

引用格式: SHEN Siyuan, LU Rongsheng, LI Shuge, et al. Deep Learning-Based Multi-Frame Fringe 3D Reconstruction for Dynamic Scenes[J]. Acta Photonica Sinica, 2026, 55(7):0712001

沈思源, 卢荣胜, 李书阁, 等. 基于深度学习的动态场景多帧条纹三维重建技术[J]. 光子学报, 2026, 55(7):0712001

基于深度学习的动态场景多帧条纹三维重建技术

沈思源¹, 卢荣胜¹, 李书阁¹, 施文松²

(1 合肥工业大学 仪器科学与光电工程学院, 合肥 230009)

(2 无锡维度机器视觉产业技术研究院有限公司, 无锡 214101)

摘 要: 条纹轮廓术通过相移条纹图案实现高分辨率的三维重建。然而, 在动态场景中, 物体的运动会引发纹理错位并引入额外的相移, 从而导致重建误差。为此, 本文提出一种基于深度学习的方法, 用于减少由刚体运动引起的动态误差。具体而言, 首先利用深度网络从连续的相移复合条纹图像中快速估计各帧的粗点云; 随后, 通过点云配准将多帧点云统一至同一坐标系, 并反投影至图像平面以实现条纹对齐; 最后, 再次利用网络提取高精度相位, 完成高保真度的三维重建。实验结果表明, 所提方法能够有效抑制动态场景中由物体运动引入的误差, 实现高精度的三维重建。

关键词: 条纹投影; 三维重建; 深度学习; 运动场景; 光学测量

中图分类号: TH741

文献标识码: A

doi:10.3788/gzxb20265507.0712001

0 引言

光学三维测量在工业质检、智能识别、文物保护等领域具有重要的应用价值^[1-4]。条纹投影轮廓术(Fringe projection profilometry, FPP)凭借低成本、高速度、高精度等优势, 被广泛运用于各类场景中。常见的 FPP 系统通过捕捉变形后的条纹, 利用相位解调算法和相位高度映射算法实现三维重建。目前傅里叶变换方法^[3](Fourier transform, FT)和相移法^[5](Phase shifting, PS)是两种主流的相位提取技术。FT 方法仅需要单个条纹图案即可计算相位信息, 但由于图像的频域混叠, 提取的相位存在较大的误差, 重建精度往往不够高^[6,7]。相移法则通过投射多模式条纹, 在时间域上增加采样率来获得更高的相位精度, 然而在动态测量中, 被测物体处于运动状态, 无法保证多个相移条纹投影到物体的同一位置, 这会引入额外的相位误差。为了补偿和缓解物体运动引起重建误差, 国内外众多学者对此进行了大量的研究^[8-13]。Liu Z 通过从运动估计值中推算出运动引起的相位误差偏移, 采用考虑相位偏移误差的通用相移算法进行相位计算^[14]。Wang Y 利用希尔伯特变换对相移后的条纹进行处理, 生成另一组相移条纹, 然后使用两组的平均值进行相位重建^[15]。Liu X 在四步相移法中计算三个相位图, 对比分析相位图的差异, 实现对非均匀表面运动相位误差的补偿^[10]。Guo W 采用了复合的双频条纹, 通过分析双频复合相移条纹的相位关系识别运动区域, 并且采用相移算法与傅里叶分析结合的方法抑制运动引起的误差^[12]。Wang H 在投影序列中添加散斑图案, 利用数字图像相关技术确定投影点位移, 通过重映射的方式校正条纹图案以计算三维形状^[16]。

近年来, 深度学习技术也被广泛运用到条纹投影三维重建中。Li Y 使用深度学习模型直接从单帧不同频率空间复用条纹图案中提取绝对相位信息, 进行动态场景的三维重建^[17]。Zhao Y 通过深度学习模型对低分辨率条纹图像进行细节恢复和超重建, 将分辨率图像恢复到高分辨率图像, 进而提升硬件系统的速度,

基金项目: “太湖之光”科技攻关(产业前瞻及关键技术研发)(G20252001)(注: 请作者在最后一页页脚同时补充基金项目的英文名称)

第一作者: 沈思源, shensiyuan@mail.hfut.edu.cn;

通讯作者: 卢荣胜, rslu@hfut.edu.cn

收稿日期: 2026-01-21; 录用日期: 2026-04-17

<http://www.photon.ac.cn>

实现动态场景的重建^[18]。Guo X使用一个深度学习模型显著提升双频时域相位解缠算法的准确度^[19]。

本文提出了一种面向动态场景的双频复合相移条纹三维重建的方法。传统多帧算法在动态场景测量中易引入运动误差,而单帧深度学习方法虽能抑制运动伪影,却难以充分利用时序信息。为此,本文设计了一个深度学习模型,先从单帧复合条纹中直接提取高精度相位,并据此快速估计深度。随后,引入点云配准算法,将相邻帧的点云对齐后反投影到图像平面,实现物体在图像平面的对齐,最后经多帧深度学习模型处理,获得高保真相位用于最终重建。实验结果表明,所提方法相较于单帧深度学习方法,有效的融合了时序信息,显著提高了动态场景下的重建精度,且避免了传统多帧方法固有的运动误差。

1 原理

1.1 双频复合条纹相移算法

标准FPP系统由一台相机与一台投影仪组成(见图1)。相移条纹法中,系统依次向物体表面投射多幅相移条纹图,相机同步采集后,利用相移公式计算包裹相位,再经相位展开即可实现高精度三维重建。

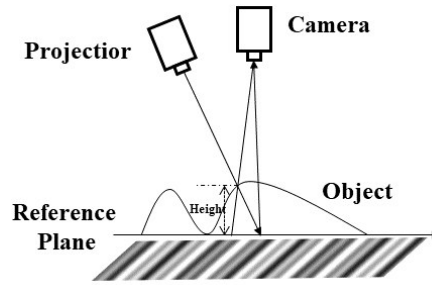


图1 条纹投影系统

Fig. 1 The system of fringe projection profilometry

为了尽可能减少所需的投影数量,可以将两个频率的信息整合到单一的图案中,双频复合相移条纹可以表示为公式(1):

$$I = A + B_1 \cos(\varphi_h + 2\pi n/N) + B_2 \cos(\varphi_l + 4\pi n/N) \quad (1)$$

其中A是平均强度,B是调制度, φ_h 和 φ_l 是编码相位的信息,x和y是相机坐标。当相机采集到变形的复合条纹后,可以通过公式(2)计算出所需的低频与高频的包裹相位。

$$\begin{aligned} \varphi_h(x, y) &= \tan^{-1} \frac{\sum_{n=0}^{N-1} I_n(x, y) \sin(2\pi n/N)}{\sum_{n=0}^{N-1} I_n(x, y) \cos(2\pi n/N)} \\ \varphi_l(x, y) &= \tan^{-1} \frac{\sum_{n=0}^{N-1} I_n(x, y) \sin(4\pi n/N)}{\sum_{n=0}^{N-1} I_n(x, y) \cos(4\pi n/N)} \end{aligned} \quad (2)$$

由公式(1)可知,公式中包含有5个未知数,因此投影模式N最小为5计算出所需要的所有信息。因为公式(2)中的反正切函数计算得到的相位范围在0到 2π 之间,并且是不连续的,所以需要通过相位展开算法消除 2π 不连续性以生成连续相位。本文采用的两个频率遵循数论相位展开算法,该方法时基于数论处理整数的可除性,以下公式可用于获得数论方法中的绝对相位图。

$$\begin{aligned} \phi_l(x, y) &= \varphi_l(x, y) + 2\pi k_l(x, y) \\ \phi_h(x, y) &= \varphi_h(x, y) + 2\pi k_h(x, y) \end{aligned} \quad (3)$$

其中两个高低频的绝对相位满足公式(4):

$$\frac{\phi_h(x, y)}{\phi_l(x, y)} = \frac{\lambda_l}{\lambda_h} \quad (4)$$

通过公式(3)和公式(4),可以计算出一个查找表,当计算出两个频率的包裹相位后,即可通过查找表快

速找到每个像素对应的相位级数,实现相位展开。

$$\frac{\lambda_h \varphi_l(x, y) - \lambda_l \varphi_h(x, y)}{2\pi} = \lambda_l k_h(x, y) - \lambda_h k_l(x, y) \quad (5)$$

$$(k_h, k_l) = LUT \left\{ \text{round} \left(\frac{\lambda_h \varphi_l - \lambda_l \varphi_h}{2\pi} \right) \right\}$$

1.2 深度学习单帧复合条纹相位提取

传统相移法依赖多帧条纹,仅适用于静态场景;在动态测量中,物体运动引入帧间运动相位误差与纹理错位,导致重建失真。单帧重建可从根本上消除帧间误差,实现运动场景的快速三维感知。为此,本文提出一种深度学习模型,从单帧复合条纹中直接预测高精度包裹相位。该方法以监督学习为核心,利用成对标注数据训练网络。模型输入为相机采集的单帧条纹,输出对应两个频率的分子与分母图像(公式6)。获得预测结果后,通过反正切运算即可快速求解包裹相位,无需额外相移。整体训练流程如图2所示

$$\varphi_h(x, y) = \tan^{-1} \frac{M_h(x, y)}{D_h(x, y)} \quad \varphi_l(x, y) = \tan^{-1} \frac{M_l(x, y)}{D_l(x, y)} \quad (6)$$

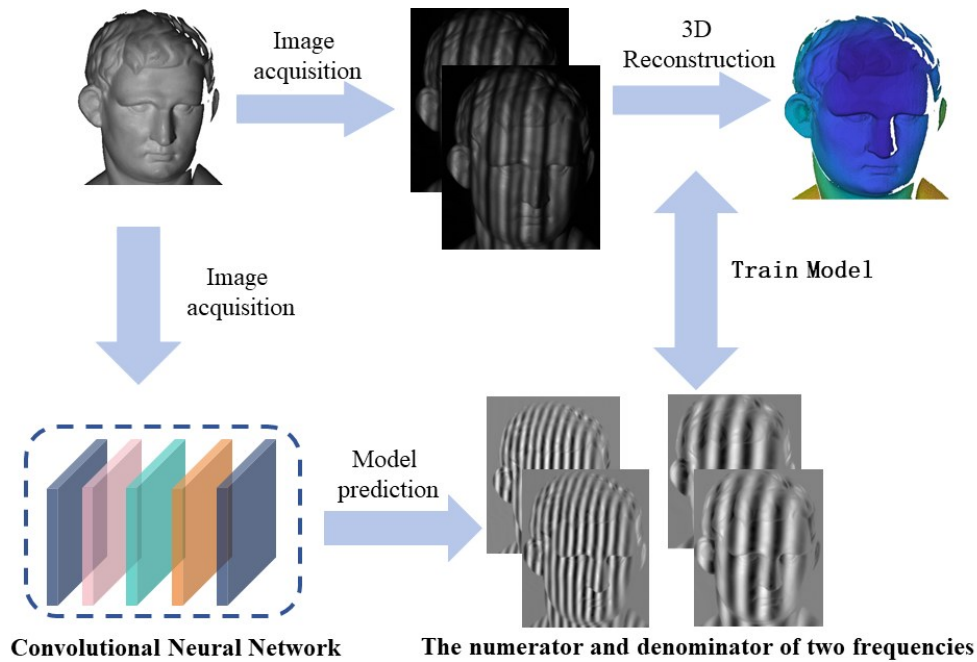


图2 单帧深度学习模型训练流程
Fig. 2 Training process of single-frame deep learning model

网络架构方面,本文提出一种改进型U-Net深度学习模型作为核心框架。U-Net最初面向生物医学图像分割设计,其对称编码器-解码器结构在密集预测任务中展现出优异性能,近年来已被成功拓展至相位恢复等计算机视觉问题。该网络通过逐级下采样捕获多尺度上下文信息,并结合跳跃连接实现高分辨率细节保持,特别适用于条纹图像的相位估计。如图3所示,所改进的U-Net由卷积块与残差块构成的编码器与对称解码器构成。首先,一个通道数为32的卷积块从图像中提取浅层特征;随后,使用步长为2的卷积与残差块交替完成四次下采样,每次下采样都将通道数增加一倍,逐步扩大感受野并压缩

空间维度,从而编码全局条纹模式。解码路径通过双线性插值和卷积逐级恢复空间分辨率,最终输出与输入尺寸一致的分子、分母预测图,以保证相位重建的高分辨率特性。跳跃连接将编码器各层特征与解码器对应层进行通道拼接,使上采样过程能够充分利用局部细节,有效抑制因下采样导致的空间信息损失。损失函数采用均方误差(Mean Squared Error, MSE)。该函数通过最小化模型预测与真值之间差值的平方均值来度量回归误差,其定义如公式(7)所示。

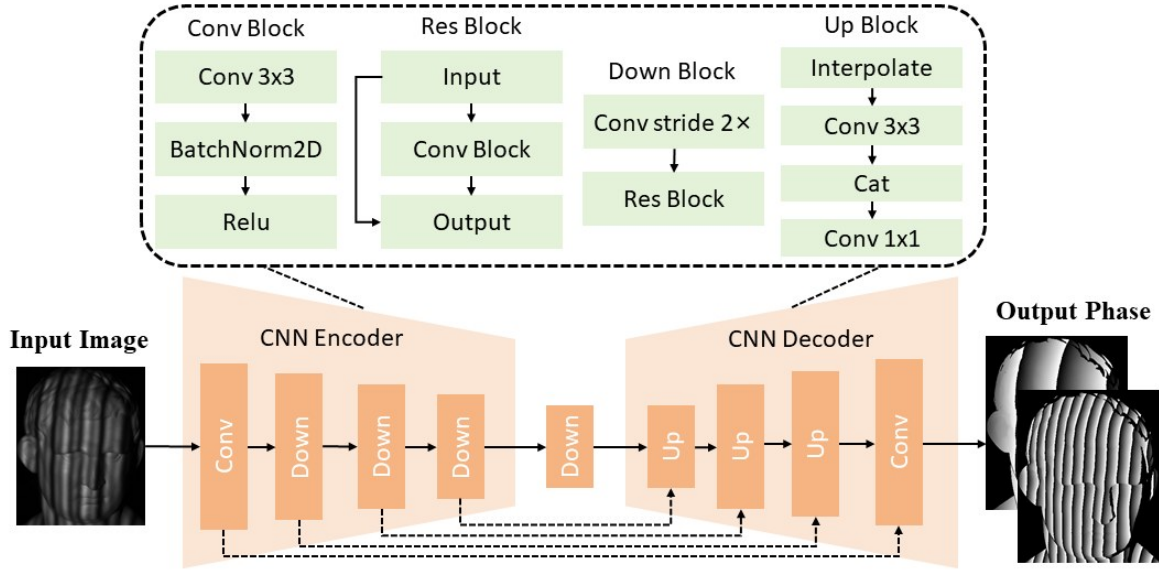


图3 单帧深度学习模型
Fig. 3 Single Frame model

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - p_i)^2 \quad (7)$$

其中N为像素点的总数, y 代表第*i*个像素点的真实标签值, p 代表第*i*个像素点的预测值。在单帧条纹相位提取任务中,模型的输出为四个通道,因此,最终的损失函数是四个通道的MSE损失的总和,以确保所有输出分量都能被同步且准确的优化。

$$Loss = L_{Ml} + L_{Dl} + L_{Mh} + L_{Dh} \quad (8)$$

1.3 基于点云配准的条纹对齐

单帧深度学习模型虽然能够生成逐帧的点云信息,但其推理过程仅依赖当前时刻采集的图像,未能充分挖掘和利用相邻帧之间蕴含的时空相关性,这在一定程度上限制了重建精度与鲁棒性。特别是在基于单帧复合条纹技术的三维重建中,通过在单次曝光中叠加高、低不同空间频率分量,实现了相位信息的空间复用,能够在一帧图像内完成包裹相位提取与初步深度估计。然而,由于多个频率分量在同一灰度动态范围内进行编码,各分量的调制度不可避免地受到压缩,导致有效信噪比下降;同时在频域分离过程中,不同频率谱带易发生混叠,引入相位误差,使细密纹理区域出现平滑化趋势,从而进一步限制了三维重建的精度。由此可见,单帧模型在动态场景或纹理复杂区域下难以充分保证高精度与鲁棒性,需要结合多帧信息或其他补偿策略以提升重建质量。为此,可以考虑结合相邻帧的信息来提升重建的质量,但受物体运动的影响,不同时刻采集到的条纹图像存在偏移,直接用于相位计算会因条纹纹理错位引入显著的误差。因此,进行条纹图案的对齐是进行高精度相位估计的关键。为实现动态场景下的帧间纹理的精准对齐,本文进一步提出一种基于点云配准的亚像素重采样的条纹图对齐策略。

首先以相邻帧的点云作为目标点云,当前帧点云作为源点云,执行点云配准,将源点云配准到目标点云的坐标系中。ICP(Iterative Closest Point)是经典的点云配准算法之一,用于把两片部分重叠的点云通过刚性变换对齐到同一坐标系上。ICP配准的目标函数为:

$$E(R, T) = \sum_{i=1}^n \|q_i - (R \cdot p_i + t)\|^2 \quad (9)$$

公式(9)中, q 和 p 分别为源点云和目标点云, n 是点云对应点总数,在目标点云中寻找欧氏距离最近的点作为源点云的对应点。通过最小二乘法求解刚体变换,使源点云与目标点云之间的欧氏距离之和最小,从而得到最优变换矩阵。

点云配准之后,即可利用公式(12)——针孔相机模型计算当前帧个点在目标帧图像中的亚像素坐标。

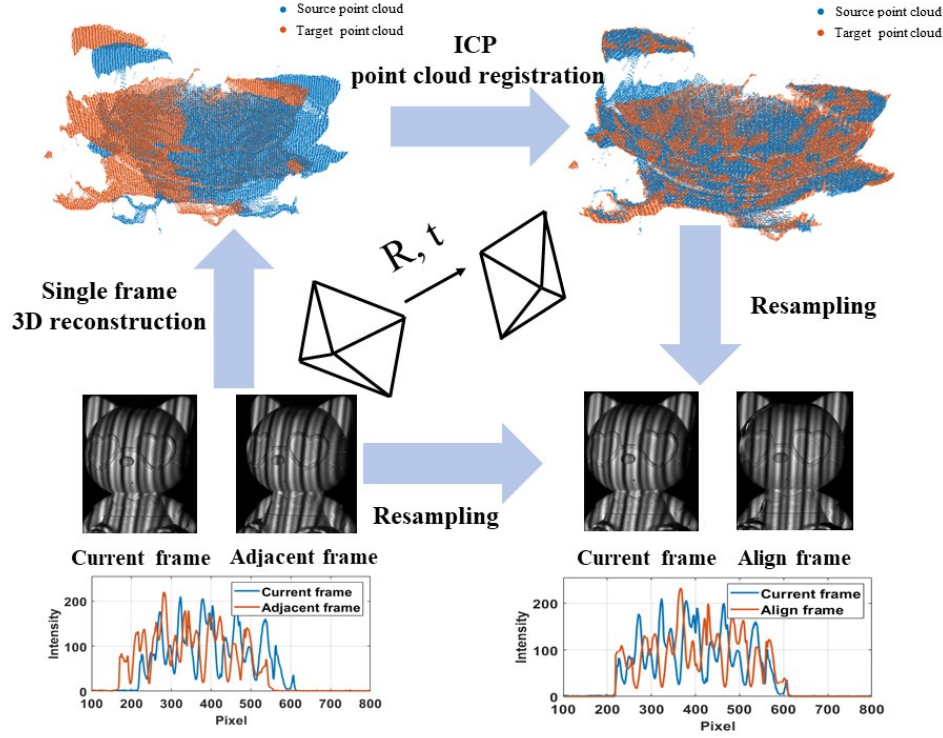


图4 基于点云配准的条纹图案对齐

Fig. 4 Fringe pattern alignment based on point cloud registration

公式(12)为的矩阵表示为:

$$\begin{bmatrix} u^c \\ v^c \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^c \\ y^c \\ z^c \end{bmatrix} \quad (12)$$

$$f_x = \frac{f}{dx} \quad f_y = \frac{f}{dy}$$

其中 (x^c, y^c, z^c) 为相机坐标系下的点的坐标, (u^c, v^c) 为像素坐标, f_x 和 f_y 是焦距的像素单位表示, f 为相机的焦距, (c_x, c_y) 为相机的主点。这些参数可以通过相机标定得到。利用亚像素坐标在目标帧的条纹图上采用双线性插值进行重采样,生成与当前帧几何完全对齐的新条纹图。图4展示了点云配准与重采样的过程。配准前,两帧中的石膏娃娃在空间中存在平移和旋转差异,导致相同位置的点在像素坐标系中出现偏移。通过点云配准与重采样算法处理后,两帧中空间上对应的点被映射到像素坐标系的统一位置,实现了帧间空间的一致性。该处理有效消除了由于物体运动引起的图像偏移,使每帧的输入在空间上保持一致,为后续相位解算提供了可靠的基础数据,从而显著提高了动态测量的精度和稳定性。

1.4 基于增强模块的多帧深度学习模型

对齐后的条纹图在几何上已与当前帧一致,但是其实际可以看作对静止的物体投射未知相移的条纹结构光,无法直接套用传统N步相移公式求解相位。为进一步挖掘时序信息,本文提出一种“单帧提取-多帧增强”的两阶段网络架构。整体框架如图5所示由单帧提取组件与多帧增强组件级联构成:

单帧提取组件与第1.2节中的单帧深度学习模型结构一致,多帧增强部分则使用了一个与单帧提取组件结构相同但输入为2通道的编码器作为共享编码器,用于提取相邻对齐帧与参考帧之间的特征。配准后的相邻四帧图像与参考帧组成四组图像对分别送入共享的编码器中得到与之对应的多尺度特征,然后经由聚合模块以残差的形式注入到单帧组件的特征图中,实现信息增强。每个聚合模块由两个门控模块构成,分别为帧内门控和帧间门控。帧内门控模块用于通过利用3x3卷积、1x1卷积及sigmoid函数生成像素级置信掩膜来对帧内的特征进行筛选;然后将门控后的四帧特征拼接,经1x1卷积层预测四通道注意力图,在

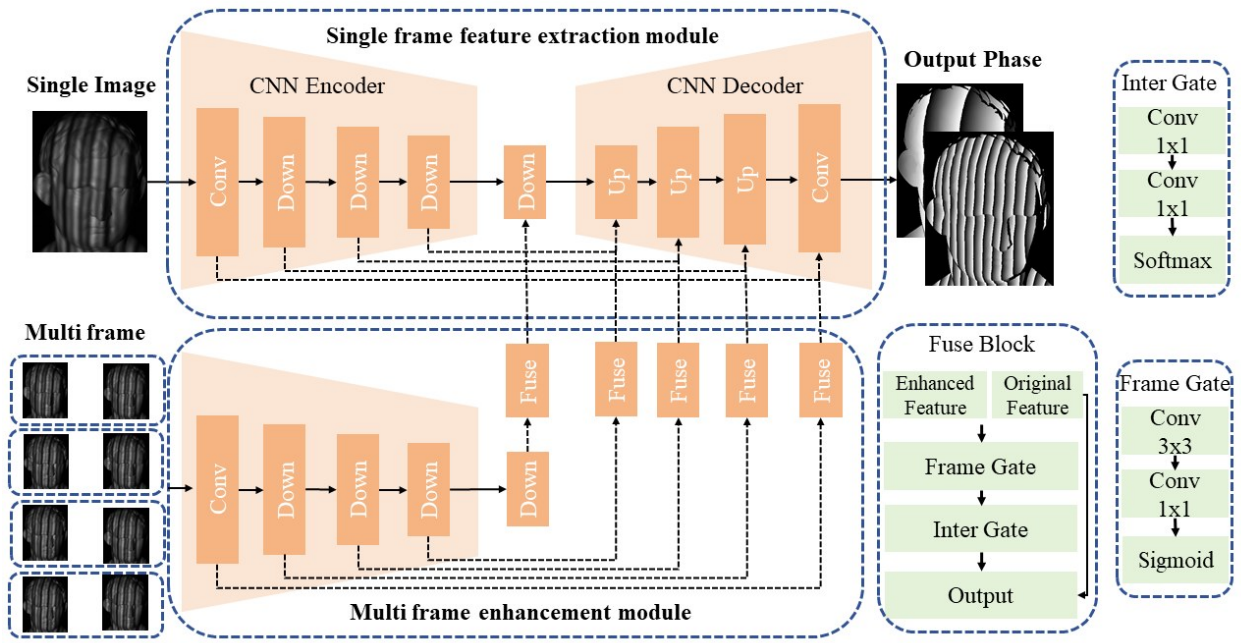


图5 双组件的多帧深度学习模型

Fig. 5 A multi-frame deep learning model with dual components

帧维度执行 Softmax 函数,使得每个像素位置的权重和为 1,从而自适应地融合四个特征图的特征,最后该特征以残差的形式注入到单帧特征图内,整个过程保持端到端可训练。网络输出仍为两个频率的分子/分母四张图,损失函数与基础模型保持一致,即四通道均方误差(MSE),可端到端最小化预测值与真值之间的相位误差。该策略使原有单帧模型具备利用多帧互补信息的能力,显著提升了动态场景下提取的相位精度与鲁棒性。

2 实验

本文的硬件系统搭建如图6所示,主要有相机、投影仪和计算单元组成。相机是一个8位工业黑白相机(Genie Nano series M800),分辨率为 608×800 ,配有一个16mm的镜头。投影仪为DLP-PRO4500,分辨率为 912×1140 。计算单元配有英特尔至强E5-2660v3处理器和英伟达RTX2080Ti显卡。在本系统中,高频条纹频率和低频条纹频率分别为16个像素和33个像素,有效测量区间为相机坐标系下400–600毫米。全部算法在Pytorch1.12.1框架下实现,训练epoch为500,初始学习率为 1×10^{-3} ,采用AdamW优化器。数据集制作分为两部分,对于单帧模型,在严格静止的条件下采集50组原始图像,以20步相移算法生成高精度的分子分母作为真值,总计1000组样本。其中900组用于训练,验证集与测试集各50组。对于多帧模型,在准静态低速运动场景下采集图像序列,利用随机间隔采样策略模拟不同运动速度,构建了10000组样本,其中涵盖了前述的1000组单帧数据集样本,划分训练集为9000组,验证集与测试集各500组。

为验证所提算法的有效性以及系统的测量精度,本文首先以标准球为对象开展静态与动态两组实验,结果如图7所示。图7(a)为测量的静止状态下的标准球,在此对多种相位测量方法进行了对比,包括传统双频复合条纹相移测量方法、移植到复合条纹处理框架下的Wang希尔伯特变换方法、单帧复合条纹深度学习模型、基于配准结果的UNet处理方法,以及本文提出的深度学习模型。从实验结果可以看出,在双频复合条纹相移测量方法中由于复合条纹降低了条纹频率的调制度,易受到噪声影响,其相位解算过程对随机噪声较为敏感,导致重建表面出现明显毛刺,拟合得到的均方根误差为0.0779mm。采用Wang的希尔伯特变换方法后,噪声在一定程度上得到了抑制,均方根误差降低至0.0625mm。单帧复合条纹深度学习模型通过单张图像直接提取相位信息,在球体表面较为光滑的条件下取得了0.0641mm的均方根误差。进一步地,基于配准结果的UNet方法和本文提出的方法分别获得了0.0618mm和0.0516mm的均方根误差,其中本文方法表现出更高的重建精度。

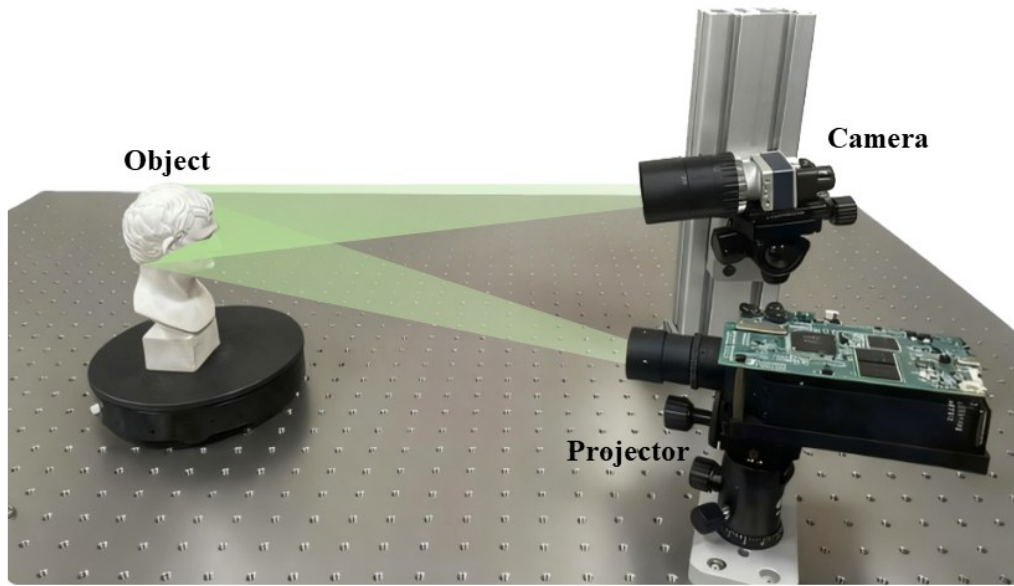


图6 条纹投影三维测量系统
Fig. 6 Fringe Projection 3D Measurement System

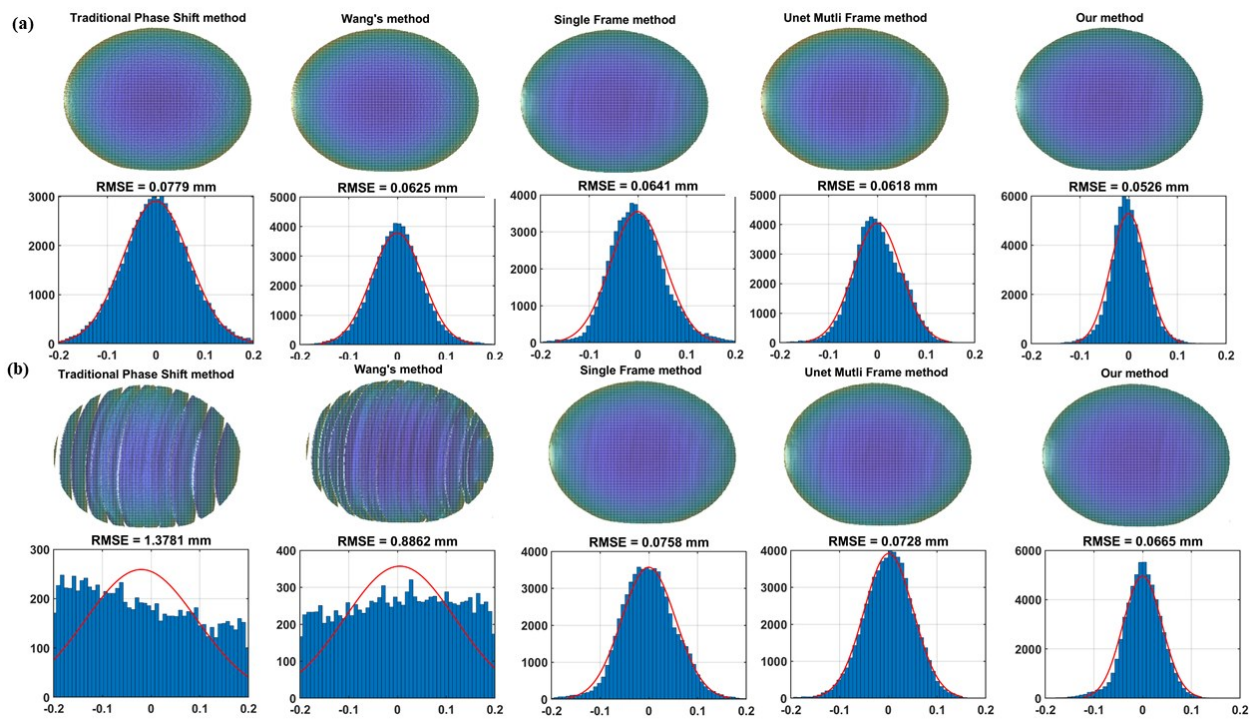


图7 静止和运动的标准球的三维重建
Fig.7 Three-dimensional reconstruction of a stationary and moving standard ball

图 7(b) 展示了标准球在运动状态下的实验结果。对于运动目标,传统双频复合条纹相移测量方法由于相邻帧之间的物体运动会引入非整数像素级位移,破坏了相移假设条件,从而在重建表面中产生明显的波纹状伪影,导致拟合球体的均方根误差显著增大。Wang 的希尔伯特变换方法主要针对微小位移情形设计,在运动幅度较小的情况下能够在一定程度上缓解由运动引起的误差,但当目标运动速度较快时,重建结果中仍然存在较为严重的波纹伪影。单帧复合条纹深度学习模型仅依赖单幅图像即可完成三维重建,从根本上避免了帧间运动带来的影响,因此在运动状态下仍能保持与静止条件下相近的重建精度,其拟合球体的均方根误差为 0.0758mm。对于本文提出的配准框架,在完成多帧条纹图像的对齐后,分别采用标准 UN-

et模型和本文设计的模型进行相位重建,其对应的均方根误差分别为0.0728mm和0.0655mm。实验结果表明,本文方法在运动场景下能够有效抑制由帧间运动引起的伪影,进一步提升三维重建精度。

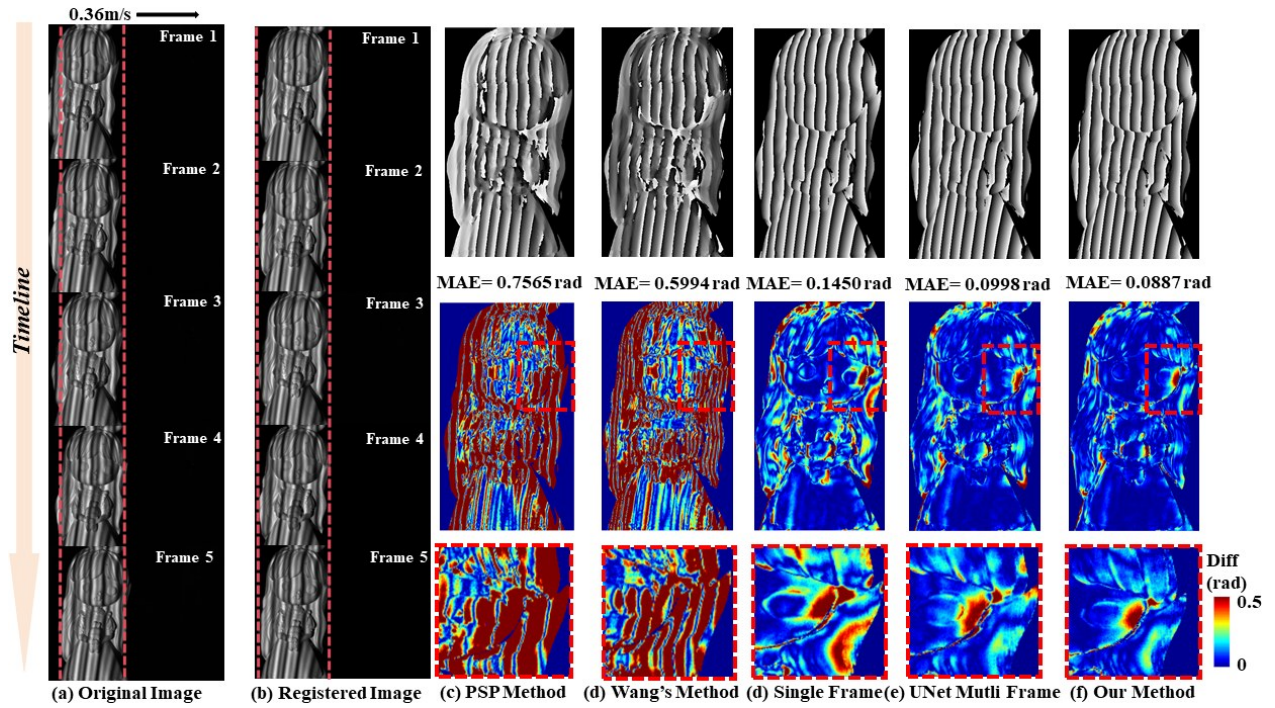


图8 石膏娃娃在不同重建方法下的相位误差对比

Fig.8 Comparison of phase errors of gypsum dolls under different reconstruction methods

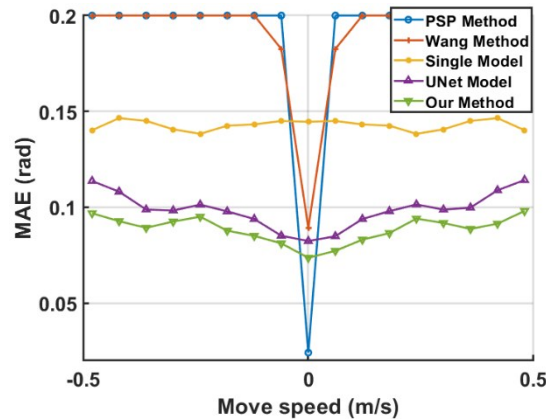


图9 不同速度下不同方法的相位误差

Fig.9 Phase errors of different methods at different speeds

为评估算法在复杂纹理表面上的鲁棒性,本文在完成标准球实验的基础上,选取纹理细节丰富的石膏娃娃作为实验对象。标准球表面光滑,便于量化误差和评估方法在静态与动态场景下的精度,而石膏娃娃则用于进一步验证算法在纹理复杂、结构多变条件下的性能。通过控制石膏娃娃的平移速度,模拟不同运动状态,从而分析算法在动态复杂场景中的三维重建鲁棒性。本文选取纹理细节丰富的石膏娃娃作为实验对象,并通过控制其平移速度来模拟不同运动状态,速度范围为 -0.48m/s 至 0.48m/s 。图8展示了在运动速度为 0.36m/s 时的实验结果。图8(a)给出了五帧复合条纹图像,由于帧间存在微小运动,物体表面纹理在不同帧之间出现明显错位。图8(b)为经过点云配准和双线性重采样后的结果,其中所有帧均被统一配准至第3帧的坐标系下,为后续的多帧信息融合奠定了基础。图8(c) - (d)定量比较了不同方法相对于20步相移测量结果的包裹相位误差。可以看出,传统相移方法由于帧间运动引入非整数像素位移,导致相位计算严重失真,整体相位误差高达 0.7565rad ,重建结果几乎完全失效。引入希尔伯特变换的方法在一定程度上

缓解了运动带来的影响,相位误差降低至 0.5994rad ,但重建质量仍然较差。单帧复合条纹深度学习模型通过利用空间上下文信息进行相位推理,使相位误差显著降低至 0.1450rad 。然而,在石膏娃娃眼周等细密纹理区域,仍然存在局部较大的相位误差。基于多帧配准融合策略的UNet模型和本文提出的模型进一步引入了多帧互补信息,使整体相位误差分别降低至 0.0998rad 和 0.0887rad ,显著提升了细节丰富区域的重建质量。相比之下,本文方法通过多帧特征与低频/高频特征融合的网络结构,实现了更充分的信息利用,因此获得了更低的相位误差。图9给出了不同运动速度下各方法的相位误差变化趋势。当物体处于静止状态时,传统相移方法具有最低的相位误差;但随着运动速度的增加,其相位误差迅速上升。希尔伯特变换方法旨在补偿由物体运动引入的未知相移,但当运动幅度较大时,帧间像素对应关系被破坏,平均强度和调制度的估计误差进一步放大,因此其性能提升仍然有限。单帧模型由于仅依赖单幅图像进行相位恢复,在不同运动速度下均保持相对稳定的重建精度。多帧UNet模型与本文提出的模型通过对配准后的多帧条纹图像进行融合,在各速度条件下均表现出更优的重建性能,尤其在细密纹理区域具有明显优势。其中,本文方法在所有速度范围内均取得了最低的相位误差,进一步验证了所提出模型结构的有效性与鲁棒性。

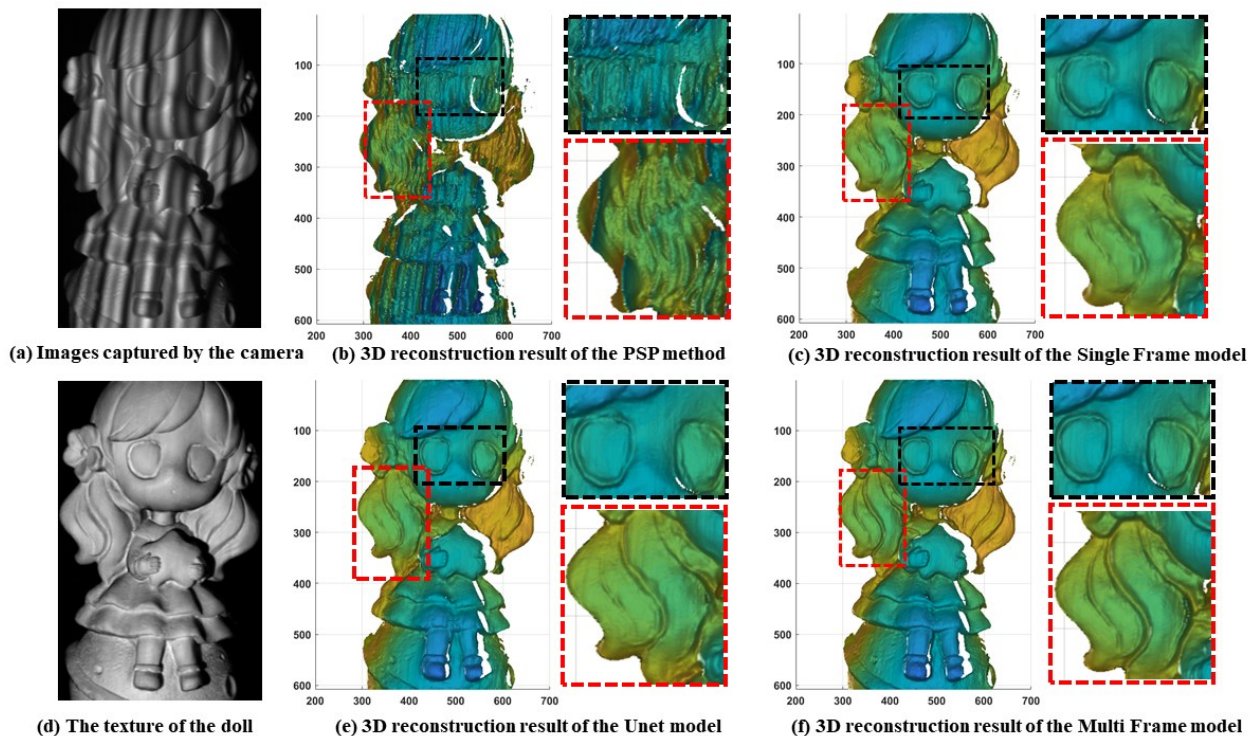


图10 动态场景中的石膏玩偶重建结果

Fig.10 Reconstruction Results of Gypsum Dolls in Dynamic Scenes

在真实动态场景中,石膏玩偶被放置于旋转台外侧,使其运动包含平移与旋转两种复杂运动,以此验证算法稳定性。图10呈现了玩偶的三维重建结果与对比分析:(a)为摄像头捕获的原始图像,清晰展示了玩偶的整体轮廓与姿态;(d)则聚焦于玩偶的原始纹理细节,如发丝、裙摆等复杂结构。针对三维重建效果,图(b)至(f)分别对应不同方法:(b)传统的标准相移方法重建结果整体形状偏差较大,细节缺失严重;(c)单帧模型虽能大致还原玩偶形态,但在眼部、发丝等细节区域存在明显不连续性与误差;(e)为使用多帧UNet模型的重建结果,其细节处理上较单帧模型有所提升,但依旧存在局部纹理模糊问题;(f)多帧模型则凭借时间冗余信息融合优势,不仅整体形状还原精准,更在红色虚线框标注的局部细节放大区域展现出卓越的纹理清晰度,有效克服了动态场景带来的挑战。对比可见,多帧模型在处理复杂动态场景时,展现出更为出色的稳定性和细节还原能力。

3 结论

本文提出了一种面向动态场景的多帧复合条纹三维重建方法。通过点云配准与图像重采样技术,将多帧时空信息有效融入深度学习框架,实现运动物体的高精度三维重建。实验结果表明,相较于传统单帧模型,所提多帧模型方法在鲁棒性与重建精度方面均表现出显著优势,尤其在复杂纹理和微小运动条件下,能够有效抑制运动伪影,提升细节还原能力。

然而,当前方法仍存在一定局限性。由于依赖点云配准技术,本方法目前仅适用于物体刚体运动的场景,对于存在明显非刚性形变或复杂形变行为的物体,如柔性材料或其他可变形结构,于其形状在运动过程中持续发生变化,难以通过点云之间建立稳定的对应关系进行可靠配准,从而也难以利用多帧数据进行联合优化。此外,当物体运动幅度较大、相邻帧之间的点云重叠区域较少时,点云配准的稳定性与准确性会明显下降,进而影响多帧优化过程的有效性,最终可能降低整体三维重建的精度与鲁棒性。

针对上述局限性,未来的研究将重点探索更适用于复杂动态场景的三维重建方法。首先,考虑到近年来深度学习大模型在视觉理解与语义感知方面的快速发展,可以利用大模型所提供的高层语义信息与全局上下文特征,构建基于语义约束的稠密像素配准方法,在图像域中建立相邻帧之间的像素级对应关系。与传统依赖几何特征或点云结构的配准方式相比,这类方法能够在弱纹理区域、局部遮挡以及复杂运动情况下仍然保持较好的匹配稳定性,从而在一定程度上替代传统点云配准过程,实现对动态场景中连续运动信息的有效建模。

在此基础上,可以进一步结合深度学习驱动的稠密运动场预测方法,对相邻帧之间的像素运动进行更加精细的描述,并将该运动信息引入三维重建流程中,以实现动态变化的补偿与优化。同时,通过在图像域直接建立连续帧之间的运动关系,可以减少对显式点云配准的依赖,从而提升算法在大幅运动或点云重叠区域较小情况下的稳定性。因此,通过融合语义信息、像素级运动估计以及几何约束,有望构建一种更加鲁棒且高效的动态三维重建框架,从而进一步拓展该方法在复杂动态场景中的应用范围,并提升其在实际三维测量与动态感知任务中的应用价值。

参考文献

- [1] Curless, Brian. From range scans to 3D models. ACM SIGGRAPH Computer Graphics. 1999;33 (4):38-41.
- [2] Srinivasan V, Liu H, Halioua M. Automated phase-measuring profilometry of 3-D diffuse objects. Applied Optics. 1984; 23 (18):3105-08. 10.1364/AO.23.003105
- [3] Takeda M, Mutoh K. Fourier transform profilometry for the automatic measurement of 3-D object shapes. Applied Optics. 1983;22 (24):3977.
- [4] Duda R.O., Hart P.E., Pattern classification and scene analysis, Pattern classification and scene analysis1973.
- [5] Zuo C, Feng S, Huang L, et al. Phase shifting algorithms for fringe projection profilometry: A review. Optics and Lasers in Engineering. 2018;109:23-59. 10.1016/j.optlaseng.2018.04.019
- [6] Han M, Chen W. Two-dimensional complex wavelet with directional selectivity used in fringe projection profilometry. Optics Letters. 2021;46 (15). 10.1364/ol.420460
- [7] Zuo C, Tao T, Feng S, et al. Micro Fourier Transform Profilometry (μ FTP): 3D shape measurement at 10,000 frames per second. Optics and Lasers in Engineering. 2018;102:70-91. 10.1016/j.optlaseng.2017.10.013
- [8] Qian J, Tao T, Feng S, et al. Motion-artifact-free dynamic 3D shape measurement with hybrid Fourier-transform phase-shifting profilometry. Opt Express. 2019;27 (3):2713-31. 10.1364/OE.27.002713
- [9] Lu L., Jia Z., Luan Y., et al. Reconstruction of isolated moving objects with high 3D frame rate based on phase shifting profilometry. Optics Communications. 2019;438:61-66. 10.1016/j.optcom.2018.12.092
- [10] Liu X, Tao T, Wan Y, et al. Real-time motion-induced-error compensation in 3D surface-shape measurement. Optics Express. 2019;27 (18). 10.1364/oe.27.025265
- [11] Duan C, Tong J., Lu L, et al. Improving the Performance of 3D Shape Measurement of Moving Objects by Fringe Projection and Data Fusion. IEEE Access. 2021;9:34682-91. 10.1109/access.2021.3061415
- [12] Guo W, Wu Z, Li Y, et al. Real-time 3D shape measurement with dual-frequency composite grating and motion-induced error reduction. Opt Express. 2020;28 (18):26882-97. 10.1364/OE.403474
- [13] Wang Y, Suresh V, Li B. Motion-induced error reduction for binary defocusing profilometry via additional temporal sampling. Optics Express. 2019;27 (17). 10.1364/oe.27.023948
- [14] Liu Z, Zibley P.C, Zhang S. Motion-induced error compensation for phase shifting profilometry. Optics Express. 2018;26 (10). 10.1364/oe.26.012632

- [15] Wang Y, Liu Z, Jiang C., et al. Motion induced phase error reduction using a Hilbert transform. *Optics Express*. 2018;26 (26). 10.1364/oe.26.034224
- [16] Wang H, Wang Y., Chen Z, et al. Real-time motion-induced error reduction for phase-shifting profilometry with projection points tracking method. *Measurement*. 2025;239. 10.1016/j.measurement.2024.115450
- [17] Li Y, Qian J, Feng S, et al. Deep-learning-enabled dual-frequency composite fringe projection profilometry for single-shot absolute 3D shape measurement. *Opto-Electronic Advances*. 2022;5 (5):210021-21. 10.29026/oea.2022.210021
- [18] Sui C, He K, Lyu C, et al. Accurate 3D Reconstruction of Dynamic Objects by Spatial-Temporal Multiplexing and Motion-Induced Error Elimination. *IEEE Transactions on Image Processing*. 2022;31:2106-21. 10.1109/tip.2022.3150297
- [19] Guo X, Li Y, Qian J, et al. Unifying temporal phase unwrapping framework using deep learning. *Opt Express*. 2023;31 (10):16659-75. 10.1364/OE.488597

Deep Learning-Based Multi-Frame Fringe 3D Reconstruction for Dynamic Scenes

SHEN Siyuan¹, LU Rongsheng¹, LI Shuge¹, SHI Wensong²

(1 *School of Instrument Science and Opto-electronics Engineering, HeFei University of Technology, HeFei 230009, China*)

(2 *Wuxi Dimension Machine Vision Industry Technology Research Institute Co., Ltd., Wuxi 214101, China*)

Abstract: Optical three-dimensional (3D) measurement techniques, particularly fringe projection profilometry (FPP), have been widely adopted in industrial inspection, intelligent recognition, and cultural heritage preservation due to their high accuracy, low cost, and fast acquisition speed. However, conventional FPP methods typically rely on multiple phase-shifted fringe patterns captured under static conditions. In dynamic scenes, object motion introduces inter-frame misalignment and additional phase shifts, leading to significant reconstruction errors. To address this challenge, we propose a novel deep learning-based multi-frame fringe 3D reconstruction method specifically designed for dynamic scenes with rigid motion. Our approach consists of three main stages. First, a deep neural network is employed to rapidly estimate coarse 3D point clouds from single-frame composite fringe images. This network is trained to predict the numerator and denominator components required for phase calculation, enabling direct phase retrieval from a single image without the need for traditional multi-step phase shifting. Second, to leverage temporal information while compensating for motion-induced misalignment, we introduce a point cloud registration step. Using the Iterative Closest Point (ICP) algorithm, we align the point clouds of adjacent frames into a common coordinate system. The aligned point clouds are then back-projected onto the image plane to generate motion-compensated fringe patterns. Finally, a multi-frame deep learning enhancement module processes the aligned fringe images to extract high-fidelity phase information. This module integrates spatial and temporal features through a dual-component network architecture, combining a single-frame phase extraction backbone with a multi-frame enhancement encoder and adaptive feature fusion mechanisms. The performance of the method was tested in both static and dynamic scenes in this article. Results demonstrate that the proposed method significantly outperforms traditional phase-shifting methods and single-frame deep learning approaches. In static scenes, the multi-frame model demonstrated superior performance by reducing the reconstruction error compared to its single-frame counterpart. In dynamic scenes, it effectively eliminated motion artifacts such as ripple distortions and achieved a substantial reduction in phase error compared to conventional methods. Moreover, the method showed strong robustness in reconstructing fine surface details, even under complex motion conditions. In conclusion, this study presents a robust and accurate solution for 3D reconstruction in dynamic environments by integrating deep learning with image alignment techniques. The proposed method successfully combines the advantages of single-frame inference speed and multi-frame temporal consistency, offering a promising direction for real-time high-precision 3D measurement in motion scenarios. Future work will focus on extending the method to non-rigid objects and improving computational efficiency for real-time applications.

Key words: Fringe Projection; 3-D reconstruction; Deep Learning; Dynamic Scence; Optical

Measurements

OCIS Codes: 120.2650; 120.5050; 120.2830

CSTR: 32255.14.gzxb20265507.0712001

Foundation item: "Light of the Taihu Lake Lake" scientific and technological breakthrough (industry foresight and key technology research and development) (G20252001)