where the textural patterns are occluded. The errors at the phase jump areas are avoided based on the adaptive global phase calculation algorithm described in Equation 3. Figure 4 presents the final reconstructed point clouds. Table 1 shows the point cloud correctness (R) and completeness (C) of the measured objects. To demonstrate that our framework can be a good complement to the conventional passive stereo sensing technique in acquiring 3D point clouds from a single shot, Table 2 shows the average performance of our approach and the passive stereo vision [Bra00] approach on the rendered and the real-world dataset.

5. Discussions

5.1. Advantages of our Framework

Measurement results of point cloud completeness and correctness demonstrate that our framework can accurately predict the local phase maps and the period order of the fringe pattern at the same time, resulting in a high-quality global phase map for 3D point cloud reconstruction. Moreover, unlike previous CNN frameworks for phase-shift 3D reconstruction, our CNNs are trained on random synthetic geometric shapes only, which significantly reduces the costs and efforts to select training data from datasets such as the Thingi10K [ZJ16].

We also discover that our CNNs are more robust in undesired lighting conditions, for example, when the scene is over-exposed. Over-exposure is a common engineering problem in cheap projector modules with a low projection framerate. In order to create a linear response when projecting a sinusoidal fringe pattern with continuously oscillating pixel intensity, the exposure time for the fringe projection pattern has to be increased, making a glossy surface more likely to have over-exposure artifacts. As such marginal over-exposed scenes are also included in the synthetic training data, the neural network can to some extent learn to extract the phase features from ambiguous over-exposed scenes as well, making it an additional advantage of our CNN framework over the conventional PS methods.

The experiment results also demonstrate that our CNN models can perform equally well in our real-world evaluation dataset and the synthetic dataset, both of which the networks have never seen before. This is unsurprising, as the training dataset simulates various material and simple textural properties of real-world objects. In addition, as a background-foreground segmentation mask is used in the loss function, the CNN models are invariant to the real-world background environment, making them more robust when it comes to transferring to real-world applications.

5.2. Limitations

Our framework in the current state has several limitations. As Figure 3 illustrates, the main source of error comes from the error of period order detection at the boundary of objects. Such error could be reduced by increasing image resolution and the resolution of textural patterns, or applying a smoothing filter on the predicted period order. However, for a stereoscopic 3D scanning framework, the final reconstruction accumulates the errors from both cameras. Future work can investigate and compare the reconstruction accuracy with a monocular SL approach using our CNN method.

In addition, our framework is not yet robust enough to perform reconstruction with heavily texture objects, as the texture of the objects could interfere with the texton patterns we use to predict the period order of the fringe pattern. However, as table 2 shows, our method is a good complement to the passive stereo algorithm, where the passive stereo is known for being able to perform effective correspondence searches on heavily texture objects rather than textureless objects.

Moreover, existing experiments only demonstrate measurement results at low image resolution (400×250 pixels). Future work can extend our framework to higher image and reconstruction resolution for accurate dense point cloud retrieval.

6. Conclusion

In this paper, we propose a novel single-shot phase-shift 3D scanning framework with a low-effort synthetic training data generation pipeline using random synthetic fractals. We demonstrate that CNNs trained on synthetic fractals only are able to perform accurate predictions of both the local phase maps and the period order of the fringe pattern in the real-world measurement setting. Compared to the conventional phase-shift algorithm with 10 periods of fringe and 15 projection patterns, our method achieves 89.5% point cloud correctness and 97.5% point cloud completeness at low image resolution. Further research will improve our pipeline by performing experiments at higher resolution and extending our methods to measure objects with heavy textures, for example, using the method proposed by Vo et al [VNS16].

ACKNOWLEDGMENTS

We acknowledge the support of the Honours for Future Researchers at the Friedrich Schiller University of Jena, funded by the German Research Foundation (DFG).

References

[BLL19] Budianto, Law W., Lun D. P. K.: Deep learning based period order detection in structured light three-dimensional scanning. In *2019 IEEE International Symposium on Circuits and Systems (ISCAS)* (2019), pp. 1–5. doi:10.1109/ISCAS.2019.8701883. 1, 2, 4

[Bra00] Bradski G.: The OpenCV Library. *Dr. Dobb's Journal of Software Tools* (2000). 5, 6, 7

[BRR11] Bleyer M., Rhemann C., Rother C.: Patchmatch stereo - stereo matching with slanted support windows. In *BMVC* (2011). 2

[BSGF10] Barnes C., Shechtman E., Goldman D. B., Finkelstein A.: The generalized patchmatch correspondence algorithm. In *ECCV* (2010). 2

[Com18] Community B. O.: *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. URL: http://www.blender.org. 4

[CSC*10] Cui Y., Schuon S., Chan D., Thrun S., Theobalt C.: 3d shape scanning with a time-of-flight camera. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2010), 1173–1180. 1

[dJD19] der Jeught S. V., Dirckx J.: Deep neural networks for single shot structured light profilometry. *Optics express 27 12* (2019), 17091–17101. 1, 2

[FCG*19] Feng S., Chen Q., Gu G., Tao T., Zhang L., Hu Y., Yin W., Zuo C.: Fringe pattern analysis using deep learning. *SPIE Advanced Photonics, 1(2)*, 025001 (2019). doi:https://doi.org/10.1117/1.AP.1.2.025001. 1, 2

[Gen11] Geng J.: Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics 3* (2011), 128–160. 1, 2

[HDL*18] Heist S., Dietrich P., Landmann M., Kühmstedt P., Notni G., Tünnermann A.: Gobo projection for 3d measurements at highest frame rates: a performance analysis. *Light, Science & Applications 7* (2018). 2

[HHJC99] Huang P., Hu Q., Jin F., Chiang F.: Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering 38* (1999), 1065–1071. 2

[HKTN17] Heist S., Kühmstedt P., Tünnermann A., Notni G.: Brdf-dependent accuracy of array-projection-based 3d sensors. *Applied optics 56 8* (2017), 2162–2170. 5, 6

[HZRS15] He K., Zhang X., Ren S., Sun J.: Deep residual learning for image recognition. *CoRR abs/1512.03385* (2015). URL: http://arxiv.org/abs/1512.03385. 5

[IOF] IOF F.: Kolibri cordless handheld optical 3d scanner. https://www.iof.fraunhofer.de/content/dam/iof/en/documents/pb/kolibri-cordless-forensics-scanner-e.pdf. Online; accessed 31 December 2021. 1, 2, 3, 4, 6

[JAFF16] Johnson J., Alahi A., Fei-Fei L.: Perceptual losses for real-time style transfer and super-resolution. *ArXiv abs/1603.08155* (2016). 1, 2, 5

[KOM*20] Kataoka H., Okayasu K., Matsumoto A., Yamagata E., Yamada R., Inoue N., Nakamura A., Satoh Y.: Pre-training without natural images. In *ACCV* (2020). 4

[Lee19] Lee S.: fast-neural-style-keras. https://github.com/misgod/fast-neural-style-keras, 2019. 5

[LIMK18] Lucas A., Iliadis M., Molina R., Katsaggelos A.: Using deep neural networks for inverse problems in imaging: Beyond analytical methods. *IEEE Signal Processing Magazine 35* (2018), 20–36. 1

[LNS16] Lin H., Nie L., Song Z.: A single-shot structured light means by encoding both color and geometrical features. *Pattern Recognit. 54* (2016), 178–189. 2

[LPC*00] Levoy M., Pulli K., Curless B., Rusinkiewicz S. M., Koller D., Pereira L., Ginzton M., Anderson S. E., Davis J., Ginsberg J., Shade J., Fulk D.: The digital michelangelo project: 3d scanning of large statues. *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (2000). 2

[NW21] Nguyen H., Wang Z.: Accurate 3d shape reconstruction from single structured-light image via fringe-to-fringe network. *Photonics 8*, 11 (2021). URL: https://www.mdpi.com/2304-6732/8/11/459, doi:10.3390/photonics8110459. 2

[NWW20] Nguyen H., Wang Y., Wang Z.: Single-shot 3d shape reconstruction using structured light and deep convolutional neural networks. *Sensors (Basel, Switzerland) 20* (2020). 1

[PMS10] Pribanić T., Mrvos S., Salvi J.: Efficient multiple phase shift patterns for dense 3d acquisition in structured light scanning. *Image Vis. Comput. 28* (2010), 1255–1266. 1, 2

[RCM*01] Rocchini C., Cignoni P., Montani C., Pingi P., Scopigno R.: A low cost 3d scanner based on structured light. *Computer Graphics Forum 20* (2001). 2

[RR93] Robinson D. W., Reid G. T.: *Interferogram Analysis, Digital Fringe Pattern Measurement Techniques*. Institute of Physics Publishing, Bristol and Philadelphia, 1993. 2, 3

[The90] Theiler J.: Estimating fractal dimension. *Journal of The Optical Society of America A-optics Image Science and Vision 7* (1990), 1055–1073. 4

[TMD09] Trummer M., Munkelt C., Denzler J.: Combined gklt feature tracking and reconstruction for next best view planning. In *DAGM-Symposium* (2009). 5, 6

[VNS16] Vo M., Narasimhan S., Sheikh Y.: Texture illumination separation for single-shot structured light reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence 38* (2016), 390–404. 7

[WGZ19] Wu Z., Guo W., Zhang Q.: High-speed three-dimensional shape measurement based on shifting gray-code light. *Optics express 27 16* (2019), 22631–22644. 2

[WL12] Weng J.-F., Lo Y.-L.: Novel rotation algorithm for phase unwrapping applications. *Optics Express Vol. 20, Issue 15* (2012). doi:https://doi.org/10.1364/OE.20.016838. 3

[WZS20] Wang Z., Zhou Q., Shuang Y.: Three-dimensional reconstruction with single-shot structured light dot pattern and analytic solutions. *Measurement 151* (2020), 107114. 2

[YWhZ*18] Yang L., Wang B., hui Zhang R., Zhou H., Wang R.: Analysis on location accuracy for the binocular stereo vision system. *IEEE Photonics Journal 10* (2018), 1–16. 2

[ZDFL95] Zhang Z., Deriche R., Faugeras O., Luong Q.: A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artif. Intell. 78* (1995), 87–119. 3

[Zha12] Zhang Z.: Microsoft kinect sensor and its effect. *IEEE Multim. 19* (2012), 4–10. 1, 2

[Zha16] Zhang S.: *High-Speed 3D Imaging with Digital Fringe Projection Techniques*. CRC Press LLC, Indiana, USA, 2016. 2

[ZJ16] Zhou Q., Jacobson A.: Thingi10k: A dataset of 10,000 3d-printing models. *arXiv preprint arXiv:1605.04797* (2016). 7