

# 考虑相机运动补偿的视觉模态测试方法

王振宇<sup>1</sup>, 王路倩<sup>1</sup>, 王兴<sup>1,2</sup>

(1. 中山大学·深圳 航空航天学院, 广东 深圳 518107;

2. 深圳市智能微小卫星星座技术与应用重点实验室, 广东 深圳 518107)

**摘要:** 基于计算机视觉的结构模态测试方法因其非接触式的特点和全场测量的优势而备受关注,但此类方法对相机运动敏感,通常要求相机固定以保证测量精度。本文提出一种考虑相机运动补偿的视觉模态测试方法,采用基于特征匹配的单应变换对原抖动视频进行校正,补偿相机的六自由度运动以得到稳定视频;在结构表面设置“虚拟视觉传感器”,利用欧拉视角的稠密光流法估计结构的振动响应;通过随机子空间法识别结构的模态参数。为验证所提方法的有效性,通过手持智能手机拍摄的方式采集模拟抖动的视频,并开展悬臂梁的单频振动试验和模态辨识试验。将视觉模态测试结果与扫描式激光多普勒测振仪的测试结果进行比较,前五阶模态频率和阻尼比的平均误差分别低于 0.4% 和 11.5%,模态振型的 MAC 矩阵对角值高于 98%。试验结果表明,所提方法有效地消除了相机运动的干扰,并获得了良好的结构模态参数估计精度。

**关键词:** 模态测试; 相机运动; 光流法; 随机子空间法

**中图分类号:** O327; TP391.4 **文献标志码:** A **文章编号:** 1004-4523(2024)06-0976-10

**DOI:** 10.16385/j.cnki.issn.1004-4523.2024.06.008

## 引言

振动试验是飞行器及其部/组件从理论研究走向工程实际的必要考核环节。为了获取结构的振动响应,通常在结构的表面粘贴“接触式”传感器,例如加速度计、速度计、应变片等。此类传感器种类丰富,技术较为成熟。近年来,随着航空航天领域轻质薄壁结构的大量使用,“接触式”传感器带来的附加刚度和附加质量对结构动力学性质的影响愈发显著。在此类工程应用中,“非接触式”传感器由于不会改变结构的动力学特性而更有优势,其中较为成熟的产品有扫描式激光多普勒测振仪<sup>[1]</sup>(Scanning Laser Doppler Vibrometer, SLDV)和数字图像相关(Digital Image Correlation, DIC)测量系统<sup>[2]</sup>等。然而,这些大型精密设备在模态测试之前需要较为复杂的试件表面处理及仪器标定过程,较为耗时且成本较高。

数字摄像机与光流算法的结合为基于单目视觉的振动检测带来了全新的方案——“视觉模态测试方法”<sup>[2]</sup>,其原理是利用结构表面光流变化获得结构的振动响应,再从表面运动响应中辨识结构的模态参数。这类“非接触式”振动测试方法不仅能够实现

测量的高空间分辨率、高频率分辨率,而且具有成本低,布署与使用灵活的优势<sup>[3-6]</sup>,其运动检测精度可达到亚像素级别<sup>[7-8]</sup>。Dong 等<sup>[9]</sup>使用基于深度学习的全场光流对真实看台进行了位移测量;Lu 等<sup>[10]</sup>提出了一种结合光流算法与模态分析的视觉模态分析技术,对四层钢框架进行了工作模态分析;王路倩等<sup>[11]</sup>使用稠密光流与运动放大算法对悬臂梁结构进行了全场工作模态分析。光流测振方法无需对结构的表面进行特殊处理,展现出极广的应用前景。然而,当前大多数文献中假设相机固定于惯性坐标系。在实际视觉模态分析中,部分场景下相机运动不可避免,例如乘坐交通工具时、地震灾害后等受限情况下手持拍摄的视频。在传统的图像和视频分析领域,对于相机运动的研究通常以平滑视频抖动、消除模糊效应为主要目标<sup>[12]</sup>,而在视觉振动检测中则需要以提取和恢复视频中结构的振动信息为目标。

相机运动对视觉振动检测带来的干扰按照运动分析视角的不同可以分成两类方法:拉格朗日方法和欧拉方法。拉格朗日方法直接跟踪视频内目标物的特征点,提取的特征点运动是目标的真实运动(在惯性系中的运动)与相机运动的叠加。因而,采用该方法进行测振首先需基于特征检测、模板匹配或信号分析对所测原始振动数据进行处理,再通过分离

**收稿日期:** 2022-06-30; **修订日期:** 2022-10-10

**基金项目:** 国家自然科学基金资助项目(52005522, 12072378); 深圳市科技计划资助项目(RCYX20210706092137055, ZDSYS20210623091808026)。

相机的运动信息获得目标的真实运动。辛宜瞳等<sup>[13]</sup>通过检测视频中距离目标物较近的不动点来近似计算相机的运动矩阵;Zhang等<sup>[14]</sup>采用了基于经验模态分解的方法消除无人机运动的干扰;Shao等<sup>[15]</sup>采用立体匹配和基于斑点的相对运动补偿方法解决了立体DIC中的相机运动问题;Chen等<sup>[16]</sup>研究了无人机桥梁检测中相机运动对DIC造成的干扰,通过检测DIC的特征点计算单应变换(Homography Transformation)矩阵对抖动视频进行校正。欧拉方法则聚焦于图像固定位置上的物体运动,无需复杂的目标特征匹配,计算效率较高。然而,相机的运动会使得图像目标位置观察到的物体发生变化,从而产生虚假的运动估计。为了抑制这一影响,欧拉方法一般首先采用视频稳定算法对抖动视频进行校正(通常是利用某种平面变换),再基于稳定视频进行振动检测。Jana等<sup>[17]</sup>研究了基于仿射变换(Affine Transformation)的相机运动抑制算法,并成功从手持抖动视频中估计了悬索桥的实时索力。然而,该算法中六参数的仿射变换模型只能抑制相机平动带来的图像扭曲和变形,尚缺乏对考虑相机多自由度空间运动的欧拉方法的研究。

此外,通过实际结构的动力学试验验证上述相机运动抑制算法的研究极为缺乏。Shao等<sup>[15]</sup>使用仿真DIC图像验证了静态位移对测量结果的影响;Chen等<sup>[16]</sup>验证了相机运动对结构单频振动检测结果的干扰;Jana等<sup>[17]</sup>验证了视频抖动导致的桥梁拉索前三阶模态频率辨识的误差,以及对实时索力检测能力的影响。针对复杂结构的模态试验,从抖动的视频中辨识出模态参数是极为严峻的挑战:一方面,模态试验中往往给被测试件施加宽频随机激励,此时结构振动响应中包含多个模态的分量。除此之外,与单频振动测试相比,各阶模态随机响应信号的幅值小、信噪比低,更容易受相机运动的影响。另一方面,模态试验中期望得到结构更多的参数,例如各阶模态的频率、阻尼比和振型。此类复杂试验尚未

见公开的文献报道。

本文提出了一种考虑相机运动补偿的视觉模态测试方法,并针对手持拍摄需求下的限制场景进行试验分析。首先采用单应变换抑制相机运动带来的视频抖动,再利用欧拉视角的稠密光流法从稳定的视频中提取结构的振动信息,最后使用协方差驱动的随机子空间法(Covariance-driven Stochastic Subspace Identification, Cov-SSI)辨识结构的多阶模态参数。与通常的仿射变换相比,本文采用的八参数单应变换更为全面地补偿了相机在三维空间中的运动<sup>[18]</sup>。为了验证所提方法的有效性,利用手持智能手机拍摄的方式获得含相机运动干扰的视频,分别开展了悬臂梁的单频振动试验和模态辨识试验。通过与扫描式激光多普勒测振仪的测试结果进行比较,验证所提方法对抑制相机运动干扰的有效性,并检验模态参数估计的精度。

## 1 视觉模态测试方法

考虑相机运动的视觉模态测试方法的步骤如图1所示,首先利用单应变换对抖动视频进行校正得到稳定视频,再通过欧拉视角的光流算法从稳定视频中获取结构的时域振动响应,最后基于Cov-SSI算法从结构响应数据中辨识得到结构的工作模态参数。

### 1.1 相机运动补偿原理

相机在三维空间中存在六自由度的刚体运动,在笛卡尔坐标系下可描述为三个方向的平移和旋转。如图2所示,相机运动对应成像平面的运动,反映为视频图像序列的抖动。若将视频首帧作为参考图像,相机运动将导致视频后续图像帧相对参考图像帧的扭曲与变形。若要消除这些影响,需借助平面变换来补偿相机的运动。如图3所示,常见的平面变换包括四参数的相似变换,六参数的仿射

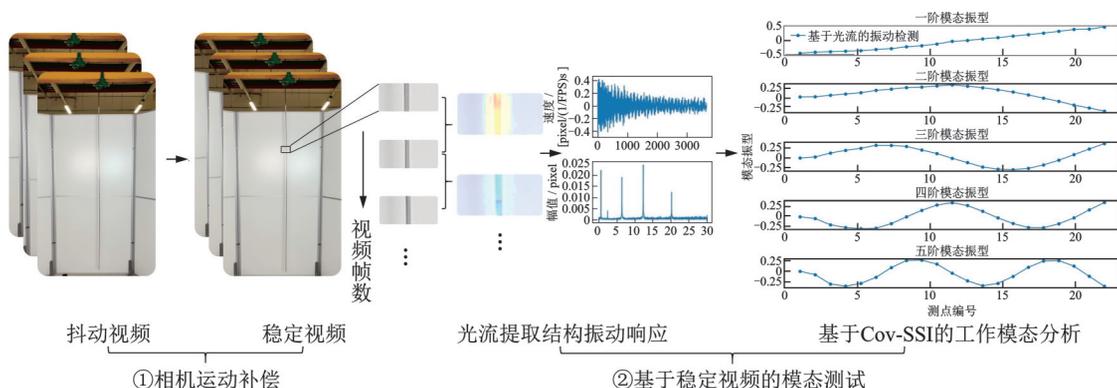


图1 考虑相机运动补偿的视觉模态测试框架

Fig. 1 Vision-based modal analysis framework in presence of camera motion compensation

变换和八参数的单应变换。其中,单应变换考虑了相机六自由度运动带来的平面变形<sup>[18]</sup>。此外,由于相机运动造成的视频抖动是全局的,即视频某帧图像不同部分之间的运动是相关的,故序列图像间存在着同样的对象和特征。

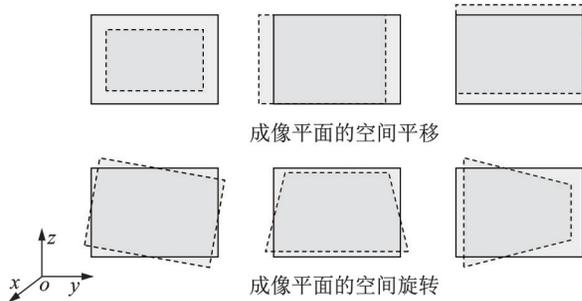


图 2 成像平面的空间运动<sup>[16]</sup>  
Fig. 2 The spatial motion of the image plane<sup>[16]</sup>

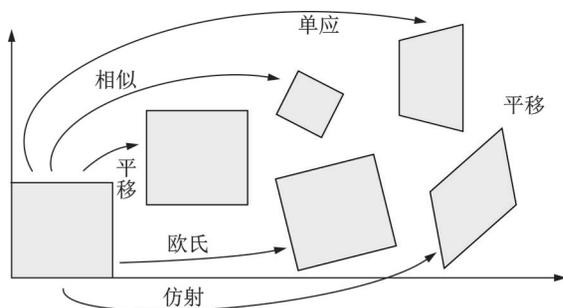
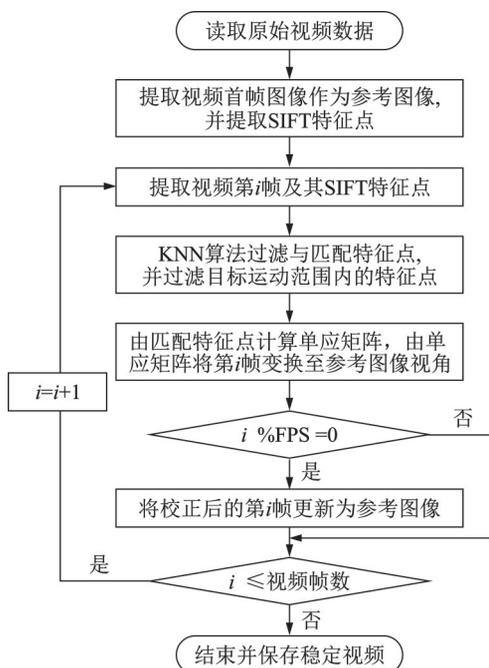


图 3 基本平面变换<sup>[18]</sup>  
Fig. 3 Basic plane transformation<sup>[18]</sup>

基于上述特点,本文采用基于特征点匹配与单应变换的相机运动补偿方法。如图 4 所示,首先将



注: %表示取余数。

图 4 相机运动补偿算法流程

Fig. 4 Camera motion compensation algorithm process

视频首帧作为参考图像并提取特征点,再依次提取后续图像帧的特征点,与参考图像的特征点进行匹配后计算图像间的单应变换矩阵。随后基于单应变换矩阵将后续图像变换至参考图像的视角,即可获得稳定的视频。考虑到在短时间内相机的运动是较小的,根据视频的拍摄帧率(Frames Per Second, FPS)更新参考图像,减少校正图像带来的误差积累。

### 1.1.1 特征点提取与匹配

特征点一般为图像中的边缘、角点和纹理变化明显的区域,由关键点(描述了特征点的位置)和描述符(描述了特征点的特征)两个量来确定。本文采用尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)算法进行特征点检测,SIFT算法综合考虑了目标区域的梯度方向,强度(灰度值)以及特征尺度信息来描述目标,其特征点描述符具有在图像缩放和旋转下保持不变的性质<sup>[19]</sup>。

检测得到参考图像帧和当前图像帧中的特征点集后,因目标结构存在柔性变形,首先过滤结构运动范围内的特征点。良好的特征点应为惯性系中的固定点,例如试验过程中的吊架、固支于地面的结构等。然后使用K近邻算法(K-Nearest Neighbors, KNN)对当前图像与参考图像中的特征点进行匹配,K近邻算法通过比较特征点描述符向量在高维特征空间中的距离对其进行匹配<sup>[20-21]</sup>。SIFT特征点的描述符为长度128的向量,设特征空间 $\chi$ 是128维实数向量空间 $\mathbb{R}^{128}$ ,对参考图像和当前图像中的任意特征点描述符向量 $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(128)})^T$ ,  $x_j = (x_j^{(1)}, x_j^{(2)}, \dots, x_j^{(128)})^T \in \chi$ ,使用欧氏距离( $L_2$ 范数)度量K近邻:

$$L_2(x_i, x_j) = \left( \sum_{l=1}^{128} |x_i^{(l)} - x_j^{(l)}|^2 \right)^{\frac{1}{2}} \quad (1)$$

根据欧氏距离度量,从参考图像的特征点集中找出与当前图像中任意特征点 $x_j$ 最邻近的两个特征点(最近邻和第二近邻),依据最近邻与第二近邻的距离比 $R$ 进行高效和正确的匹配。筛选 $R < 0.8$ 的匹配项时能够消除90%的错误匹配项,同时丢弃少于5%的正确匹配项。因为当最近邻是正确匹配时,最近邻要比第二近邻更为接近;而当最近邻为错误匹配时,由于特征空间的高维性,在相近的距离内可能存在许多错误匹配<sup>[19]</sup>。

图5展示了抖动视频中典型的两帧图像所匹配到的特征点,可以发现特征点间的连线反映了图像各个部分相似的运动趋势。

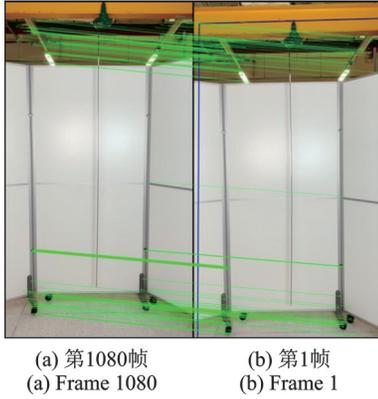


图5 两帧图像间匹配的特征点

Fig. 5 Matched feature points between two frames

### 1.1.2 单应矩阵的计算

从当前图像和参考图像中匹配到足够数量的特征点对后,如图6所示,即可根据匹配点对 $x_i(x_i, y_i), x'_i(x'_i, y'_i)$ 的坐标计算图像(成像平面)之间的单应变换矩阵 $H^{[22]}$ :

$$x' = Hx \quad (2)$$

其中:

$$H = \begin{pmatrix} A & t \\ v^T & s \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ v_1 & v_2 & s \end{pmatrix}, x = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, x' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}.$$

式中 矩阵 $A$ 反映了尺度、旋转和剪切等仿射变换参数;参数 $t$ 反映了平移变换参数;参数 $v$ 反映了单应变换的非线性效应;参数 $s$ 是与 $v$ 相关的缩放因子,通常归一化使 $s$ 为1。故该变换矩阵 $H$ 由8个参数确定。

考虑一般情况,任意一组匹配点 $x_i(x_i, y_i), x'_i(x'_i, y'_i)$ 间有下式成立:

$$\begin{pmatrix} x'_i \\ y'_i \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \quad (3)$$

得到:

$$\begin{cases} x'_i = \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}} \\ y'_i = \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}} \end{cases} \quad (4)$$

式中  $h_i$ 为变换矩阵的未知系数。

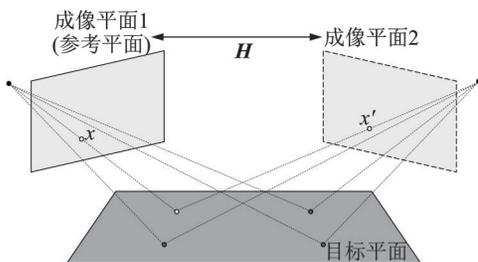


图6 成像平面间的单应变换

Fig. 6 Homography transformation between image planes

计算当前图像平面与参考图像平面之间的变换矩阵 $H$ ,使得反向投影错误率达到最小,即

$$\min \sum_i \left[ (x'_i - a)^2 + (y'_i - b)^2 \right] \quad (5)$$

式中  $a = \frac{h_{11}x_i + h_{12}y_i + h_{13}}{h_{31}x_i + h_{32}y_i + h_{33}}, b = \frac{h_{21}x_i + h_{22}y_i + h_{23}}{h_{31}x_i + h_{32}y_i + h_{33}}$ 。

使用随机抽样一致算法(Random Sample Consensus, RANSAC)<sup>[23]</sup>对式(5)进行单应变换矩阵的参数估计,避免错误匹配点对的干扰。得到后续图像相对参考图像的单应变换矩阵 $H$ 后,即可通过反向变换将后续图像变换至参考图像的视角,从而将抖动视频变换至稳定视频。

## 1.2 基于稳定视频的模态参数辨识

将抖动视频校正为稳定视频后,第二步使用基于欧拉视角的稠密光流提取目标结构的振动响应,并辨识结构的模态参数。

### 1.2.1 运动提取

视频数据是按时序顺序捕获的图像帧,每帧图像为离散的 $m \times n$ 的二维矩阵,可以认为是投影平面上关于灰度值的二维函数 $I(x, y)$ 。图像的灰度值受到对象表面特征和光照等因素的影响,当对象移动时,图像序列的灰度值也将发生相应变化,因此某一像素的灰度值也是时间的函数,表示为 $I(x, y, t)$ 。光流将图像中像素灰度值的变化与像素的运动联系起来,反映了物体的三维运动场在二维成像平面上的投影,即物体的表观运动<sup>[24]</sup>。

结构动力学测试场景通常能够满足光流法的亮度恒定和小运动假设。欧拉视角的光流法在图像上选择固定区域、固定大小的“虚拟视觉传感器”,使用Gunnar Farneback稠密光流法获取每个传感器所测结构的表观运动,该方法仅利用相邻帧进行帧间的运动估计,对连续帧之间快速变化且难以预测的位移具有较好的适用性。其基本思想是用多项式展开拟合每个像素的某个邻域。以二次多项式为为例,某个像素 $x=(x, y)$ 及其邻域在某时刻 $t_1$ 的多项式拟合可表示为<sup>[25]</sup>:

$$f_1(x) = x^T A_1 x + b_1^T x + c_1 \quad (6)$$

根据像素及其邻域的灰度值以加权最小二乘法估计式(6)的拟合系数。像素及其邻域经过全局运动 $d$ 后在下一时刻 $t_2$ 可表示为:

$$\begin{aligned} f_2(x) &= f_1(x - d) = \\ &= (x - d)^T A_1 (x - d) + b_1^T (x - d) + c_1 = \\ &= x^T A_1 x + (b_1 - 2A_1 d)^T x + \\ &= d^T A_1 d - b_1^T d + c_1 \end{aligned} \quad (7)$$

同时可表示为:

$$f_2(x) = x^T A_2 x + b_2^T x + c_2 \quad (8)$$

由式(7)和(8)对应系数相等可求得像素及其邻域的运动:

$$d = -\frac{1}{2} A_1^{-1} (b_2 - b_1) \quad (9)$$

更详细的叙述参见文献[25]。受视频拍摄噪声、单应变换误差和光流检测算法精度的影响,直接采用欧拉视角的光流算法估计的结构振动响应存在一些异常值。首先对原始信号进行时域滤波来检测和替换异常值,计算每个采样值偏离中位数绝对值的标准差,如果采样值与中位数相差三个以上标准差,则该采样值将被中位数替换<sup>[26]</sup>。

### 1.2.2 工作模态分析

在视觉模态测试中,通过光流算法估计得到的结构运动响应是离散化的,考虑环境噪声干扰得到系统离散时间状态空间模型<sup>[27-29]</sup>:

$$\begin{cases} x_{k+1} = Sx_k + w_k \\ y_k = Cx_k + v_k \end{cases} \quad (10)$$

式中  $w_k$  为环境噪声干扰;  $v_k$  为“虚拟”传感器测量误差带来的噪声干扰,此处包括视频拍摄、单应变换、光流检测等导致的测量误差;  $x_k$  为第  $k$  采样时刻的系统状态向量;  $y_k$  表示第  $k$  采样时刻的输出向量,为“虚拟”传感器观测所得到的结构响应数据;  $S \in \mathbb{R}^{2N \times 2N}$  表示系统状态矩阵;  $C \in \mathbb{R}^{M \times 2N}$  表示系统输出矩阵;  $M$  为测点数量;  $N$  为系统阶数。

通过基于稠密光流的运动估计得到结构上  $M$  个“虚拟视觉传感器”所测时域速度响应,将  $N$  个采样点(取决于视频帧率及拍摄时长)、 $M$  个测点的时域数据矩阵  $Y$  作为 Cov-SSI 的输入辨识结构模态参数:

$$Y = (y_1 \ y_2 \ \cdots \ y_N) \quad (11)$$

由振动理论可知,基于 Cov-SSI 求得系统状态矩阵  $S$  和系统输出矩阵  $C$  后,即可求得系统的模态参数。对系统状态矩阵  $S$  进行特征值分解:

$$S = \Psi \Lambda \Psi^{-1} \quad (12)$$

式中 对角矩阵  $\Lambda = \text{diag}(\mu_r)$ , 其中  $\mu_r$  为矩阵  $\Lambda$  的各阶特征值。

各阶模态频率、阻尼比与模态振型如下式所示:

$$\omega_r = \sqrt{\text{Re}^2\left(\frac{\ln \mu_r}{\Delta t}\right) + \text{Im}^2\left(\frac{\ln \mu_r}{\Delta t}\right)} \quad (13)$$

$$\xi_r = -\frac{\text{Re}\left(\frac{\ln \mu_r}{\Delta t}\right)}{\omega_r} \quad (14)$$

$$\Phi = C\Psi \quad (15)$$

式中  $\Delta t$  表示采样时间间隔。

## 2 试验验证

为验证本文方法的有效性与精确度,开展了悬

臂梁的单频振动测量试验和工作模态分析试验,对比了相机运动干扰下直接视觉方法<sup>[11]</sup>(直接对抖动视频进行欧拉视角的振动检测与参数辨识)和补偿视觉方法(补偿相机运动后对稳定视频进行振动检测与参数辨识)的测试结果,并将 SLDV 的测试结果作为参考基准进行了对比分析。

被测悬臂梁结构为  $(17.8 \times 39.2 \times 2016) \text{ mm}^3$  的钢尺,如图 7 所示,钢尺顶部使用台虎钳固定于龙门架上,固接位置距离钢尺底部为 2000 mm。



(a) 视觉模态测试试验环境

(a) Experimental environment of vision-based modal analysis



(b) SLDV 试验环境

(b) Experimental environment of SLDV

图 7 试验环境

Fig. 7 Experimental environment

考虑相机运动补偿的视觉模态测试环境如图 7(a) 所示,使用无频闪光源进行照明以避免室内频闪灯的干扰,使用智能手机作为视频采集设备,采用手持拍摄的方式获取受相机运动干扰的抖动视频。经有限元仿真与理论计算,该悬臂梁结构前五阶频率在 30 Hz 内。根据香农-奈奎斯特采样定理,为避免时域混叠,视频拍摄帧率需大于结构模态频率的 2 倍,设置视频拍摄帧率为 60 FPS,分辨率为  $1920 \times 1080 \text{ pixel}$ ,录制时长为 60 s。

图 7(b) 所示为采用 SLDV 作为测量设备的试验场景,采用了德国 Polytec 公司生产的 PSV-500 全场扫描式激光多普勒测振仪,其速度分辨率优于  $0.02 \mu\text{m/s/1 Hz}$ ,扫描角度范围为  $40^\circ \times 50^\circ$ 。激光扫描头距离结构约 5 m,高度约为 1.3 m,与悬臂梁中部对齐。悬臂梁从底端开始间距 100 mm 粘贴 21 个反光片作为 SLDV 的测点。

如图 8 所示,视觉模态测试过程中,在被测结构

上设置了与SLDV反光片相同位置的21个“虚拟视觉传感器”,传感器的大小为 $80 \times 30$  pixel。取每个传感器水平方向的像素光流均值作为被测结构对应位置的动态响应。

如图9所示,通过单应变换将第1080帧校正至第1帧视角,可以看到在 $1920 \times 1080$ 分辨率下变换前手持拍摄带来的帧间图像变形幅度在几十像素量级,变换后帧间图像的扭曲与变形基本被消除,变形幅度在像素级。

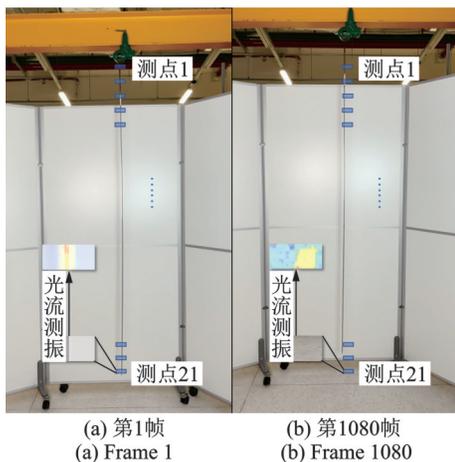


图8 “虚拟视觉传感器”及局部稠密光流

Fig. 8 “Virtual vision sensors” and local dense optical flow

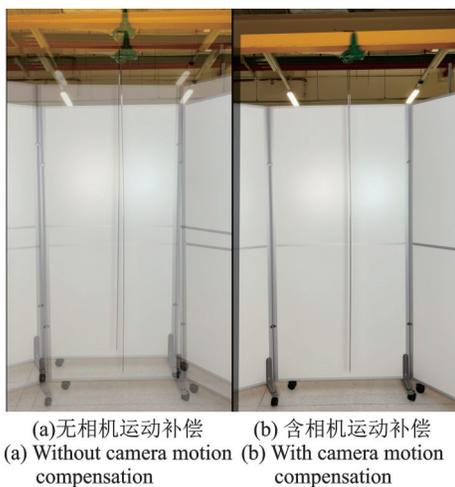


图9 第1帧与第1080帧的叠加图

Fig. 9 Overlay of frame 1 and frame 1080

### 2.1 单频振动测量试验

使用信号发生器(Tektronix, AFG2021)和激振器(东华测试, DH40020)在悬臂梁固接端根部施加定频为10 Hz的正弦激励。

图10显示了21号测点的时频域数据,并对比了直接视觉方法与补偿视觉方法的振动检测结果。可以看出,由于手持拍摄导致的视频微幅和低频抖动,直接视觉方法存在较大的低频干扰,同时其估计得

到的测点运动幅值与补偿视觉方法存在较大差异。这是因为直接视觉方法选择了固定区域和大小的“虚拟视觉传感器”,由于相机运动的干扰,被测结构在水平方向可能会移动至传感器以外,如图9(b)所示。

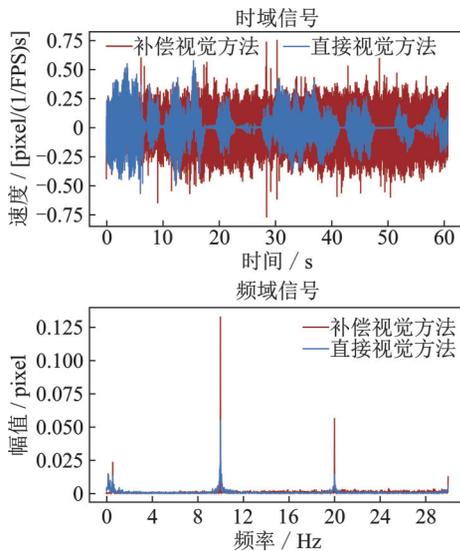


图10 补偿视觉方法与直接视觉方法<sup>[11]</sup>估计测点21的时频信号

Fig. 10 Time and frequency signals of measuring point 21 estimated using the compensation vision method and direct vision method<sup>[11]</sup>

表1对比了直接视觉方法、补偿视觉方法和SLDV测试方法对梁的振动频率和幅值的估计结果。从频率估计结果来看,直接视觉方法的检测结果为10.003 Hz,补偿视觉方法的结果为10.005 Hz(频域分辨率为0.0083 Hz),二者的频率测量精度都在理想范围内。这反映了相机的低频运动并不会显著影响结构单频运动下的频率检测精度。从幅值估计结果来看,直接视觉方法所得21个测点的幅值与SLDV所测结果的均方根误差(Root Mean Square Error, RMSE)达20.81%,而补偿视觉方法对应的RMSE降至12.12%,幅值误差降低了41.76%,取得了较大的改善。这一改善从各测点的幅值对比中(图11)可以更加清晰地观察到。

表1 检测精度:直接视觉方法<sup>[11]</sup>、补偿视觉方法与SLDV测试方法对比

Tab. 1 Detection accuracy: direct vision method<sup>[11]</sup> and compensation vision method compared to SLDV measurement

测试方法	21个测点平均频率/Hz	误差/%	21个测点幅值RMSE%
直接视觉方法	10.003	0.03	20.81
补偿视觉方法	10.005	0.05	12.12

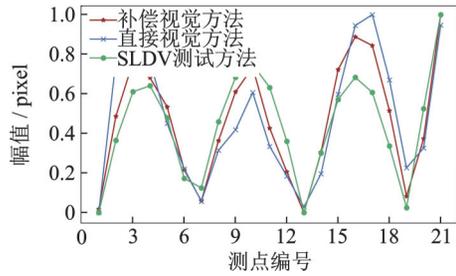


图 11 补偿视觉方法、直接视觉方法<sup>[11]</sup>与 SLDV 测试方法对梁振动的幅值估计

Fig. 11 Amplitude estimation of compensation vision method, direct vision method<sup>[11]</sup> and SLDV measurement

2.2 悬臂梁工作模态分析试验

给予悬臂梁根部附近初始力激励,直接视觉方法与补偿视觉方法所测结构响应的时频数据如图 12 所示。同样的,由于手持拍摄导致的相机微幅、低频的六自由度空间运动,直接视觉方法所测振动响应存在大量低频噪声,对识别结构的低阶模态频率造成一定干扰,同时各阶模态频率的幅值也存在较大差异。表 2 和 3 对比了直接视觉方法、补偿视觉方法与 SLDV 测试方法得到的模态频率和阻尼比。以 SLDV 测试方法得到的结果为基准,直接视觉方法估计的悬臂梁前五阶模态频率的平均误差为

6.59%,其中主要误差来源为第一阶模态,误差达 32.14%;而对高阶模态频率和阻尼比的估计相对准确,进一步说明手持拍摄带来的运动干扰对低阶模态的辨识影响更为显著;前五阶模态阻尼比的平均误差为 367%,误差较大。采用补偿视觉方法前五阶模态频率的平均误差仅为 0.39%,低阶模态频率误差极大的降低,而前五阶模态阻尼比的平均误差为 11.45%。由以上分析可知,利用相机运动补偿算法,悬臂梁视觉模态测试结果中模态频率误差降低了 94.08%,模态阻尼比误差降低了 96.88%。

图 13 为悬臂梁工作模态振型的分析结果,直接视觉方法的低阶振型识别结果相比 SLDV 的结果在一些测点(测点 1, 2, 3, 11, 12, 21)存在较大突变。如前文所述,当相机存在运动时被测结构可能移动至“虚拟视觉传感器”以外。而补偿视觉方法与 SLDV 的振型分析结果具有极佳的一致性。采用模态置信准则(Modal Assurance Criteria, MAC)评价直接视觉方法与补偿视觉方法所测振型与 SLDV 所测振型的相似度。如图 14(b)所示,补偿视觉方法与 SLDV 的前五阶 MAC 矩阵对角线元素分别为 99.19%, 99.39%, 98.81%, 98.64% 和 98.18%,非对角线元素趋近于零,均值仅为 1.10%。

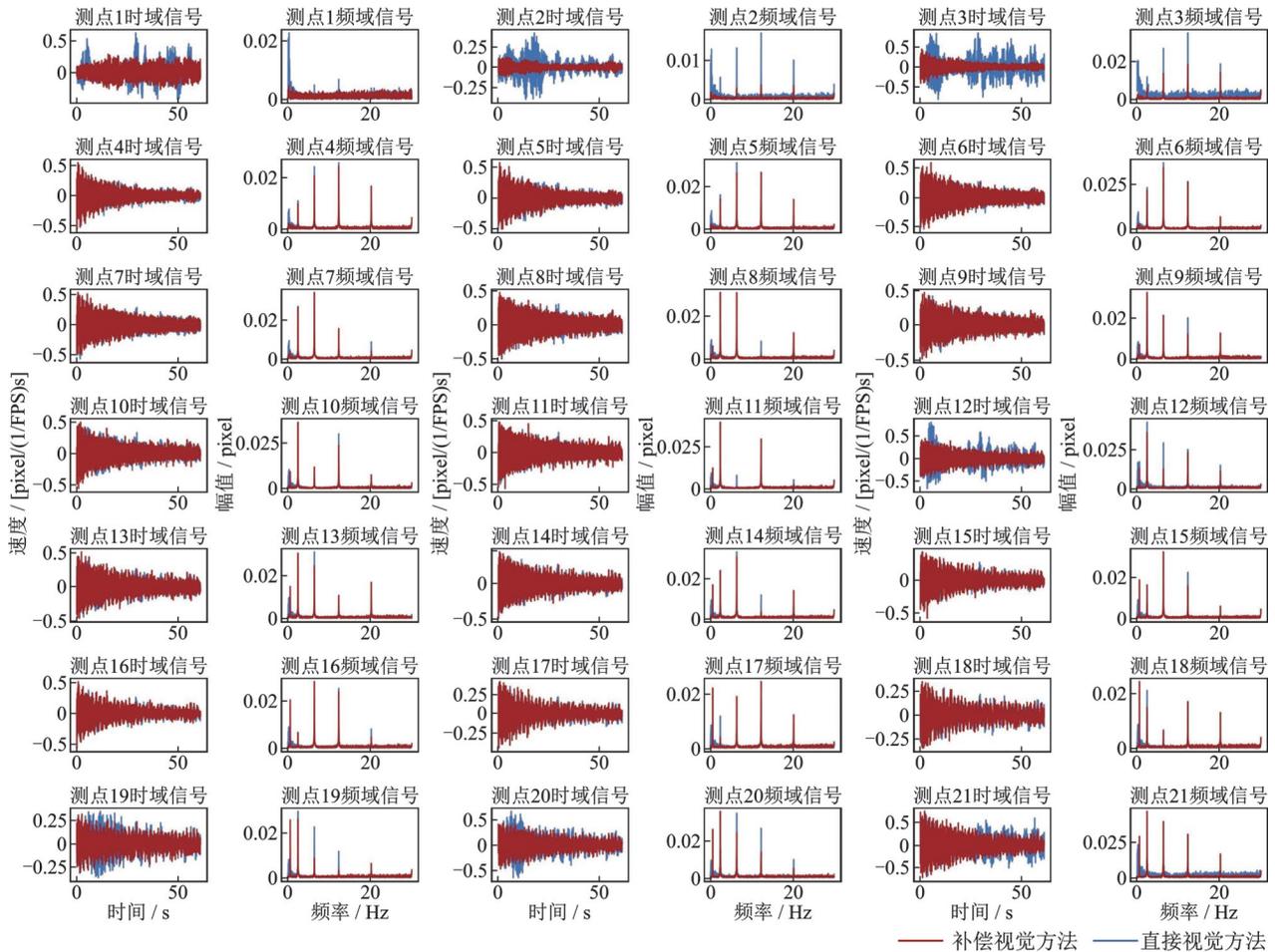


图 12 21 个测点的时频信号:补偿视觉方法与直接视觉方法<sup>[11]</sup>

Fig. 12 Time and frequency signals of 21 measure points: compensation vision method and direct vision method<sup>[11]</sup>

表 2 模态频率:直接视觉方法<sup>[11]</sup>、补偿视觉方法与 SLDV 测试方法对比

Tab.2 Modal frequency: direct vision method<sup>[11]</sup> and compensation vision method compared to SLDV measurement

模态参数	SLDV/ Hz	直接视 觉方法/ Hz	直接视 觉方法 误差/%	补偿视 觉方法/ Hz	补偿视 觉方法 误差/%
一阶频率	0.56	0.38	32.14	0.55	1.79
二阶频率	2.42	2.41	0.41	2.42	0.00
三阶频率	6.41	6.40	0.16	6.41	0.00
四阶频率	12.37	12.37	0.00	12.36	0.08
五阶频率	19.99	19.94	0.25	20.01	0.10
频率平均 误差			6.59		0.39

表 3 阻尼比:直接视觉方法<sup>[10]</sup>、补偿视觉方法与 SLDV 测试方法对比

Tab.3 Damping ratio: direct vision method<sup>[10]</sup> and compensation vision method compared to SLDV measurement

模态参数	SLDV/ %	直接视 觉方 法/%	直接视 觉方法 误差/%	补偿视 觉方 法/%	补偿视 觉方法 误差/%
一阶阻尼比	1.25	9.75	680.00	1.66	32.80
二阶阻尼比	0.22	2.44	1009.09	0.25	13.64
三阶阻尼比	0.52	0.96	84.62	0.51	1.92
四阶阻尼比	0.45	0.47	4.44	0.41	8.89
五阶阻尼比	1.02	1.60	56.86	1.02	0.00
阻尼比平均 误差			367.00		11.45

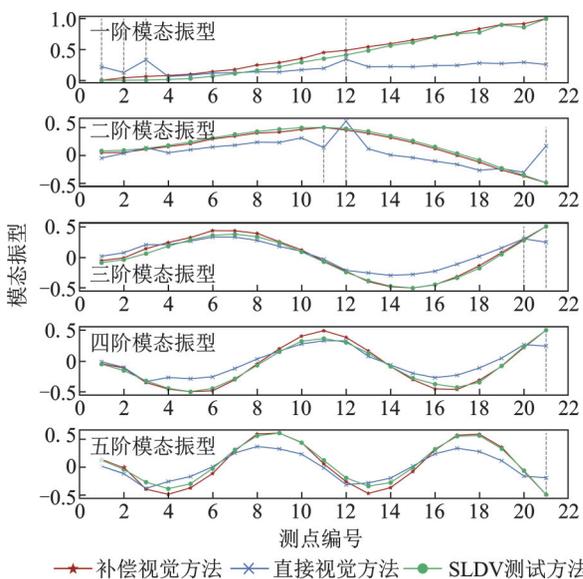
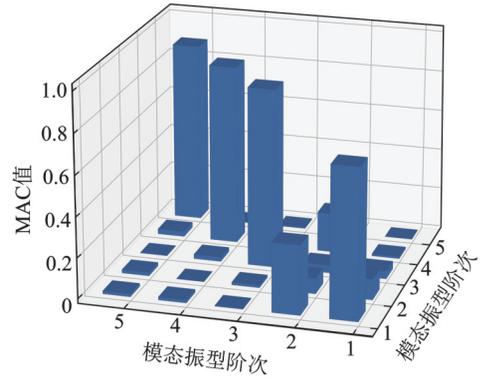
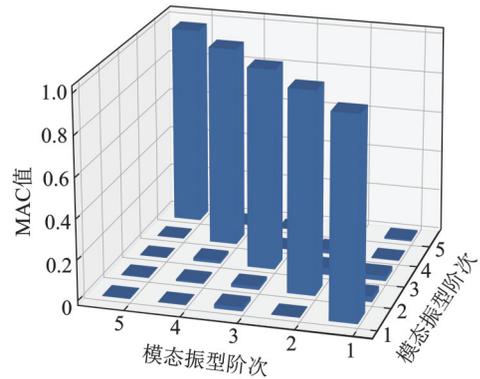


图 13 前五阶模态振型:补偿视觉方法、直接视觉方法<sup>[11]</sup>对比 SLDV 测试方法

Fig. 13 First five-order modal shapes: compensation vision method and direct vision method<sup>[11]</sup> compared to SLDV measurement



(a) 直接视觉方法<sup>[11]</sup>与 SLDV 测试方法  
(a) Direct vision method<sup>[11]</sup> and SLDV measurement



(b) 补偿视觉方法与 SLDV 测试方法  
(b) Compensation vision method and SLDV measurement

图 14 前五阶模态振型的 MAC 矩阵

Fig. 14 MAC matrix of first five-order modal shapes

### 3 结 论

本文提出了一种考虑相机运动补偿的视觉模态测试方法,该方法基于单应变换抑制相机运动带来的视频抖动,利用稠密光流法获取结构振动响应,通过随机子空间法辨识结构模态参数。依据手持拍摄条件下悬臂梁单频振动试验和模态辨识试验表明本文所提方法在类似限制场景下能够有效地抑制相机运动干扰。以 SLDV 测试方法获得的结果为基准,考虑相机运动补偿的视觉模态测试方法对悬臂梁的前五阶模态频率辨识误差低于 0.4%,模态阻尼比辨识误差低于 11.5%,MAC 矩阵对角值高于 98%。本方法有效抑制了相机运动对欧拉视角的视觉模态测试的干扰,较为精准地获得了结构的模态参数。

本文提出的方法具有成本低、布署灵活、无需在结构表面进行人工标记等优势,但仍在以下方面值得进一步研究:(1)视频校正后图像序列间的运动仍为像素级别,对亚像素级别的微弱振动的检测精度值得探究。(2)基于平面变换的相机运动补偿在某些复杂环境(如特征点过少且具有显著景深变化时)存在较大幅度的图像变形问题,如何减少该类校正误差需深入探究。

## 参考文献:

- [1] Chen D M, Xu Y F, Zhu W D. Identification of damage in plates using full-field measurement with a continuously scanning laser Doppler vibrometer system[J]. *Journal of Sound and Vibration*, 2018, 422: 542-567.
- [2] 王兴, 王路倩. 一种基于光流法的结构微幅振动工作模态分析方法: 202111410401.0[P]. 2022-03-15.
- [3] Baqersad J, Poozesh P, Niezrecki C, et al. Photogrammetry and optical methods in structural dynamics—a review[J]. *Mechanical Systems and Signal Processing*, 2017, 86(B): 17-34.
- [4] Bhowmick S, Nagarajaiah S, Lai Z. Measurement of full-field displacement time history of a vibrating continuous edge from video[J]. *Mechanical Systems and Signal Processing*, 2020, 144: 106847.
- [5] Luan L L, Zheng J W, Wang M L, et al. Extracting full-field subpixel structural displacements from videos via deep learning[J]. *Journal of Sound and Vibration*, 2021, 505: 116142.
- [6] Javed A, Lee H, Kim B, et al. Vibration measurement of a rotating cylindrical structure using subpixel-based edge detection and edge tracking[J]. *Mechanical Systems and Signal Processing*, 2022, 166: 108437.
- [7] Javh J, Slavič J, Boltežar M. The subpixel resolution of optical-flow-based modal analysis[J]. *Mechanical Systems and Signal Processing*, 2017, 88: 89-99.
- [8] Diamond D H, Heyns P S, Oberholster A J. Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence[J]. *Measurement*, 2017, 95: 166-172.
- [9] Dong C Z, Celik O, Catbas F N, et al. Structural displacement monitoring using deep learning-based full field optical flow methods[J]. *Structure and Infrastructure Engineering*, 2020, 16(1): 51-71.
- [10] Lu Z, Lin G, Wang L. Output-only modal parameter identification of structures by vision modal analysis[J]. *Journal of Sound and Vibration*, 2021, 497: 115949.
- [11] 王路倩, 王振宇, 王兴. 结构全场工作模态分析的计算机视觉测振方法[J]. *振动工程学报*, 2024, 37(2): 226-236.
- Wang L Q, Wang Z Y, Wang X. Full-field structural modal analysis based on visual vibrometry[J]. *Journal of Vibration Engineering*, 2024, 37(2): 226-236.
- [12] Guilluy W, Oudre L, Beghdadi A. Video stabilization: overview, challenges and perspectives[J]. *Signal Processing: Image Communication*, 2021, 90: 116015.
- [13] 辛宜瞳, 曹文霞, 陈鲸, 等. 抗相机运动干扰的欧拉视频振动检测[J]. *计算机工程与应用*, 2022, 58(17): 277-284.
- Xin Y T, Cao W X, Chen J, et al. Eulerian video vibration detection against camera motion interference[J]. *Computer Engineering and Applications*, 2022, 58(17): 277-284.
- [14] Zhang C, Tian Y, Zhang J. Complex image background segmentation for cable force estimation of urban bridges with drone-captured video and deep learning[J]. *Structural Control and Health Monitoring*, 2022, 29(4): e2910.
- [15] Shao X, He X. Camera motion-induced systematic errors in stereo-DIC and speckle-based compensation method[J]. *Optics and Lasers in Engineering*, 2022, 149: 106809.
- [16] Chen G, Liang Q, Zhong W, et al. Homography-based measurement of bridge vibration using UAV and DIC method[J]. *Measurement*, 2021, 170: 108683.
- [17] Jana D, Nagarajaiah S. Computer vision-based real-time cable tension estimation in Dubrovnik cable-stayed bridge using moving handheld video camera[J]. *Structural Control and Health Monitoring*, 2021, 28(5): e2713.
- [18] Szeliski Richard. 计算机视觉——算法与应用[M]. 艾海舟, 兴军亮, 等译. 北京: 清华大学出版社, 2012.
- [19] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [20] 李航. 统计学习方法[M]. 2版. 北京: 清华大学出版社, 2019.
- [21] Cover T, Hart P. Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21-27.
- [22] Hartley Richard, Zisserman Andrew. 计算机视觉中的多视图几何[M]. 韦穗, 杨尚骏, 章权兵, 等译. 合肥: 安徽大学出版社, 2002.
- [23] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [24] 伯特霍尔德·霍恩. 机器视觉[M]. 王亮, 蒋欣兰, 译. 北京: 中国青年出版社, 2014.
- [25] Farneback G. Two-Frame Motion Estimation Based on Polynomial Expansion[M]. Berlin, Heidelberg: Springer, 2003: 363-370.
- [26] Liu H, Shah S, Jiang W. On-line outlier detection and data cleaning[J]. *Computers and Chemical Engineering*, 2004, 28(9): 1635-1647.
- [27] Brincker R, Andersen P. Understanding stochastic subspace identification[C]//Conference Proceedings: IMAC-XXIV: A Conference & Exposition on Structural Dynamics. St. Louis, Missouri, US: Society for Experimental Mechanics, 2006.

- [28] Van Overschee P, De Moor B L. Subspace Identification for Linear Systems: Theory-Implementation-Applications[M]. New York: Springer, 1996.
- [29] Zabel V. Operational modal analysis-theory and aspects of application in civil engineering[D]. Weimar: Bauhaus-Universität Weimar, 2019.

## Vision-based structural modal analysis in presence of camera motions

WANG Zhen-yu<sup>1</sup>, WANG Lu-qian<sup>1</sup>, WANG Xing<sup>1,2</sup>

(1.School of Aeronautics and Astronautics, Shenzhen Campus of Sun Yat-sen University, Shenzhen 518107, China;

2.Shenzhen Key Laboratory of Intelligent Microsatellite Constellation, Shenzhen 518107, China)

**Abstract:** Vision-based modal analysis techniques have gained attention due to their non-contact, full-field measurement capabilities, making them particularly suitable for the dynamic testing of large-scale or thin-walled structures. However, these techniques often require cameras to be fixed to the ground to avoid coupling with the vibrations of the test structure, a requirement that can be too restrictive in real-world applications. This paper proposes a method to compensate for camera motion using homography transformation, followed by the extraction of the test structure's movement by applying the dense optical flow method to the stabilized video. The procedure involves transforming the video captured by a moving camera using feature matching algorithms, where a homography matrix compensates for six degrees of camera motions. Several "virtual vision sensors" are selected on the edges of the structure, and their vibrations are estimated using optical flow methods. Structural modal parameters are then extracted from the output-only data using stochastic subspace identification algorithms. The proposed procedure was applied to videos recorded using a moving smartphone to conduct an operating modal analysis of a 2 m cantilevered beam. To validate the procedure, the vision-based analysis results were compared with measurements taken with a Scanning Laser Doppler Vibrometer. The results show an average discrepancy of 0.4% and 11.5% for the first five natural frequencies and damping ratios of the beam, respectively. The mode shapes also show strong correlation between the two measurement techniques, as indicated by the diagonal MAC values greater than 98%. Therefore, the proposed procedure effectively cancels out camera motions and achieves accurate estimation of structural modal parameters.

**Key words:** modal analysis; camera motion; optical flow method; stochastic subspace identification

**作者简介:** 王振宇(1997—),男,硕士研究生。E-mail: wangzhy237@mail2.sysu.edu.cn。

**通讯作者:** 王 兴(1989—),男,博士,副教授。E-mail: wangxing5@mail.sysu.edu.cn。