

Supplementary Information for Digital-Twin-Driven Unambiguous Structured Light 3D Imaging with Physics-Aware Learning

Yiheng Liu,^{†,‡} Wenwu Chen,^{†,‡} Jinyang Jiang,^{†,‡} Shengqi Yu,^{†,‡} Ziheng Jin,^{†,‡}
Xinsheng Li,^{†,‡} Shijie Feng,^{*,†,‡} Qian Chen,^{†,‡} and Chao Zuo^{†,‡}

[†]*Smart Computational Imaging Laboratory (SCILab), Nanjing University of Science and
Technology, Nanjing, Jiangsu Province 210094, China*

[‡]*Jiangsu Key Laboratory of Visual Sensing & Intelligent Perception, Nanjing, Jiangsu
Province 210094, China*

E-mail: shijiefeng@njust.edu.cn

Abstract

This document provides supplementary information for the paper titled "Digital-Twin-Driven Unambiguous Structured Light 3D Imaging with Physics-Aware Learning." We detail the specific parameter configurations of the digital twin system, its calibration process, and ablation studies on the integration of digital twin technology, physical priors, and the Fourier loss function.

Contents

1. Construction of the Digital Twin System
2. Calibration of the Digital Twin System
3. Fringe-Order Map Computation for Synthetic Data
4. Comparative Study on Network Performance Under Different Training Strategies
5. Ablation Study on Composite Loss Design

1. Construction of the Digital Twin System

Based on the derivations in the main text and the literature,¹ we established the parameter mapping relationship between the real system and the digital twin system. The intrinsic and extrinsic matrices of the virtual camera and projector were provided in Table S1, including focal lengths, principal points, skew factors, rotation angles, and translation vectors.

Table S1: Mapping between real-system calibration matrices and digital twin system parameters

	Virtual camera	Virtual projector
Intrinsic matrix		
Focus	$f_u^c k, f_v^c l$	$f_u^p k, f_v^p l$
Principle Points	(u_0^c, v_0^c)	(u_0^p, v_0^p)
Skew factor	λ_c	λ_p
Extrinsic matrix		
Location	$-R_c^T T_c$	$-R_p^T T_p$
Rotated angles	$\psi^c = \arctan\left(\frac{r_{32}^c}{r_{33}^c}\right)$	$\psi^p = \arctan\left(\frac{r_{32}^p}{r_{33}^p}\right)$
	$\theta^c = \arctan\left(\frac{-r_{31}^c}{\sqrt{(r_{32}^c)^2 + (r_{33}^c)^2}}\right)$	$\theta^p = \arctan\left(\frac{-r_{31}^p}{\sqrt{(r_{32}^p)^2 + (r_{33}^p)^2}}\right)$
	$\phi^c = \arctan\left(\frac{r_{21}^c}{r_{11}^c}\right)$	$\phi^p = \arctan\left(\frac{r_{21}^p}{r_{11}^p}\right)$

In this study, the system parameters of the real camera and projector are shown in Table S2.

Table S2: Intrinsic and extrinsic parameters of the real camera and projector

Parameter	Real camera	Real projector
Principal point	[309.047, 236.166]	[443.108, 588.109]
Focal length	[2592.156, 2592.696]	[1732.166, 3463.328]
Skew factor	-0.2332	-0.9514
Rotation matrix	$\begin{bmatrix} 0.990 & -0.006 & -0.135 \\ -0.016 & -0.996 & -0.0774 \\ 0.135 & 0.079 & -0.987 \end{bmatrix}$	$\begin{bmatrix} 0.999 & -0.031 & 0.019 \\ -0.030 & -0.996 & -0.076 \\ 0.021 & 0.075 & -0.996 \end{bmatrix}$
Translation vector	[4.866, -26.432, 847.611]	[2.963, -25.620, 801.679]

Therefore, according to the mapping relationship in table S1, we set the digital twin system parameters as the values in the Table S3.

Table S3: Rotation and location parameters of the virtual camera and projector

Parameter	Virtual camera	Virtual projector
Rotated angles	$[4.4822^\circ, 7.8134^\circ, -0.3567^\circ]$	$[4.3868^\circ, -1.0881^\circ, -1.8107^\circ]$
Location(mm)	$[10.1985; -93.2745; 835.7891]$	$[-20.8421; -86.2501; 797.1710]$

The fundamental intrinsic parameters of both cameras and projectors also include focal length. In this study, key dimensional specifications were obtained by referring to the official technical documentation of the devices used. According to the datasheet, the pixel size p_c of the camera (acA640-750 μm) is $[4.8 \mu m, 4.8 \mu m]$. Based on this information, the camera's focal length can be calculated accordingly

$$\begin{cases} f_u^c = K_{11}^c \times p_{11}^c = 12.4423mm, \\ f_v^c = K_{22}^c \times p_{12}^c = 12.4449mm. \end{cases} \quad (S1)$$

Due to differences between the arrangement of the DMD's micro-mirror array and the pixel arrangement of the camera, the focal length of the projector is estimated using the physical dimensions of its sensor. According to the technical manual of the DLP4500, the projector sensor size is 10.8 mm×6.8 mm. Given the projector's resolution of 912×1024, the equivalent pixel sizes in the horizontal and vertical directions can be derived by dividing the sensor dimensions by the corresponding resolution

$$p_p = [\frac{10.8mm}{912}, \frac{6.8mm}{1140}] = [11.84\mu m, 6.64\mu m]. \quad (S2)$$

Thus, the projector's focal length under the inverse camera model can be derived as

$$\begin{cases} f_u^p = K_{11}^p \times p_{11}^p = 20.5089mm, \\ f_v^p = K_{22}^p \times p_{12}^p = 22.9965mm. \end{cases} \quad (S3)$$

2. Calibration of the Digital Twin System

To visually verify the equivalence between our constructed digital twin system and the real system, we performed calibration on the digital twin system. Specifically, a virtual circular calibration board with a 9×11 grid layout was created in Blender. The spacing between the circular markers was set to 20 mm, with large markers having a diameter of 5 mm and small markers 2 mm. As shown in Fig. S1, to ensure calibration consistency under coordinate transformations, the center of the virtual calibration board was placed at the origin (0,0,0) of the digital twin system's world coordinate system. Subsequently, Zhang's calibration method² was adopted, aligning with the calibration workflow of the real system. During the virtual calibration process, horizontal and vertical phase-shifted sinusoidal fringe patterns with phase shift steps $N=\{12,12,12\}$ and frequencies $f=\{1,8,48\}$ were projected onto the virtual calibration board. The board was tilted at angles $(0 \pm 10^\circ, 0 \pm 20^\circ, 0)$, and nine sets of images under different poses were captured to compute the calibration parameters for the camera and projector in the digital twin system.

The intrinsic and extrinsic matrices of the projector and camera were calibrated using the MATLAB Calibration Toolbox³ and further optimized through bundle adjustment.⁴ This optimization accounted for measurement errors in the calibration board and potential manufacturing discrepancies. As a result, we obtained the calibrated parameters for the digital twin system, which are shown in Table S4.

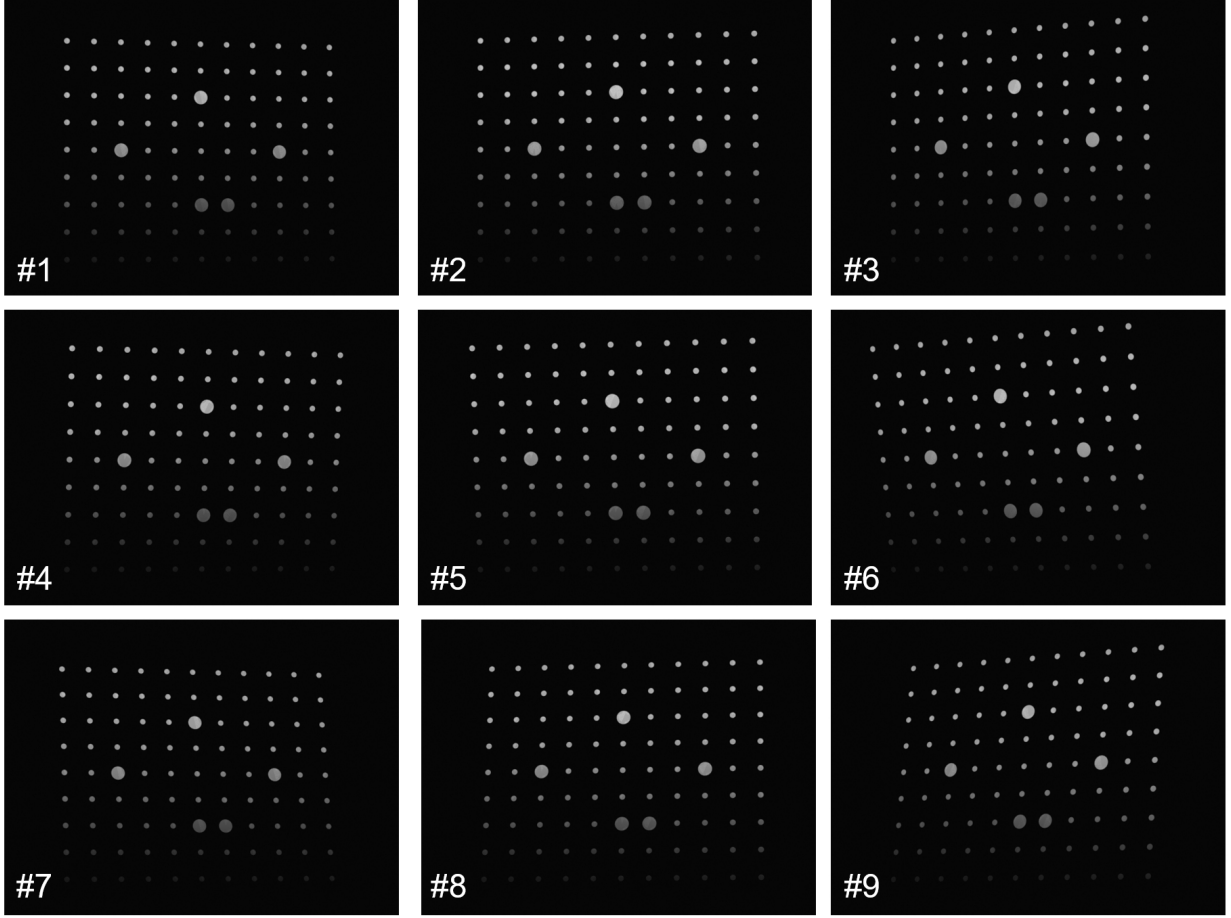


Figure S1. 9 images of the calibration board with different poses in the calibration of the digital twin system.

Table S4: Calibration parameters of camera and projector in digital twin system

Parameter	Digital twin camera	Digital twin projector
Principal point	[310.331, 241.586]	[437.855, 559.994]
Focal length	[2601.299, 2601.946]	[1730.956, 3435.8356]
Skew factor	-0.2568	0.0144
Rotation matrix	$\begin{bmatrix} 0.991 & -0.006 & -0.132 \\ -0.017 & -0.996 & -0.079 \\ 0.132 & 0.081 & -0.988 \end{bmatrix}$	$\begin{bmatrix} 0.999 & -0.031 & 0.017 \\ -0.030 & -0.997 & -0.071 \\ 0.019 & 0.071 & -0.997 \end{bmatrix}$
Translation vector	[4.281; -25.448; 845.854]	[2.405; -27.955; 800.933]

3. Fringe-Order Map Computation for Synthetic Data

In our work, the fringe order for synthetic data is computed using exactly the same algorithm as employed in traditional temporal phase unwrapping (TPU) methods. The specific process is as follows:

Step 1: image rendering. After calibrating the camera-projector pair, we instantiate the digital twin in Blender and import the selected Thingi10K model. The projector sequentially casts twelve phase-shifted sinusoidal patterns onto the object, and the virtual camera captures twelve phase-shifted images. The phase shifted sinusoidal fringe pattern can be represented as

$$I_n(x, y) = A^p(x, y) + B^p(x, y) \cos \left(2\pi f x - \frac{2\pi n}{N} \right), \quad (4)$$

where $A^p(x, y)$ is DC component, $B^p(x, y)$ is amplitude, and f is the frequency of the projected fringe pattern. Each intensity image follows the standard phase-shifting image formation

$$I_n^c(x, y) = A^c(x, y) + B^c(x, y) \cos \left[\psi(x, y) - \frac{2\pi n}{N} \right], \quad (5)$$

Step 2: Wrapped phase recovery. For each 12-frame stack, we compute the wrapped phase $\psi(x, y)$ by the least-squares formula

$$\psi(x, y) = \arctan \frac{\sum_{n=1}^N I_n^c(x, y) \sin \left(\frac{2\pi n}{N} \right)}{\sum_{n=1}^N I_n^c(x, y) \cos \left(\frac{2\pi n}{N} \right)}. \quad (6)$$

Step 3: Compute the fringe-order map. We adopt the same formulas as the traditional temporal methods to obtain $k(x, y)$ from the pairs of wrapped phases. For multi-frequency (MF) method, it can be calculated as

$$k_h^{MF}(x, y) = \text{Round} \left[\frac{f_h \psi_l(x, y) - f_l \psi_h(x, y)}{2\pi f_l} \right]. \quad (7)$$

where f_h and f_l represent the high- and low-frequency gratings respectively. For multi-wavelength (MW) method, it can be calculated as

$$k_h^{MW}(x, y) = \text{Round} \left[\frac{f_h \psi_{eq}(x, y) - f_{eq} \psi_h(x, y)}{2\pi f_{eq}} \right], \quad (8)$$

where ψ_{eq} and f_{eq} represent the equivalent wrapped phase and equivalent grating frequency. For number-theoretic (NT) method, it can be calculated as

$$(k_h^{NT}, k_l^{NT}) = \text{LUT} \left[\text{Round} \left(\frac{\lambda_l \psi_l - \lambda_h \psi_h}{2\pi} \right) \right], \quad (9)$$

where LUT represents the lookup table.

4. Comparative Study on Network Performance Under Different Training Strategies

To further validate the contribution of the digital twin technology and physical priors employed in our proposed method to network performance enhancement, we constructed four experimental groups for comparison: Traditional TPU algorithm (Traditional), Digital Twin-driven network (DT-TPU), Real-data-driven conventional U-Net models without physical priors (UNet-TPU), and Digital-Twin-Driven Physics-Aware network (DP-TPU). Networks were tested under both seen frequency $f_h = 48$ and unseen frequency $f_h = 96$, with phase unwrapping errors evaluated across three modalities (MF, MW, NT). Table shows that the DT-TPU network, leveraging the high-fidelity, noise-free virtual label data generated by the digital twin system, achieves higher unwrapping accuracy compared to the traditional TPU algorithm under the seen frequency $f_h = 48$. Across all three modalities (MF, MW, NT), the unwrapping accuracy exceeds 97%. However, a significant performance gap persists between DT-TPU and the UNet-TPU network trained on real-world data, indicating that despite the utility of digital twin data in improving training precision, non-negligible discrepancies

Table S5: Comparison of phase unwrapping errors under different training strategies for seen and unseen frequencies.

frequency	Method	MF	MW	NT
F=48 seen	Traditional	3.39%	4.12%	3.73%
	DT-TPU	2.37%	2.63%	2.46%
	UNet-TPU	1.51%	1.68%	1.59%
	DP-TPU	0.97%	1.07%	1.01%
F=96 unseen	Traditional	11.40%	15.01%	12.14%
	DT-TPU	100%	100%	100%
	UNet-TPU	100%	100%	100%
	DP-TPU	3.04%	4.99%	4.63%

remain between simulated and real-world scenarios. When the projection frequency shifts to the unseen frequency $f_h = 96$, the spatial distribution of sinusoidal fringes changes drastically. Both DT-TPU and UNet-TPU networks experience rapid performance degradation, completely failing in cross-frequency generalization tasks with 100% error rates. This outcome underscores the limitations of purely data-driven networks in maintaining robustness against frequency variations and highlights their constrained generalization capabilities.

After integrating physical priors, the network’s generalization ability improves significantly. At the seen frequency $f_h = 48$, the DP-TPU network further reduces errors, outperforming even the UNet-TPU. Specifically, for MF, the error decreases from 1.51% to 0.97%; for NT, the error decreases from 1.59% to 1.01%; and for MW, the error decreases from 1.68% to 1.07%. All unwrapping accuracies remain above 98.9%, demonstrating the critical role of physical priors in mitigating domain gaps between simulation and reality and enhancing unwrapping precision. More importantly, with physical priors, the network achieves effective generalization for the first time at the unseen frequency $f_h = 96$, reducing errors from 100% to lower than 5%, while all other methods fail completely under this condition. This result unequivocally establishes the pivotal value of physical priors in improving cross-frequency generalization capabilities.

5. Ablation Study on Composite Loss Design

To systematically evaluate the effectiveness of the composite loss function in improving phase unwrapping accuracy and frequency generalization, we designed two comparative experiments under identical network architecture, training data, and optimizer settings. The first model used only Mean Squared Error (MSE) as the loss function, while the second incorporated a Fourier Loss term to form a joint optimization strategy (MSE + FL). Both models were trained using synthetic data.

The results show that across all testing modalities (MF, MW, NT) and fringe frequencies (ranging from 16 to 96), the composite loss consistently outperformed the MSE-only model, with particularly significant improvements at higher frequencies. For example, in the low-frequency range ($f_h \leq 32$), the performance of the two models was relatively similar, with negligible differences in all TPU methods. This indicates that in scenarios with wider fringes and smoother phase transitions, the MSE loss alone is sufficient to guide effective learning.

However, as the frequency increases, the advantages of the composite loss become more evident. Specifically, At the highest frequency $f_h = 96$, the error of MF decreased from 3.08% (MSE) to 2.98% (MSE + FL); for MW, it dropped from 5.05% to 4.64%; and for NT, it decreased from 4.92% to 4.46%. These results demonstrate that Fourier loss helps the network learn the global spectral structure of the phase map in the frequency domain, thereby improving its ability to model high-frequency components and maintain the continuity of fringe structures. Such frequency-domain consistency serves as a valuable complement to conventional MSE loss, particularly in capturing fine details that are often missed. This is especially beneficial in high-frequency scenes where fringe patterns are dense and rapidly varying, effectively suppressing phase-jump errors.

Table S6: Comparison of phase unwrapping errors under different loss function configurations.

Method	MF		MW		NT	
	MSE	MSE + FL	MSE	MSE + FL	MSE	MSE + FL
16	0.11%	0.10%	0.22%	0.20%	0.16%	0.15%
32	0.33%	0.32%	0.72%	0.71%	0.51%	0.51%
48	0.68%	0.66%	0.83%	0.79%	0.70%	0.69%
64	1.41%	1.35%	1.90%	1.85%	1.59%	1.53%
80	2.16%	2.07%	2.62%	2.57%	2.44%	2.35%
96	3.08%	2.98%	5.05%	4.64%	4.92%	4.46%

References

- (1) Zheng, Y.; Wang, S.; Li, Q.; Li, B. Fringe projection profilometry by conducting deep learning from its digital twin. *Optics express* **2020**, *28*, 36568–36583.
- (2) Zhang, Z. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence* **2002**, *22*, 1330–1334.
- (3) Fetić, A.; Jurić, D.; Osmanković, D. The procedure of a camera calibration using Camera Calibration Toolbox for MATLAB. 2012 Proceedings of the 35th International Convention MIPRO. 2012; pp 1752–1757.
- (4) Huang, L.; Zhang, Q.; Asundi, A. K. Camera calibration with active phase target: improvement on feature detection and optimization. **2013**,