

# 基于深度光流的结构简谐位移响应监测方法

郭晋<sup>1</sup>, 许维炳<sup>1\*</sup>, 黎家玮<sup>2</sup>, 孙玉龙<sup>3</sup>, 周大兴<sup>3</sup>, 周春娟<sup>1,4</sup>, 姬勇刚<sup>5</sup>, 曹康<sup>5</sup>

<sup>1</sup>北京工业大学建筑工程学院, 北京

<sup>2</sup>北京市公路事业发展中心, 北京

<sup>3</sup>中铁建设集团有限公司, 北京

<sup>4</sup>陕西省建筑科学研究院有限公司, 陕西 西安

<sup>5</sup>云南省交通投资建设集团有限公司, 云南 昆明

收稿日期: 2026年5月22日; 录用日期: 2026年6月14日; 发布日期: 2026年6月25日

## 摘要

传统接触式位移测量方法存在对场地要求高、现场施工周期长、经济性差等问题。鉴于此, 本文通过图像预处理、相机标定、畸变校正, 并结合稠密光流估计网络, 提出一种基于深度学习的结构简谐位移监测方法。结果表明, 所提出方法对规则正弦振动条件下的结构简谐位移响应时程具有较好的识别效果, 视觉位移时程与激光位移计结果具有较高一致性。视觉位移时程相对于激光位移计基准结果的RMSE为0.208~0.246 mm, 峰值误差为0.692~0.780 mm, 相对误差为13.84%~15.60%。

## 关键词

结构简谐位移响应, 计算机视觉, 深度学习光流网络

# Structural Harmonic Displacement Response Monitoring Method Based on Deep Optical Flow

Jin Guo<sup>1</sup>, Weibing Xu<sup>1\*</sup>, Jiawei Li<sup>2</sup>, Yulong Sun<sup>3</sup>, Daxing Zhou<sup>3</sup>, Chunjuan Zhou<sup>1,4</sup>, Yonggang Ji<sup>5</sup>, Kang Cao<sup>5</sup>

<sup>1</sup>College of Architecture and Civil Engineering, Beijing University of Technology, Beijing

<sup>2</sup>Beijing Highway Development Center, Beijing

<sup>3</sup>China Railway Construction Group Co., Ltd., Beijing

<sup>4</sup>Shaanxi Academy of Building Science Co., Ltd., Xi'an Shaanxi

<sup>5</sup>Yunnan Communications Investment & Construction Group Co., Ltd., Kunming Yunnan

Received: May 22, 2026; accepted: June 14, 2026; published: June 25, 2026

\*通讯作者。

文章引用: 郭晋, 许维炳, 黎家玮, 孙玉龙, 周大兴, 周春娟, 姬勇刚, 曹康. 基于深度光流的结构简谐位移响应监测方法[J]. 土木工程, 2026, 15(6): 50-56. DOI: 10.12677/hjce.2026.156154

## Abstract

Traditional contact-based displacement measurement methods suffer from stringent site requirements, long on-site installation periods, and poor cost-effectiveness. To address these limitations, this study proposes a deep learning-based monitoring method for structural harmonic displacement response through image preprocessing, camera calibration, distortion correction, and dense optical flow estimation. The results show that the proposed method achieves good identification performance for structural harmonic displacement responses under regular sinusoidal excitation, and the vision-based displacement measurements are in good agreement with those obtained from the laser displacement meter. The RMSE between the vision-based displacement time histories and the laser-based reference results ranges from 0.208 to 0.246 mm, the peak error ranges from 0.692 to 0.780 mm, and the relative error ranges from 13.84% to 15.60%.

## Keywords

Structural Harmonic Displacement Response, Computer Vision, Deep Learning-Based Optical Flow Network

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

结构动态位移响应能够直接反映结构变形状态、刚度退化及动力性能变化,是表征结构服役状态与安全水平的重要指标[1][2]。然而,受现场安装空间及运营条件等限制,拉线位移计、百分表等传统接触式测量方法在现场实施难度较大;GPS、全站仪和激光多普勒等非接触式方法虽可避免与结构直接接触,但仍存在采样频率有限、布设复杂或设备成本较高等不足。与之相比,基于计算机视觉的位移监测方法具有非接触、低成本、布设灵活和易于远距离观测等优势[3]。

近年来,基于计算机视觉的结构位移监测方法已在桥梁、振动台模型、拉索及地下交通枢纽等对象中得到研究。Yoon 等[4]利用消费级相机实现了无靶标结构动力识别;Feng 等[5]构建了非接触视觉位移传感系统;Won 等[6]提出基于参考帧的 Deepflow 位移测量方法,通过稠密光流计算与信号滤波减弱漂移问题;Dong 等[7]进一步将深度学习全场光流引入结构位移监测,说明稠密运动场能够为位移提取提供更丰富的信息。尽管如此,现场监测中普遍存在复杂背景、局部反光、弱纹理等问题,传统视觉方法在复杂场景下仍易出现误匹配、漂移累积及识别稳定性不足等现象。

针对上述问题,Xu 等[8]提出的光流估计网络重新表述为全局匹配问题,并借助 Transformer 架构[9]特征间的全局关联能力,使其在大位移、弱纹理及复杂背景条件下表现出更好的适应性。基于此,本文采用基于全局匹配的深度学习光流估计框架,构建结构简谐位移响应的单目视觉监测方法,并通过试验验证其对结构端部简谐位移响应的识别精度,为结构简谐振动响应的非接触测量提供试验依据和方法参考。

## 2. 基于深度光流的简谐位移响应监测方法

### 2.1. 相机标定

基于计算机视觉的结构位移响应监测系统中,原始图像中的像素坐标仅表征二维图像平面信息。为

了定量分析结构在三维物理空间中的位移行为，需要建立由二维像素坐标系到三维世界坐标系的映射关系。相机作为数据采集传感器，其成像几何模型的准确性直接影响位移测量精度，因此相机标定的核心目的在于求解相机内参数、外参数及畸变参数，为像素位移向真实物理位移的转换提供依据。

相机成像的方法基于小孔成像模型，如图 1 所示。

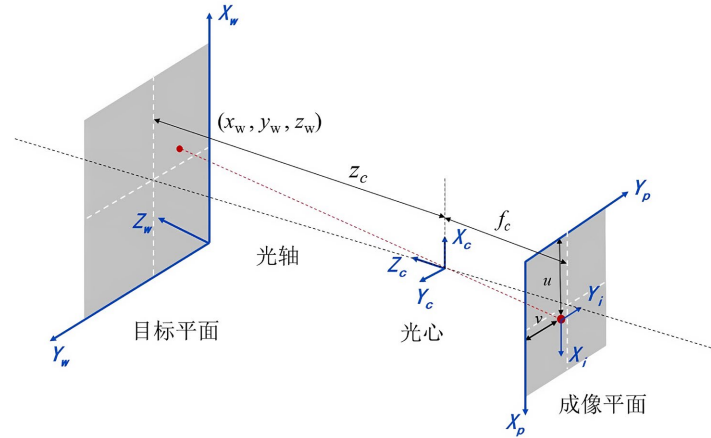


Figure 1. Schematic of the camera model

图 1. 相机模型示意图

相机模型涉及四个独立的坐标系，分别为世界坐标系( $X_w, Y_w, Z_w$ )、相机坐标系( $X_c, Y_c, Z_c$ )、图像坐标系( $x, y$ )和像素坐标系( $u, v$ )。设空间中存在一个目标点，其世界坐标为( $x_w, y_w, z_w$ )，则其像素坐标可通过下式计算：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} s_x & 0 & c_x \\ 0 & s_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_c/z_c & 0 & 0 & 0 \\ 0 & f_c/z_c & 0 & 0 \\ 0 & 0 & 1/z_c & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (1)$$

其中  $f_c$  表示相机焦距， $c_x$  和  $c_y$  表示图像坐标系原点的像素坐标； $s_x$  和  $s_y$  表示单位长度的像素数， $r_{ij}$  ( $i, j = 1, 2, 3$ ) 为世界坐标系与相机坐标系之间的刚体旋转运动参数，它们构成一个  $3 \times 3$  的旋转矩阵  $R$ 。 $t_i$  ( $i = 1, 2, 3$ ) 表示世界坐标系与相机坐标系之间的刚体平移运动参数，它们构成一个  $3 \times 1$  的平移向量  $t$ 。 $R$  矩阵和  $t$  向量统称为相机外参数。等式右边第一个矩阵称为相机内参数矩阵，描述相机焦距、主点位置及像素尺度等成像几何关系。

目前常用的相机标定方法为张正友标定法，其引入了相机径向与切向畸变模型。径向畸变可用三个系数( $k_1, k_2, k_3$ )描述，切向畸变可用两个系数( $p_1, p_2$ )描述， $r$  为归一化像平面上目标点到主点的径向距离。则畸变后的归一化坐标( $x_{dist}, y_{dist}$ )可表示为：

$$\begin{cases} x_{dist} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2 (r^2 + 2x^2) \\ y_{dist} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_2 xy + p_1 (r^2 + 2y^2) \end{cases} \quad (2)$$

本文使用黑白方块间隔组成的棋盘格作为标定板，分别从不同角度、位置和姿态拍摄标定物，并输入 Matlab 中的 camera calibration 工具箱完成标定。试件端部主要沿单一方向发生平面内位移，本文依据试验场景中已知尺寸参照物与成像尺度关系确定像素—物理位移换算系数，并据此将光流输出转换为结构端部实际位移时程。

## 2.2. 基于深度学习的光流估计网络

GMFlow (Learning Optical Flow via Global Matching)是一种结合全局匹配和 Transformer 架构的深度学习光流估计网络,其网络架构如图 2 所示。

该网络架构主要由三部分组成:Transformer 特征增强层、进行相关性计算和 softmax 的全局特征匹配层,以及一个光流传播的自注意力层。对于输入的两个视频帧  $I_1$  和  $I_2$ ,首先使用权重共享卷积网络提取下采样的密集特征  $F_1, F_2 \in \mathbb{R}^{H \times W \times D}$ ,其中  $H$ 、 $W$ 、 $D$  分别代表图像的高度、宽度和特征维度。

与传统依赖局部代价体回归的光流方法不同,GMFlow 将光流估计视为全局匹配问题,因此对特征的判别性和跨帧一致性提出了更高要求。为获得更适合全局匹配的高质量特征表示,GMFlow 在卷积骨干网络提取初始特征后,引入 Transformer 对两帧特征进行增强。

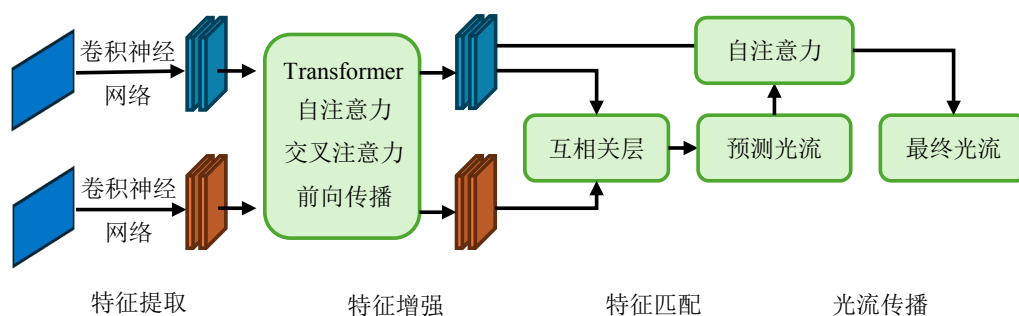


Figure 2. GMFlow network architecture diagram  
图 2. GMFlow 网络架构图

在获得增强特征后,GMFlow 不再采用传统局部代价体回归方式,而是通过全局匹配直接建立两帧图像间的像素对应关系。具体而言,网络通过计算两帧特征之间的全局相关性,并结合 softmax 归一化获得匹配概率分布,进而对第二帧像素坐标进行加权求和,得到第一帧像素在第二帧中的对应位置。最终,初始光流由对应坐标与原始坐标之差确定,其表达式为:

$$\hat{V} = \hat{G} - G \in \mathbb{R}^{H \times W \times 2} \quad (3)$$

其中,表示匹配得到的对应像素坐标矩阵, $G$  表示原始像素坐标矩阵, $\hat{G}$  为加权后目标帧像素坐标矩阵, $\hat{V}$  表示初始光流结果。该方法具有可微性,能够支持端到端训练,并可实现亚像素级匹配。

## 3. 方法验证

### 3.1. 试验工况

Table 1. Experimental condition design  
表 1. 试验工况设计

工况号	激励波形	激励频率/Hz	激励幅值/mm
1	正弦波	1	5
2	正弦波	5	5

为验证所提基于 GMFlow 的结构简谐位移响应监测方法,本文在室内开展悬臂梁振动台试验。试件采用单柱悬臂简化形式,由工字钢梁与圆钢柱焊接而成,悬臂段长 1757 mm,工字钢腹板厚度 10 mm,圆钢柱高度 1520 mm,直径 200 mm,底部通过厚钢板与振动台刚性连接。试验激励工况设置见表 1。视

频图像序列通过智能手机进行拍摄，相机分辨率设置为 4 K (3840 × 2160)，采样频率为 60 Hz，通过三脚架布置在试件前方约 2.0 m 处，摄像头正对悬臂梁端部区域。为获取对应位置结构动力响应，在悬臂梁端部布置 CMOS 微型激光位移传感器，通过有线方式接入 JMTEST 动态信号测试分析系统。

### 3.2. 数据处理

为保证视觉测量结果的准确性与稳定性，需对视频数据进行图像预处理。对采集到的原始视频逐帧解算，并依据悬臂端部的空间位置对图像进行裁剪，仅保留包含结构主要振动部位及其邻近区域的感兴趣区域(region of interest, ROI)。采用 ROI 处理的目的在于减少无关背景、边界扰动及局部光照变化对后续光流估计的影响，同时降低网络推理的计算复杂度。将裁剪后的图像序列进一步采用双边滤波对图像进行降噪处理，该方法在抑制噪声的同时能够较好保留边缘与纹理细节。纹理筛选采用 Sobel 算子计算图像灰度梯度幅值，卷积核尺寸取 3，并保留梯度幅值排序前 30%的像素；同时规定有效纹理像素数量不少于 300。并根据相机标定结果进行畸变矫正与归一化，以提高图像质量并增强帧间灰度与几何一致性，为后续稠密光流求解提供稳定输入。

图像预处理完成后，将参考帧与各时刻当前帧输入 GMFlow 网络，获得参考状态至当前状态的稠密光流场，并据此重构悬臂端部动态位移时程。

### 3.3. 结果分析

为定量评价视觉识别结果与激光位移计基准结果之间的一致性，采用峰值误差(Peak Error)、相对误差(Relative Error)、均方根误差 RMSE (Root Mean Square Error)进行评价，其中，RMSE 用于衡量整个位移时程的总体偏差，峰值误差反映对极值响应的重构能力。计算公式如下：

$$E_p = \max(|d_{l,i} - d_{v,i}|) \quad (4)$$

$$E_r = \frac{|d_{l,i} - d_{v,i}|}{d_{l,i}} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (d_{l,i} - d_{v,i})^2} \quad (6)$$

其中  $N$  为视频帧的总数， $d_l$  表示第  $i$  时刻位移计的监测位移值， $d_v$  表示第  $i$  时刻视觉监测的位移值。图 3 和图 4 给出了两种工况下的位移时程对比图。

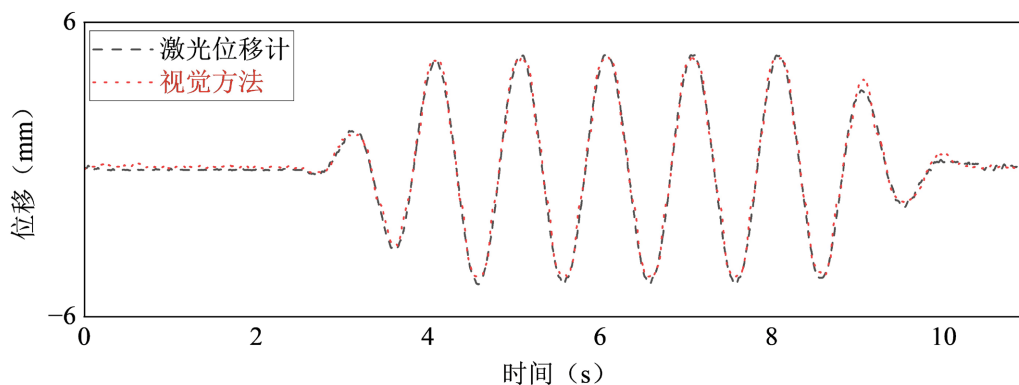
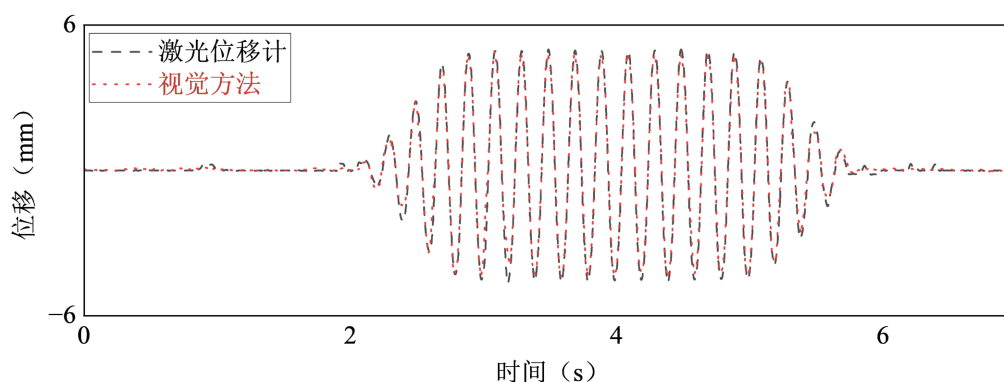


Figure 3. Comparison of displacement time histories under Condition 1

图 3. 工况 1 位移时程对比图



**Figure 4.** Comparison of displacement time histories under Condition 2

**图 4.** 工况 2 位移时程对比图

由图 3 和图 4 可知，所提出方法对简谐振动条件下的结构位移响应时程具有较好的重构效果，视觉识别位移结果与激光位移计位移结果之间保持了较高的一致性。但视觉测量结果在局部峰值附近仍存在一定幅值偏差，该方法在极值响应表征方面与整体时程重构精度之间仍存在一定差异。表 2 给出了两组工况下的误差统计结果。

**Table 2.** Error results based on computer vision methods

**表 2.** 基于计算机视觉方法的误差结果

工况号	RMSE/mm	峰值误差/mm	相对误差/%
1	0.208	0.692	13.84
2	0.246	0.780	15.60

由表 2 可知，视觉位移相对于激光位移计位移的 RMSE 分别为 0.208 mm 和 0.246 mm，均小于 0.25 mm，所提出方法在整个位移时程范围内具有较小的总体偏差。对应的峰值误差分别为 0.692 mm 和 0.780 mm，相对误差分别为 13.84% 和 15.60%，表明视觉测量结果能够较好重构结构端部位移峰值，但在峰值幅值测量方面存在一定偏差。此外，与时程整体误差相比，峰值误差相对更为突出，反映出视觉方法在极值响应重构方面仍受局部匹配精度、图像质量及瞬时位移变化速率等因素影响。

#### 4. 结论与展望

针对传统接触式测量方法布设困难、复杂场景下视觉识别稳定性不足等问题，本文提出了基于 GMFlow 的结构简谐位移响应识别方法，并通过室内实验室试验验证。主要结论如下：

1) 构建了基于深度学习全局匹配光流的简谐位移响应识别流程。该方法通过图像预处理、相机标定与畸变校正，建立了由视频图像到结构实际位移时程的转换路径。

2) 试验结果表明，在实验室受控条件和规则简谐激励条件下，所提方法能够较好重构悬臂端部动态位移时程。视觉识别位移结果与激光位移计位移结果之间保持了较高的一致性，表明该方法在理想化实验室条件下能够有效表征规则简谐激励条件下结构端部的动态响应过程。

3) 误差统计结果显示，视觉位移相对于激光位移计基准结果的 RMSE 为 0.208~0.246 mm，峰值误差为 0.692~0.780 mm，相对误差为 13.84%~15.60%，表明所提方法在简谐激励条件下具有较好的位移时程重构能力，但在峰值幅值表征方面仍存在一定偏差。

4) 本文方法仍存在一定局限性。当前试验中 ROI 区域主要依赖人工选取，尚未实现目标区域的自动

识别与长期稳定跟踪；相机在试验过程中默认保持固定，尚未系统考虑现场环境中相机自身振动、支架扰动和视角变化对测量结果的影响；此外，本文主要识别二维图像平面内位移，尚不能直接获得结构三维空间位移响应。

后续研究可围绕方法的工程适用性进一步展开。引入目标检测方法，实现测区的自动识别与持续跟踪，以降低人工 ROI 选取对测量结果的影响；结合参考靶标、背景特征点匹配或图像稳像方法，对相机振动及视角扰动进行修正；此外，可融合双目视觉或多相机信息，拓展结构三维位移识别能力。在复杂光照、背景干扰及实际工程环境下开展验证，以进一步评估该方法的鲁棒性与适用范围。

## 参考文献

- [1] 郑佳艳, 毛若愚, 吴桐, 等. 基于欧拉放大和灰度均值差法的桥梁动态响应研究[J]. 公路工程, 2021, 46(3): 45-53+187.
- [2] 蒋华, 龚德书. 基于近景摄影测量的建筑物结构动态位移监测系统[J]. 南通大学学报(自然科学版), 2010, 9(4): 19-23.
- [3] 舒小华, 王珺, 闫松. 基于计算机视觉的结构位移测量技术及应用进展综述[J]. 火箭推进, 2024, 50(6): 40-51.
- [4] Yoon, H., Elanwar, H., Choi, H., Golparvar-Fard, M. and Spencer, B.F. (2016) Target-Free Approach for Vision-Based Structural System Identification Using Consumer-Grade Cameras: Target-Free Vision-Based Structural System Identification. *Structural Control and Health Monitoring*, **23**, 1405-1416. <https://doi.org/10.1002/stc.1850>
- [5] Feng, D., Feng, M., Ozer, E. and Fukuda, Y. (2015) A Vision-Based Sensor for Noncontact Structural Displacement Measurement. *Sensors*, **15**, 16557-16575. <https://doi.org/10.3390/s150716557>
- [6] Won, J., Park, J.W., Park, K., Yoon, H. and Moon, D. (2019) Non-Target Structural Displacement Measurement Using Reference Frame-Based Deepflow. *Sensors*, **19**, Article 2992. <https://doi.org/10.3390/s19132992>
- [7] Dong, C.Z., Celik, O., Catbas, F.N., O'Brien, E.J. and Taylor, S. (2020) Structural Displacement Monitoring Using Deep Learning-Based Full Field Optical Flow Methods. *Structure and Infrastructure Engineering*, **16**, 51-71. <https://doi.org/10.1080/15732479.2019.1650078>
- [8] Xu, H., Zhang, J., Cai, J., Rezaatofghi, H. and Tao, D. (2022) GMFlow: Learning Optical Flow via Global Matching. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 8111-8120. <https://doi.org/10.1109/cvpr52688.2022.00795>
- [9] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017) Attention Is All You Need. *Advances in Neural Information Processing Systems*, **30**, 5998-6008.