buildings

MDPI

Review

Vision-Based Displacement Measurement for Structural Health Monitoring: A Metrology-Oriented Review of Uncertainty Quantification

Arman Neyestani [1,*], Francesco Picariello [2], Ioan Tudosa [3], Michela Monaco [3], Luca De Vito [3] and Mauro D'Arco [1]

[1] Department of Information Technology and Electrical Engineering, University of Naples Federico II, 80138 Naples, Italy; mauro.darco@unina.it
[2] Department of Engineering and Science, University Mercatorum, 00186 Rome, Italy; francesco.picariello@unimercatorum.it
[3] Department of Engineering, University of Sannio, 82100 Benevento, Italy; itudosa@unisannio.it (I.T.); monaco@unisannio.it (M.M.); devito@unisannio.it (L.D.V.)
* Correspondence: neyestani@ieee.org

Abstract

This paper presents a metrology-oriented review of vision-based displacement and deformation measurement for civil structural health monitoring (SHM), with an emphasis on field robustness and uncertainty quantification (UQ). The review focuses on image- and video-based methods that convert visual information into quantitative physical measurements, such as displacement, strain, or derived dynamic indicators. The literature is organized according to the main stages of the measurement chain: image formation, image-plane motion estimation, and geometric conversion to metric motion. Within this framework, measurement pipelines are interpreted through three levels of geometric mapping, namely, a scalar scale-factor model, a planar homography-based model, and a full Jacobian-based model. The review synthesizes major method families, including marker-based and markerless tracking, feature-based tracking, optical flow, digital image correlation (DIC), phase-based motion magnification, edge-based estimators, fixed- and moving-camera configurations, UAV-based acquisition with ego-motion compensation, hybrid vision–sensor fusion, and deep-learning-enhanced pipelines. A structured taxonomy of uncertainty sources is then presented along the processing chain, covering camera geometry and calibration, imaging noise and blur, quantization, timing and synchronization, environmental disturbances, optical turbulence and heat haze, platform motion, algorithmic failure modes, and reference-sensor uncertainty. The paper also compares UQ practices, including GUM-aligned analytical propagation, Monte Carlo methods, DIC-specific error budgets, bootstrap and resampling strategies, and probabilistic deep learning. The main contribution of this review is to connect computer-vision-based displacement pipelines with metrological requirements by explicitly linking measurement models, uncertainty sources, UQ methods, and field-validation evidence within a unified framework. A practical uncertainty-budget template is compiled to support traceable reporting across different pipelines and deployment scenarios. The paper concludes with prioritized research gaps and future directions, including standardized benchmarks and datasets, traceable UQ for moving-camera systems, multi-sensor fusion with end-to-end uncertainty propagation, long-term drift characterization, optical-turbulence and adverse-weather modeling, validated subpixel limits at extreme range, probabilistic deep learning–metrology integration, and standardized reporting practices.

check for updates

Academic Editors: Paweł Karol Frankowski and Marcin Mąka

Received: 5 May 2026
Revised: 19 June 2026
Accepted: 20 June 2026
Published: 4 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: vision-based displacement measurement; structural health monitoring (SHM); measurement uncertainty; uncertainty quantification

1. Introduction

Civil infrastructure systems such as bridges, viaducts, and high-rise buildings increasingly rely on structural health monitoring (SHM) to support performance assessment, life-cycle asset management, and resilience-oriented design. Over the past decade, vision-based sensing has emerged primarily as a compelling complement to conventional displacement-monitoring instrumentation, and in selected scenarios as a practical alternative when contact sensors or dense pointwise instrumentation are difficult to deploy. These conventional systems include contact sensors, such as LVDT, displacement transducers, and accelerometers, as well as non-contact geodetic systems, such as global navigation satellite system (GNSS) receivers. At larger spatial scales and longer temporal horizons, satellite-based SAR/InSAR and DInSAR techniques provide complementary information on slowly varying deformation patterns and ground/structural displacement trends. Such remote-sensing measurements are not the focus of the present review, which concentrates on camera-based image/video measurement, but they are relevant as part of broader multi-sensor SHM architectures in which vision-based sensing, contact sensors, geodetic systems, and remote sensing operate at different spatial and temporal resolutions [1,2]. Compared with many traditional approaches, vision-based methods can provide non-contact measurements with high spatial density and relatively low deployment cost for large-scale structures [3–5]. However, converting image-plane information into metrologically reliable displacement and deformation measurements at the millimeter level under real field conditions remains a non-trivial challenge [6].

Despite substantial progress in computer vision algorithms, most civil-engineering applications of vision-based displacement measurement still report performance mainly through empirical error metrics, such as the root-mean-square error (RMSE) relative to a reference sensor, without a traceable, metrology-grade uncertainty assessment. The lack of explicit measurement models, comprehensive uncertainty budgets, and standardized reporting practices hampers the integration of vision-based sensors into safety-critical decision processes, algorithm tuning and validation against reference measurements, and long-term asset management. Furthermore, environmental and operational variability, including illumination changes, weather, temperature variations, camera vibration, and long standoff distances, can significantly affect the robustness and validity of vision-based systems deployed in the field [7,8].

In this context, metrological requirements refer to the conditions that allow a vision-based displacement estimate to be interpreted as a traceable measurement rather than only as an algorithmic output. These requirements include an explicit measurement model, calibration of the relevant optical and geometric parameters, identification and quantification of dominant uncertainty sources, propagation of input uncertainties to the measurand, consideration of reference-sensor uncertainty during validation, and standardized reporting of combined and expanded uncertainty. Addressing these requirements is particularly important when vision-derived displacements are used in safety-critical SHM decisions or in comparisons with conventional sensors.

This review targets these gaps by focusing on vision-based displacement and deformation measurement for civil SHM with an emphasis on

(i) metrology-grade uncertainty quantification (UQ) and (ii) robustness and validity under real field conditions. Rather than providing a generic overview of computer vision

in SHM, the paper concentrates on methods that deliver quantitative displacement (e.g., in millimeters) or strain (dimensionless, e.g., mm/mm or με), and that include either explicit uncertainty modeling, detailed calibration and error analysis, or field validation on full-scale structures. The time window considered spans approximately from 2010 to 2026, during which both traditional computer vision techniques (e.g., template matching, feature tracking, and DIC) and more recent deep-learning-based approaches were applied to SHM problems. The review also distinguishes between static and dynamic monitoring scenarios, because the advantages and limitations of vision-based techniques, as well as their dominant uncertainty sources and validation requirements, vary depending on the type of test.

To make the scope of the review explicit, the paper is organized around four core research questions. The questions are stated compactly below, while their scope and implications are progressively developed.

- 1: What are the dominant measurement models and processing pipelines used for vision-based displacement and deformation measurement in civil SHM?
- 2: What are the main sources of uncertainty in these pipelines, and how are they modeled, propagated, and reported?
- 3: Which uncertainty-quantification approaches are used in the literature, and what are their assumptions, strengths, and practical limitations for SHM applications?
- 4: What evidence is available on field robustness and validation of vision-based systems on full-scale civil structures, including reported failure modes and mitigation strategies?

The contribution of this paper is threefold. First, it provides a structured taxonomy of vision-based displacement and deformation measurement methods tailored to civil SHM, explicitly linking algorithmic families to their underlying measurement models. Second, it synthesizes a comprehensive uncertainty taxonomy and develops a metrology-oriented uncertainty budget framework for vision-based displacement, suitable for adaptation to different pipelines and deployment scenarios. Third, it reviews and critically compares existing UQ methodologies and field deployments, identifying practical limitations, recurring failure modes, and open research needs. In doing so, the paper aims to bridge the gap between computer vision research and metrological requirements in civil engineering, and to support the design of future vision-based SHM systems that are both field-robust and traceably reliable. The central position of this review is that vision-based SHM methods cannot be evaluated only as computer-vision algorithms; they must also be treated as measurement systems with explicit measurands, validity conditions, calibration chains, uncertainty budgets, and reference-comparison protocols. This perspective motivates the organization of the review and the proposed uncertainty-budget template. This paper is a structured review rather than an experimental validation study. The proposed projection–registration–mapping framework and uncertainty-budget template are not introduced as a proprietary measurement algorithm, but as a synthesis tool for organizing existing vision-based SHM methods and for supporting traceable uncertainty reporting across different pipelines.

The novelty of this review lies in reframing existing vision-based SHM pipelines as measurement systems that must satisfy metrological requirements. The proposed projection–registration–mapping framework provides the organizing structure for this reframing: it links algorithmic choices to the physical measurand, identifies where uncertainty enters the measurement chain, and supports the construction of comparable uncertainty budgets across different pipelines. In this sense, the framework offers a metrology-oriented synthesis and reporting tool, rather than yet another displacement estimation algorithm.

The framework should help clarify how vision-based displacement estimates can move from algorithmic outputs toward traceable SHM measurements.

Review Methodology and Selection Criteria

The reviewed studies were selected through a targeted literature search focused on vision-based displacement and deformation measurement for civil SHM. The search was conducted using major scientific databases and indexing platforms, including Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, MDPI, Google Scholar, and publisher-specific searches for journals in structural health monitoring, measurement science, civil infrastructure monitoring, and computer vision. The search combined keywords related to the measurement modality, application domain, and uncertainty treatment. Typical search strings included combinations of "vision-based displacement measurement", "image-based displacement", "video-based structural monitoring", "computer vision structural health monitoring", "digital image correlation civil structures", "phase-based motion magnification SHM", "UAV displacement measurement bridge", "camera-based displacement uncertainty", "vision-based SHM uncertainty quantification", "measurement uncertainty", "GUM", "Monte Carlo uncertainty propagation", and "field validation".

The search emphasized peer-reviewed journal articles and selected conference papers published approximately between 2010 and 2026. Earlier references were retained when they introduced foundational measurement models, metrological concepts, DIC uncertainty formulations, calibration methods, or widely used computer-vision algorithms that remain necessary for interpreting current practice. Recent studies from 2023–2025 were prioritized when they reported new developments in moving-camera and UAV-based measurement, heat haze and optical-turbulence mitigation, deep-learning-assisted pipelines, motion magnification, hybrid vision–sensor fusion, or uncertainty-aware SHM applications.

Studies were included when they satisfied at least one of the following criteria: (i) they reported quantitative displacement, deformation, strain, modal, or damage-related indicators derived from image or video data; (ii) they included an explicit measurement model, calibration procedure, uncertainty analysis, or comparison with reference sensors; (iii) they demonstrated laboratory validation, full-scale field deployment, or realistic operational conditions; or (iv) they contributed to one of the method families considered in this review, including marker-based tracking, markerless feature tracking, optical flow, DIC, phase-based motion magnification, moving-camera/UAV systems, hybrid vision–sensor fusion, or deep-learning-enhanced pipelines.

Studies were excluded from the core review when they focused only on qualitative visual inspection, generic crack classification, semantic damage detection, or image segmentation without a direct connection to displacement/deformation measurement, uncertainty quantification, calibration, or feature verification for measurement pipelines. Studies outside camera-based or image/video-based measurement, such as radar remote sensing or purely geodetic monitoring, were not treated as core literature, unless they were needed to contextualize reference sensing, sensor fusion, or complementary SHM architectures. The selected papers were then organized according to the projection–registration–mapping framework, the dominant uncertainty sources, the UQ method used, and the type of validation evidence reported.

The remainder of the paper is organized as follows. Section 2 describes the dominant measurement models and pipelines employed in vision-based displacement and deformation sensing. Section 3 introduces a detailed taxonomy of uncertainty sources along the measurement chain. Section 4 reviews UQ approaches, including GUM-aligned analytical and Monte Carlo methods, DIC error budgets, bootstrap techniques, and probabilistic deep learning. Section 5 presents a metrology-grade uncertainty budget template for vision-

based displacement measurements. Section 6 discusses implications for SHM practice, including the distinction between static and dynamic displacement monitoring, as well as issues of traceability and standardization, while Section 7 outlines open research gaps and future directions. Finally, Section 8 concludes the paper.

2. Measurement Models and Pipelines

The examples discussed in this section are selected as representative cases for each measurement-pipeline family, rather than as an exhaustive catalog of all available applications. They are used to illustrate how different algorithmic choices instantiate the projection–registration–mapping framework and how these choices affect uncertainty propagation, validation requirements, and field robustness. Vision-based displacement and deformation sensing in civil SHM is implemented through a variety of measurement pipelines that differ in camera configuration (monocular versus stereo), reliance on artificial targets, tracking algorithms, and whether the camera is fixed or moving. In this review, two orthogonal classification axes are distinguished as shown in Figure 1. The X-axis describes the scene instrumentation: marker-based pipelines rely on installed artificial targets, whereas markerless (often termed targetless in SHM) pipelines track natural features or textures on the structure and include feature tracking, optical flow, edge/phase methods, and DIC. The Y-axis describes the deployment mode: fixed-camera versus moving-camera acquisition (including UAV platforms), where the latter requires explicit ego-motion compensation and introduces additional uncertainty terms in the measurement model. In addition, depending on the application, these pipelines may be employed for either static or dynamic monitoring; rather than constituting a separate classification axis, this distinction is treated in the present review as an application-dependent regime that influences uncertainty sources, temporal requirements, and validation strategies.

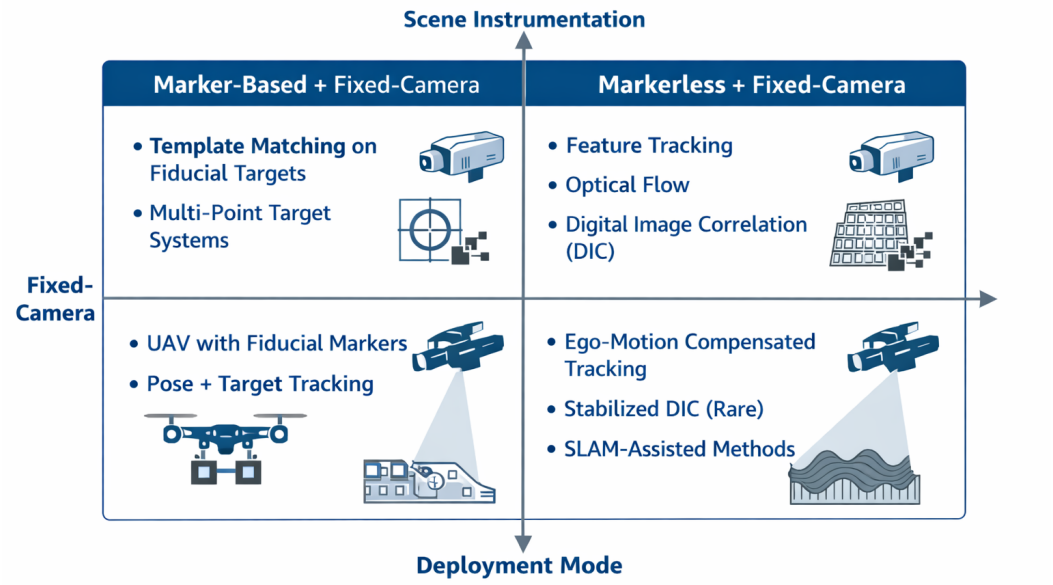


Figure 1. Orthogonal classification axes for vision-based displacement measurement in SHM, separating scene instrumentation (marker-based vs. markerless) from deployment mode (fixed-camera vs. moving-camera, including UAV).

Vision-based displacement monitoring in SHM covers both quasi-static and dynamic regimes. Although both follow the same general measurement chain, from pixel-domain motion estimation to geometric mapping, their dominant uncertainty sources and validation requirements are not the same. For this reason, the static-dynamic distinction is treated

in this review not as a separate classification axis, but as an application-dependent condition that affects uncertainty quantification, measurement validity, and field robustness.

In quasi-static applications, the main concerns are usually geometric mapping, calibration, long-term drift, and stability of the camera setup, since even small camera repositioning or gradual optical changes can accumulate into apparent structural displacement over long monitoring periods [9,10]. In dynamic applications, time-dependent effects become more important, including frame rate limitations, synchronization errors, motion blur, rolling-shutter distortion, and camera or platform vibration, often motivating explicit stabilization, ego-motion compensation, or fusion with inertial sensing [11]. In addition, target-free dynamic monitoring under moving loads and recent data-driven robustness studies show that distortion handling, scale consistency, atmospheric turbulence, occlusion, and camera shake can strongly affect practical field performance [12,13].

Static and dynamic tests are the principal load tests carried out on civil infrastructures. In current practice, they are not jointly performed on the same structure, so correlations are, in general, not available, except for testing in a controlled environment [4,14]. Nevertheless, data fusion of the high-frequency displacements evaluated in a dynamic test with the low-frequency displacements obtained in a static test should be considered if both tests are performed. Quasi-static tests involve costly, time-consuming, and complex experimental setups (i.e., the weight vehicle columns on a standard road bridge). In general, the dataset from a quasi-static loading test consists of displacements and strains at a finite number of sections, determined by the available number of sensors. Dynamic tests involve ambient and free vibration tests and provide accelerations and strain data; only in a limited number of cases has the detection of displacements been added [10]. The ambient vibration ones are used to calibrate the modal parameters—i.e., natural frequencies and mode shapes. Typically, this information allows the calibration of numerical models of the structure (bridge or building), usually performed via finite element (FE) analysis. Free vibration tests provide accurate estimates of damping coefficients and inelastic properties of the structure. Numerical analysis is an iterative process that uses data from experimental tests to obtain time histories, frequency response curves, or mode shapes and frequencies, thereby reducing differences between the numerical and experimental modal parameters [15].

The iterative process allows the FE model to be refined to better represent the structure's actual behavior. They are used for structural evaluations, both to assess the behavior of a newly built structure, as well as fatigue strength in an existing one, to provide early warning signs of structural failure, and to design interventions on a damaged structure. This last case needs accurate evaluation of the structural inelastic behavior, to be determined after accurate dynamic and static numerical analyses [16].

Vision-based measurement systems typically provide displacement time histories and can be employed in both static and dynamic tests. However, their sampling frequencies are generally lower than those required for high-fidelity dynamic analyses, and estimating accelerations from double-differentiated displacement signals may significantly amplify measurement noise.

This section introduces a unified measurement-model framework and then classifies dominant pipelines into marker-based, targetless feature-tracking, and optical-flow methods, DIC, and moving-camera configurations. The focus is on how pixel-level observations are transformed into metric displacement or strain, and on the modeling assumptions that later underpin UQ.

2.1. Unified Measurement-Model Framework

Most vision-based displacement systems can be described by a two-stage model: (i) an image formation stage, where three-dimensional structural points $\mathbf{X}$ are projected

to image coordinates $\mathbf{p}$ via a camera model, and (ii) a motion estimation stage, where changes in image coordinates $\Delta\mathbf{p}$ between frames are converted to physical displacement $\mathbf{d}$ or deformation measures. For a calibrated pinhole camera with radial and tangential distortion [17], the forward projection from a 3D point $\mathbf{x}$ in the world frame to a distorted pixel coordinate $\mathbf{p}$ can be written as

$$\mathbf{p} = \Pi\big(\mathbf{K}, \mathbf{R}, \mathbf{t}, \mathbf{x}, \boldsymbol{\theta}_{\text{dist}}\big) \tag{1}$$

where $\mathbf{K}$ denotes the intrinsic parameters (pixel size, focal lengths, and principal point), $(\mathbf{R}, \mathbf{t})$ are the extrinsic parameters (rotation and translation from the reference coordinate system to the camera frame), and $\boldsymbol{\theta}_{\text{dist}}$ collects the lens distortion coefficients. Under small displacements, the physical displacement increment $\mathbf{d}$ of a tracked point, expressed in the reference frame and, in the most general case, represented as a 3D vector, can be related to the corresponding pixel/subpixel image shift $\Delta\mathbf{p}$ through a local inverse mapping evaluated at a nominal operating point. Equivalently, this relation may be written as a first-order linearization of the pixel-to-metric mapping, which is valid for sufficiently small $\Delta\mathbf{p}$.

$$\mathbf{d} \approx \mathbf{J}_{\text{inv}}(\mathbf{K}, \mathbf{R}, \mathbf{t}, \mathbf{x}_0, \boldsymbol{\theta}_{\text{dist}})\, \Delta\mathbf{p}, \tag{2}$$

where $\mathbf{J}_{\text{inv}}$ has units of physical length per pixel and is obtained by local linearization of the camera projection model (including lens distortion) and, when applicable, planar constraints (e.g., homography-based mapping). Here, $\mathbf{x}_0$ denotes the nominal 3D point or reference configuration at which the linearization is performed. [14,18,19]. The complete measurement pipeline then consists of camera calibration (estimation of $\mathbf{K}$, $\boldsymbol{\theta}_{\text{dist}}$, and possibly $(\mathbf{R}, \mathbf{t})$), image acquisition, image registration or tracking to estimate $\Delta\mathbf{p}$, and geometric conversion to $\mathbf{d}$.

> **Notation and conventions.** $\Delta\mathbf{p} \in \mathbb{R}^2$ denotes the pixel displacement vector (registration output), $\Delta\mathbf{p} = [\Delta u, \Delta v]^\top$. The projection onto the measurement axis is $\Delta p_{\text{axis}} = \mathbf{a}^\top \Delta\mathbf{p}$ with $\|\mathbf{a}\| = 1$. $\mathbf{d}$ denotes the physical quantity (metric units): $d \in \mathbb{R}$ for pointwise displacement, $\mathbf{d} \in \mathbb{R}^2$ for planar motion, and $\mathbf{d}(\mathbf{x})$ for full-field measurements. $\boldsymbol{\theta}_g$ collects geometric parameters used in mapping (intrinsics/extrinsics/pose/scene geometry). We refer to mapping **Level 0** (scale), **Level 1** (homography), and **Level 2** (full Jacobian). Throughout the paper, we use residual to denote the model mismatch term (unmodeled effects) in the measurement equation. We use uncertainty for quantified dispersion (standard uncertainty u, combined u_c, and expanded U) associated with input quantities and model parameters. We use error only when a reference is available, and its uncertainty is stated, i.e., for the difference between the estimate and a reference value (measurement agreement).

2.1.1. Levels of Geometric Mapping: From a Single Scale Factor to a Full Jacobian

Regardless of the selected registration/tracking strategy, vision-based displacement estimation can be written as a two-stage measurement model: (i) a registration stage that estimates the pixel-domain displacement $\Delta\mathbf{p}$ from image data, and (ii) a geometric mapping stage that converts $\Delta\mathbf{p}$ to the physical measurand (displacement) $\mathbf{d}$. To unify the notation across the method families considered in this review, we express the mapping as

$$\mathbf{d} = g\big(\Delta\mathbf{p};\, \boldsymbol{\theta}_g\big), \qquad \boldsymbol{\theta}_g = \{\mathbf{K}, \mathbf{R}, \mathbf{t}, \mathbf{x}, \boldsymbol{\theta}_{\text{dist}}\}, \tag{3}$$

where $\boldsymbol{\theta}_g$ collects the geometric parameters used in the mapping, including camera intrinsics $\mathbf{K}$, pose $(\mathbf{R}, \mathbf{t})$, scene or reference geometry $\mathbf{x}$ (e.g., a 3D point location), distortion parameters $\boldsymbol{\theta}_{\text{dist}}$, and any additional constraints such as planarity. The term $\boldsymbol{\theta}_{\text{dist}}$ denotes the vector of coefficients of the adopted lens-distortion function, for example, a polynomial

radial/tangential distortion model. In practice, distortion is often corrected (undistortion) before applying planar or scale mappings.

In practice, $g(\cdot)$ is often implemented through one of the following approximation levels, which form a hierarchy from simple to metrology-grade models:

Level 0: Scalar Scale Factor (Locally Linear, Fixed Geometry)

If perspective effects are negligible over the region of interest (ROI) and motion is predominantly along a known axis, the mapping can be approximated by a constant scale factor S:

$$d \approx S\,\Delta p_{\text{axis}}, \qquad \Delta p_{\text{axis}} = \mathbf{a}^{\top}\Delta\mathbf{p},\ \|\mathbf{a}\| = 1, \tag{4}$$

where Δp_{axis} is the pixel displacement projected onto the measurement axis. This approximation is common in long-distance, small-ROI monitoring when camera pose is fixed and out-of-plane motion is negligible [20].

Level 1: Planar/Projective Mapping (Homography/Affine)

If the ROI is approximately planar but perspective variations across the ROI are non-negligible, a planar mapping based on a homography (or an affine approximation) is more appropriate. In displacement form, this is commonly expressed through a local linear mapping [21], i.e., the Jacobian of the planar/projective transformation evaluated at the operating point:

$$\mathbf{d} \approx \mathbf{J}_{\text{H}}\,\Delta\mathbf{p}, \tag{5}$$

where $\mathbf{J}_{\text{H}}$ denotes a local Jacobian (typically 2×2) of the planar mapping, capturing spatially varying scale within the ROI.

Level 2: Full Jacobian (Pinhole + Pose, General Local Linearization)

For general 3D settings, wide ROIs, significant perspective or out-of-plane effects, or moving-camera configurations, a physically consistent local linearization can be expressed in terms of the Jacobian of the displacement mapping with respect to the image-plane subpixel/pixel shift [21,22]:

$$\mathbf{d} \;\approx\; \mathbf{J}_{\Delta\mathbf{p}}\big|_{\xi_0}\,\Delta\mathbf{p}, \tag{6}$$

where $\Delta\mathbf{p} = [\Delta u\ ,\ \Delta v]^{\top}$ denotes the measured subpixel displacement in the image. The Jacobian $\mathbf{J}_{\Delta\mathbf{p}}$ is obtained by locally linearizing the (generally nonlinear) mapping $\mathbf{d} = g(\Delta u, \Delta v; \mathbf{x})$ around a nominal configuration $\boldsymbol{\xi}_0$, which collects the relevant geometric and calibration parameters, yielding

$$\mathbf{J}_{\Delta\mathbf{p}} \approx \left.\begin{bmatrix} \dfrac{\partial g_x}{\partial \Delta u} & \dfrac{\partial g_x}{\partial \Delta v} \\ \dfrac{\partial g_y}{\partial \Delta u} & \dfrac{\partial g_y}{\partial \Delta v} \\ \dfrac{\partial g_z}{\partial \Delta u} & \dfrac{\partial g_z}{\partial \Delta v} \end{bmatrix}\right|_{\xi_0} \begin{bmatrix} \Delta u \\ \Delta v \end{bmatrix}. \tag{7}$$

This form makes explicit that the approximation is linear in $\Delta\mathbf{p}$ at the operating point.

Validity Conditions and Practical Guidance

Level 0 is valid when (i) the ROI is small enough that local scale is approximately constant, (ii) camera pose and intrinsics are stable, (iii) depth variation and out-of-plane motion within the ROI are negligible relative to the required accuracy, and (iv) the displacement direction is known or constrained. When any of these conditions is violated, the appropriate upgrade depends on the geometric origin of the mismatch. If the ROI remains

approximately planar and the main limitation is spatially varying perspective across that plane, a Level 1 mapping based on a homography (or affine approximation) is generally sufficient. By contrast, if the scene cannot be approximated as planar under a single global linearization, or if out-of-plane motion, depth variation, or time-varying camera pose must be explicitly represented, the mapping should be upgraded to Level 2, i.e., a full Jacobian-based local linearization of the underlying 3D pixel-to-metric model [23]. Importantly, the mapping level determines which uncertainty contributors dominate. Level 1 already depends on uncertainty in the estimated planar geometric mapping, including the uncertainty and covariance of homography- or affine-model parameters. Level 2 extends this dependency to a fuller 3D geometric description, including camera intrinsics and distortion parameters, pose parameters, and reference 3D geometry (e.g., $\mathbf{K}$, $\boldsymbol{\theta}_{\text{dist}}$, $\mathbf{R}$, $\mathbf{t}$, and $\mathbf{x}$, where $\mathbf{x}$ denotes the nominal 3D point or reference geometry in the world/reference frame), as well as uncertainties in timing and synchronization. As summarized in Figure 2, the reviewed method families share the same general measurement-chain structure, while differing mainly in the feature or ROI definition, registration strategy, and geometric mapping level.

	Geometry/calib.	Feature/ROI	Registration (Δp)	Mapping ($S/H/\mathbf{J}$)	Output
Template	Fixed pose or known scale	Marker/patch template	Correlation subpixel peak	Often Level 0/1 (S or H)	Pointwise $d(t)$
KLT	Fixed pose or calibrated	Corners/keypoints (sparse)	KLT tracks (track set)	Level 0/1 common Level 2 if needed	Tracked-point $d(t)$
Flow	Fixed pose or calibrated	Dense grid implicit	Optical flow vector field	Level 1/2 (large ROI)	Field $d(\mathbf{x}, t)$
DIC	Mono/stereo calibrated	Speckle/texture subsets	Subset/global corr. warp	Mono: Level 0/1 Stereo: Level 2	Disp. + strain
Moving cam	Time-varying pose IMU/GNSS/SLAM	Structure + background ref	Ego-motion est. + compensation	Typically Level 2 ($\mathbf{J}$)	Residual $d(t)$

Figure 2. Family-wise overlay of vision-based displacement pipelines, highlighting shared stages (geometry, feature/ROI, registration, mapping) and the blocks that differ across method families. The mapping level ($S/H/\mathbf{J}$) should be selected based on geometry validity conditions (Section 2.1.1).

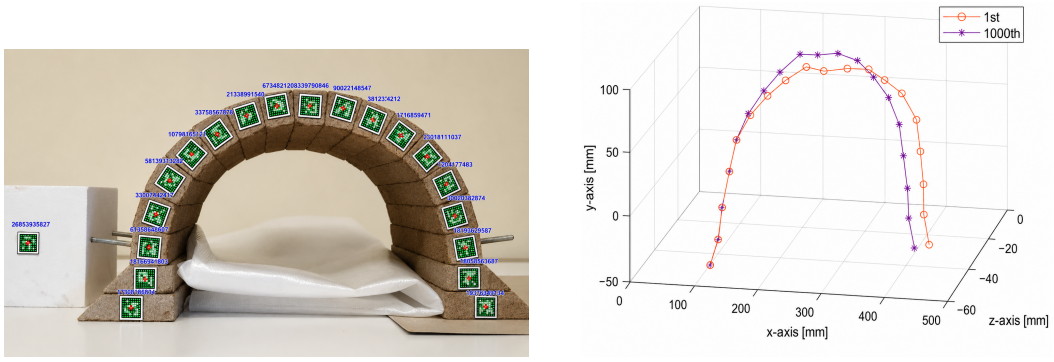
2.2. Marker-Based Monocular Pipelines

Marker-based systems attach high-contrast artificial targets (e.g., printed patterns, circular patches, retro-reflective markers) to selected structural points, and use a fixed monocular camera to track the motion of these targets over time. Early work proposed a vision-based displacement sensor using a robust object-search algorithm with subpixel image registration and time-synchronized acquisition [18]. In that pipeline, the main steps are as follows:

1. **Camera setup and calibration**: the camera is mounted on a stable support at a known distance from the structure; intrinsic parameters and lens distortion are calibrated using checkerboard or calibration targets; a scale factor S is obtained either from known target dimensions or from a reference object in the scene.
2. **Target detection and ROI definition**: image regions containing markers are defined as ROI; marker centers or corners provide canonical tracking points.
3. **Template matching and subpixel registration**: for each ROI, a reference template at the initial time is correlated with subsequent frames using normalized cross-correlation or motion-estimation schemes; subpixel shifts $\Delta\mathbf{p}$ are obtained from the correlation peak or from a parametric motion model [18,24].
4. **Pixel-to-metric conversion**: the measured pixel displacement is converted into physical displacement either through a constant scale factor derived from calibration (Level 0) or, when planar perspective effects must be accounted for, through a homography-based mapping (Level 1).

This basic template-matching pipeline was adopted and extended in many SHM studies. Multi-point displacement measurement systems were developed for building

frames and bridges by combining upsampled cross-correlation and orientation code matching to track multiple markers in a single field of view [14]. A non-contact, multi-point system was demonstrated on a cable-stayed footbridge, with markers placed on the deck and cables, providing millimeter-level accuracy relative to accelerometers and laser displacement sensors [25]. At a smaller scale but with a metrological emphasis (Figure 3), low-cost marker-based tracking systems for 3D-scaled masonry models were proposed, leveraging DeepTag fiducial markers to achieve reliable pose and displacement tracking in controlled laboratory conditions [26,27]. These works explicitly model the geometric mapping between marker pose and structural displacement, enabling later uncertainty analysis.



(**a**) (**b**)

Figure 3. An example of a marker-based technique for span monitoring using the DeepTag method: (**a**) marker-based tracking configuration on the scaled masonry model; (**b**) 3D tracking result comparing the initial frame (orange line) and the 1000th frame (purple line) [26].

In Figure 4, this family primarily instantiates the registration block $\mathcal{R}(\cdot)$ via template correlation and subpixel peak estimation, yielding pointwise $\Delta\mathbf{p}$. In many SHM deployments with fixed cameras and limited perspective variation, Level 0 mapping ($d \approx S\,\Delta p$) is adequate; however, when the ROI is large, or the observation geometry is oblique, Level 1/2 mappings should be preferred to avoid geometric bias.

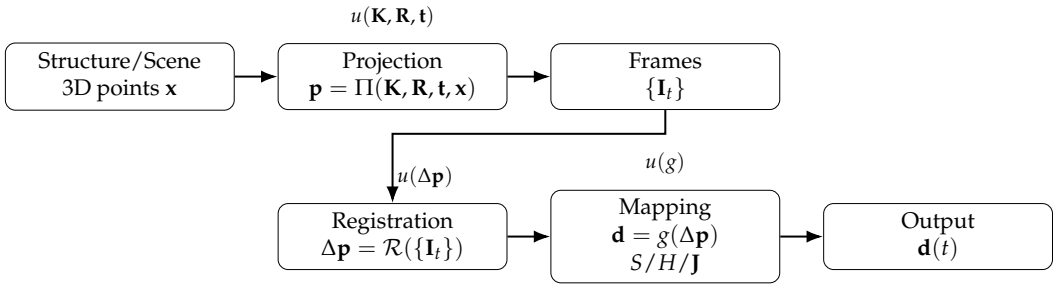


Figure 4. Unified projection-plus-registration measurement model. Pixel motions $\Delta\mathbf{p}$ are estimated via registration/tracking and mapped to metric displacements $\mathbf{d}$ through a geometry-dependent mapping $g(\cdot)$ with multiple approximation levels (scalar scale S, planar mapping H, or full Jacobian $\mathbf{J}$).

A vision-based system, with AprilTags as target markers, was used on the historic bowstring iron truss structure (Figure 5) of the pedestrian Blackfriars Bridge in London, Ontario, Canada, an 1875 wrought iron bowstring arch recently subjected to extensive repair interventions [28]. The ten markers were placed on the tied arch bridge deck, in the middle of the span (Figure 5, red frame). A system of seven accelerometers was installed on the entire span of the base beam, at deck level. The displacement measurement and the modal analysis were limited to the bridge deck only. During the tests, the bridge was in service, so ambient vibrations were recorded.

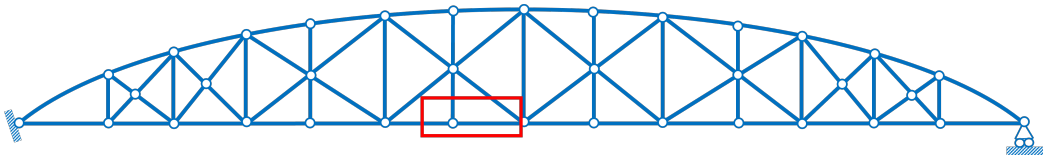


Figure 5. The truss structure of the Blackfriars Bridge. In the red area, the part of the deck arranged with markers.

Vision-based measurement systems provide displacement time histories by tracking markers over time, enabling the extraction of dynamic information such as modal parameters. Based on these displacement measurements, the authors compared the identified modal properties with reference values obtained from accelerometers, reporting a maximum relative error of 5.22% for the natural frequencies of the first three deck mode shapes. Overall, the modal parameters were well captured; however, the authors also highlighted practical challenges in data extraction due to environmental factors such as reflection, refraction, and wind effects. Moreover, the displacement was evaluated only in the central part of the bridge, where vertical displacements with limited rotations are expected. It should be noted that a bowstring bridge like the Blackfriars Bridge is a rigid structure, which presents large rotations at the supports and limited flexural displacements of the deck. Monitoring of civil infrastructure could be limited to particularly vulnerable or damaged parts of a structure. In this case, a detailed evaluation of displacements should be performed, since a simple 2D analysis [28] is insufficient to completely describe the behavior of the substructure. In [9], the 6-DOF displacement measurement of a bridge bearing is presented, in which an algorithm for camera ego-motion compensation is developed. The AprilTags employed in the field tests are placed on the upper and lower rigid parts of the bearing, and measurements were taken over two days and four hours to account for strains due to thermal variations. In the examined case, it is extremely important to detect rotations, since the Korean railway bridge design standards fix a limit for the bearing rotation. The method proposed by the authors to evaluate the 6-DOF displacements achieves low RMSE, with maximum values of 0.17 mm for displacement and 0.46 mrad for rotation.

2.3. Markerless Feature Tracking and Optical Flow

Targetless methods avoid installing artificial markers by tracking natural features on the structural surface, such as edges, bolts, or texture patterns. Their pipelines differ in feature representation and tracking algorithms but share common stages of camera calibration, feature detection, feature tracking, and geometric conversion [8].

In sparse feature tracking, salient points are detected using corner or blob detectors and then tracked using Kanade–Lucas–Tomasi (KLT) or related optical-flow algorithms. Different multi-object tracking algorithms were evaluated for multi-point displacement extraction on laboratory frames, including KLT-based tracking and correlation filters [29]. For field-scale bridges, natural features on the deck or towers can be tracked to infer displacement, provided that robust feature descriptors and outlier rejection schemes (e.g., RANSAC, forward–backward consistency) are employed.

Dense optical-flow methods estimate a motion vector for every pixel between frames, supporting full-field displacement estimation. Displacement identification algorithms were developed in which optical flow and spatiotemporal context are used to track non-target features across image sequences under operational conditions [30,31]. Robust and versatile dynamic displacement monitoring of natural feature targets has also been demonstrated on large-scale structures using illumination-robust optical flow and adaptive ROI management [32]. In these pipelines, image displacement fields are converted to metric displacement via homography or pinhole geometry, with performance validated against contact sensors. Non-target displacement measurement using reference-frame-based Deep-

Flow provides additional algorithmic detail and full-field examples that fit into the same measurement-model structure [33].

Recent metrology-oriented studies have explored deep-learning-assisted tracking for scaled masonry models and for defect-related feature extraction. Concrete crack detection and segmentation using UAV imagery and deep learning [34] and triplet-loss-based crack verification for SHM and digital twin applications [35] do not directly output displacement, but they highlight how feature extraction and verification networks can be integrated into displacement pipelines, for example, to ensure that tracked features correspond to structurally meaningful locations.

This family modifies the registration block by estimating $\Delta\mathbf{p}$ either sparsely (KLT tracks) or densely (optical-flow fields). As a result, the displacement output may be pointwise (tracked features) or spatially distributed. Mapping is often Level 1 when the ROI can still be approximated as planar, even if it is relatively large. By contrast, Level 2 becomes necessary when the measurement geometry is genuinely three-dimensional, for example, in the presence of significant out-of-plane motion, depth variation across the ROI, or explicit time-varying camera pose; in such cases, calibration and pose uncertainties must also be propagated through the full Jacobian-based mapping.

A dynamic displacement measurement system and a modal identification method for a simply supported beam using a single camera are presented in [12]. The simply supported 6 m steel beam with a hollow rectangular cross section, which could easily be considered a large-scale sample, was subjected to a moving load, i.e., a motorized scaled vehicle. Vision-based measurements, taken to target points on the structure, were compared with those given by five standard contact-type LVDT sensors, placed at a constant distance of 1 m.

As shown in Figure 6, while good agreement could be observed for the central and lateral sensors (S3–S1), displacements detected by sensor S2 show a significant deviation from those detected via the vision-based system. This issue is evident from the comparison of the first and second mode shapes identified by the two measurement systems, and is confirmed by the higher RMSE reported by the authors.

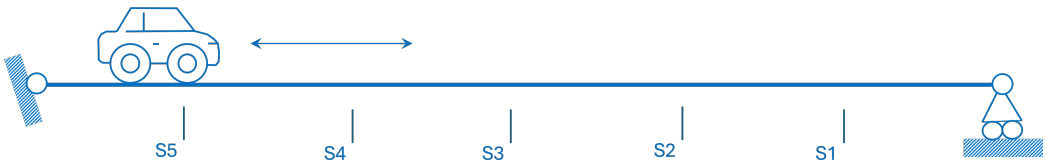


Figure 6. Scheme of the structure, loading, and position of LVDT sensors.

2.4. Digital Image Correlation and 3D Photogrammetric Pipelines

DIC treats displacement estimation as an image-registration problem between reference and deformed images using speckle patterns [36]. In subset-based two-dimensional DIC, the pipeline involves: (i) preparing a random speckle pattern on the surface, (ii) acquiring reference and deformed images with a calibrated camera, (iii) defining subsets around grid points in the reference image, (iv) estimating the displacement and deformation gradient that best maps reference subsets to deformed subsets by maximizing cross-correlation or minimizing a sum-of-squared-differences criterion, and (v) interpolating subpixel positions using, for instance, Newton–Raphson or inverse compositional Gauss–Newton schemes. A detailed review of two-dimensional DIC computational methods summarizes the influence of subset size, speckle statistics, interpolation schemes, and image noise on displacement and strain resolution, and provides theoretical and experimental error estimates [37] as shown in Figure 7. Additional studies quantify DIC displacement accuracy and investigate

the influence of speckle and recording setup, establishing widely used error models for displacement and strain [36,38].

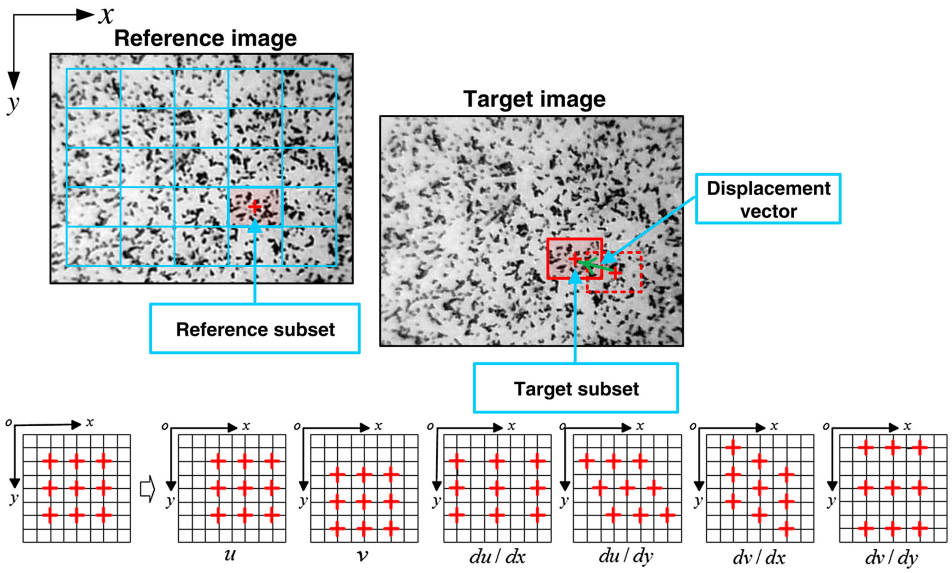


Figure 7. Example schematic showing an undeformed and a corresponding deformed subset in 2D-DIC, illustrating six different types of linear transformations applied to the reference subset [37]. The red "+" symbols indicate the subset center/reference points.

Stereo DIC and photogrammetric approaches extend these concepts to three dimensions using calibrated stereo camera pairs. A photometric DIC framework based on textured digital twins was proposed, which unifies global stereo DIC and detailed camera modeling, including distortion and radiometric effects [39]. In structural engineering, digital photogrammetry was used for 3D structural vibration measurement, reconstructing the motion of building and beam models from multi-camera configurations [40]. These pipelines share the same core elements: camera calibration, subset or global correlation to obtain image displacements, and triangulation or projection-based conversion to 3D displacement and strain. Microscopic vision-based displacement sensing with patterned optical discs has also been investigated to achieve high resolution and large measuring range, illustrating the breadth of vision-based displacement concepts [41].

DIC instantiates registration through subset/global correlation with an explicit warp model, producing displacement (and potentially strain) fields. Mono DIC commonly uses Level 0/1 mappings when the measurement is restricted to in-plane motion under weak perspective, so that a scale-factor or planar mapping is sufficient. By contrast, stereo/3D DIC inherently requires Level 2 mapping, because the displacement is obtained through full 3D reconstruction and cannot be represented by a purely planar image-to-metric transformation. In this case, calibration and triangulation uncertainties can also be propagated through the same 3D mapping.

2.5. Phase-Based Motion Magnification and Edge-Based Pipelines

Phase-based motion magnification (PBMM) and related phase-domain methods provide a different measurement route from intensity-correlation methods such as DIC. Rather than estimating displacement directly by subset matching or image registration, PBMM represents image content using a multi-scale, multi-orientation complex steerable pyramid and relates small motions to local phase variations. Temporal band-pass filtering of the phase component at frequencies of interest can amplify subtle vibrations in selected bands, after which displacement time histories may be recovered from the phase signal itself or by combining PBMM with edge tracking or DIC-like registration [42]. PBMM is therefore better

viewed as a phase-based motion extraction and amplification framework that can support displacement estimation, particularly for small-amplitude vibrations, rather than simply a video-input variant of DIC. In such cases, the measurement model includes both the phase-to-pixel-motion relation and the subsequent mapping from image-domain motion to physical displacement, including calibration of the amplification and geometric mapping. Recent building-scale applications further show that phase-based motion magnification can support post-event and damage-oriented SHM when combined with independent vibration and geometric references. For example, advanced video-based processing was applied to low-cost high-speed footage of a two-story reinforced-concrete frame tested under seismic loading on a shaking table, with the video-derived dynamic response compared against accelerometers and a passive 3D motion-capture system [43]. Such studies are particularly relevant because they connect magnified video measurements not only to displacement or vibration visualization, but also to modal-parameter estimation, pre- and post-damage comparison, and independent geometric benchmarking. From a metrological perspective, they highlight the need to propagate image-level and amplification-related uncertainty through subsequent dynamic-identification steps.

Edge-based pipelines avoid explicit template matching by tracking structural edges at subpixel precision. High-speed edge tracking using a modified Taylor-approximation optical-flow approach was demonstrated, focusing on algorithmic robustness and computational efficiency for displacement sensing [24]. Advanced edge-detection schemes have also been proposed for non-contact structural displacement monitoring, where subpixel edge positions are converted into displacement measures along girder or deck boundaries, providing high sensitivity without installing markers.

2.6. Moving-Camera and Ego-Motion-Compensated Pipelines

While most early vision-based SHM implementations assume a fixed camera, recent work has investigated moving-camera configurations, including deck-mounted cameras on bridges and cameras mounted on UAVs. In these settings, camera ego-motion must be estimated and compensated to obtain structural displacement in a global frame [44,45].

For fixed but vibrating camera mounts, background features within the same field of view can be used to estimate camera motion via homography or pose estimation, which is then subtracted from the apparent motion of structural features. A non-target vision sensor for remote measurement of bridge dynamic response explicitly addresses camera movement by referencing stationary background objects, enabling non-contact monitoring at long standoff distances [19,46]. More recently, reference-free ego-motion compensation was introduced for miter gates, where relative deformations between structural points are estimated while eliminating rigid-body motion of both the structure and camera system [47].

UAV-based pipelines integrate onboard inertial measurement unit (IMU) and GNSS/RTK measurements with vision to determine camera pose and perform ego-motion compensation. The typical steps include: (i) estimating UAV pose using inertial and navigation sensors, (ii) projecting camera rays into a global frame using the calibrated extrinsic transformation between camera and IMU, (iii) tracking structural features in the video, and (iv) solving for structural displacement consistent with both image observations and estimated camera pose. Existing studies on UAV-based crack detection and scene reconstruction illustrate how displacement sensing and defect monitoring can be combined in a single platform [34]. Feasibility studies on UAV-based multi-point displacement measurement of bridges follow the same measurement-model structure, even when detailed uncertainty propagation is not yet fully developed.

For moving cameras, the observed pixel motion mixes structural motion with ego-motion. Therefore, the pipeline must explicitly estimate and compensate camera motion

before interpreting the residual $\Delta\mathbf{p}$ as structural displacement. The pose parameters are then time-dependent states, $\mathbf{R}(t)$ and $\mathbf{t}(t)$, produced by an ego-motion estimator. As a result, the displacement can no longer be represented by a fixed planar mapping, and a Level 2 model is typically required, since the pixel-to-metric conversion depends explicitly on the full geometry and on time-varying camera pose. In this setting, uncertainty contributions also shift toward pose estimation, synchronization, and residual motion-compensation errors.

2.7. Hybrid Vision–Sensor Fusion Pipelines

Hybrid pipelines combine vision with other sensing modalities to overcome limitations of individual sensors and enhance robustness. Time-synchronized, wireless vision sensor networks were developed to compute multi-point displacements using low-cost cameras, synchronized clocks, and accelerometer references, with field testing on civil structures [44]. In these systems, vision provides spatially rich displacement data at lower sampling rates, while accelerometers supply high-frequency information; the two are fused through state-space models or Kalman filters, effectively defining a combined measurement model with explicitly modeled sensor noise and synchronization errors.

Structural dynamic displacement measurement based on vision and vibration data fusion has also been proposed, in which camera-based displacement and accelerometer-based response are integrated to produce continuous, drift-free displacement time histories [4]. Marker-based tracking systems for masonry models emphasize rigorous calibration and evaluation protocols that are directly applicable to SHM fusion scenarios, ensuring that displacement estimates are consistent with geometric constraints and reference measurements [26,27].

A data fusion method is presented and discussed in [11], where a field test was conducted on the Dongshuimen Bridge on the Yangtze River, known as one of the two twin River bridges in Chongqing, China. The midspan section of the two-deck steel truss girder cable-stayed bridge was monitored during a 30-min time interval, and ambient vibrations were detected. A stationary target was considered to account for camera vibrations, although negligible ground vibration is expected, since the employed tilt camera is 300 m from the measurement points. During the field tests, trains traveled on the lower deck, where an accelerometer was placed at midspan, while trucks, cars, and pedestrians traveled on the upper deck. The authors propose a linear relationship between vision-based displacements and those evaluated from accelerometric measurements by fitting the data within the same frequency band. The authors' proposal could be a first step toward the use of surveillance cameras, which are generally mounted on the bridge structure and therefore subjected to bridge vibrations, to perform SHM.

2.8. Deep-Learning-Based Pipelines

Deep learning is increasingly used within vision-based SHM pipelines, either to enhance image quality (denoising, super-resolution, heat-haze mitigation), to provide robust feature detection and tracking, or to directly infer displacement-related quantities from video. These approaches generally wrap a neural network module around the projection-plus-registration measurement model described in Section 2, rather than replacing it entirely.

One common usage is pre-processing to improve effective spatial resolution and robustness at long standoff distances. Super-resolution networks and learned denoisers were investigated for bridge monitoring, where the network enhances frames before classical template matching or feature tracking is applied. In long-span bridge applications, enhanced template-matching pipelines have achieved improved displacement resolution and robustness to blur and low contrast compared to purely classical methods, while retaining explicit geometric calibration and pixel-to-metric conversion [48,49]. Similar ideas

appear in methods that learn to suppress optical-turbulence distortions or heat-haze effects prior to tracking, thereby reducing image-level uncertainty and improving the stability of subpixel localization [49,50].

A second usage is deep feature detection and tracking. Instead of hand-crafted features, convolutional networks are trained to detect and describe keypoints on structural surfaces, or to segment regions of interest (e.g., edges, joints, cracks) that are subsequently tracked by classical or learned trackers. DeepTag-based marker detection for scaled masonry models exemplifies this pattern: a network detects fiducial markers and estimates their poses, which are then mapped to structural displacements through calibrated geometry [26,27]. In the context of UAV-based inspection, deep networks perform crack detection and segmentation on aerial images [34], and their outputs can be used to define or verify feature locations to be tracked over time for displacement or deformation measurement. Triplet-loss-based crack verification further illustrates how deep representations can be used to ensure the identity and consistency of tracked features across views or time, supporting robust long-term monitoring [35]. An example of this deep representation-based verification strategy is shown in Figure 8.

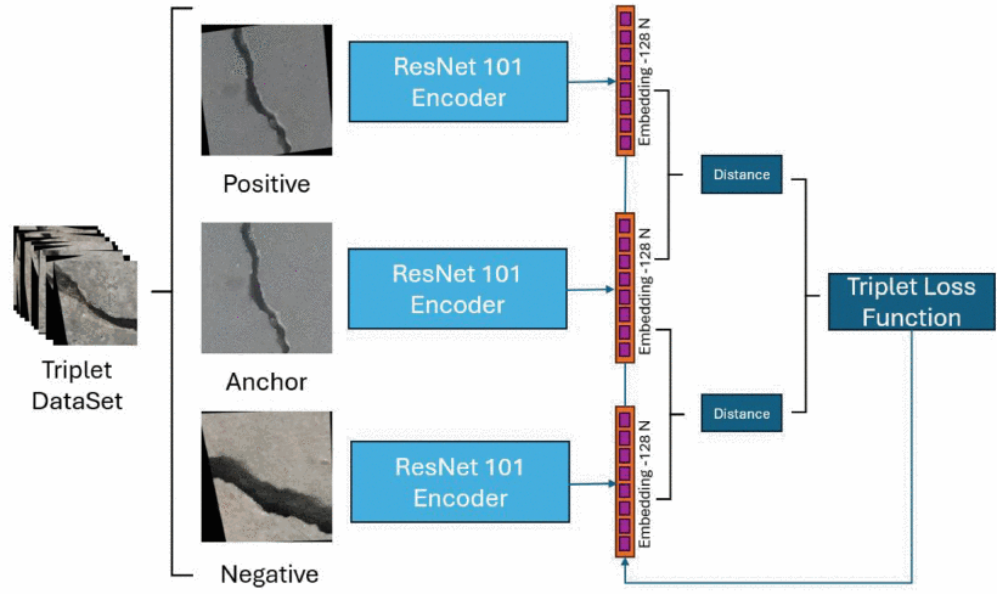


Figure 8. Diagram of a Siamese network used for crack verification, showing a triplet dataset fed into three ResNet-101 encoders to generate embeddings, which are subsequently compared to calculate the triplet loss [35].

A third class of approaches uses deep networks for end-to-end regression from image sequences to displacement time histories or dynamic parameters. In such pipelines, the network implicitly learns both feature representations and motion mapping from training data labeled with reference displacements (e.g., from LVDTs or laser sensors). While early examples focus more on modal identification and damage detection than on displacement itself, recent work on displacement identification under operational conditions has started to incorporate temporal convolutional or recurrent architectures on top of optical-flow or feature trajectories [30,31]. These methods can capture complex spatiotemporal dependencies and can be trained to be robust to illumination changes and occlusions, but they often behave as black boxes and require careful calibration and validation to ensure traceable displacement measurements.

Across all these deep-learning-based pipelines, the core geometric relationship between image motion and physical displacement still depends on camera calibration and the underlying projection model. Deep networks primarily modify or enhance how $\Delta\mathbf{p}$ (or directly $\mathbf{d}$) is estimated from the image data, and may introduce new sources of epistemic and aleatoric uncertainty related to training data, model architecture, and domain shifts. These uncertainties must be considered alongside classical components when constructing metrology-grade uncertainty budgets, as discussed later in Sections 3 and 4.

3. Uncertainty Sources: A Comprehensive Taxonomy

The accuracy and reliability of vision-based displacement and deformation measurements are governed by a diverse set of uncertainty sources spanning optics, imaging, geometry, time, environment, algorithms, and reference sensors. This section provides a structured taxonomy of these sources, with emphasis on how they arise in the measurement chain introduced in Section 2 and how they have been characterized in recent literature. Also, the block structure in Figure 4 provides a natural scaffold for uncertainty attribution: calibration and imaging terms enter the projection block, algorithmic terms enter the registration block, and geometric-assumption terms enter the mapping block, while Figure 9 demonstrates uncertainty sources in vision-based displacement.

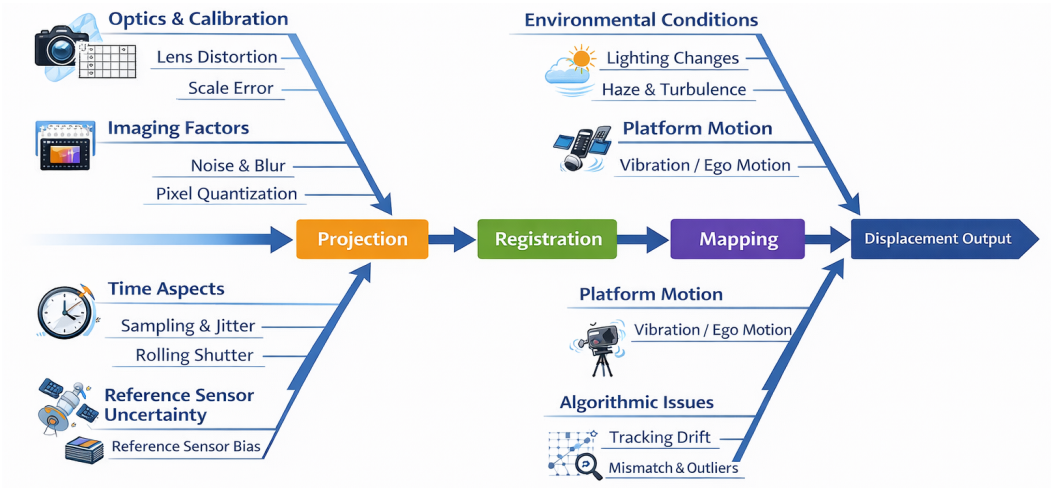


Figure 9. Fishbone diagram summarizing key uncertainty sources in vision-based displacement measurement and their linkage to the processing chain (projection–registration–mapping) leading to the displacement output.

3.1. Camera Model and Calibration Uncertainties

Camera model uncertainties originate from imperfect knowledge of intrinsic parameters, lens distortion, extrinsic pose, and the scale factor that converts image motion into physical displacement. In monocular bridge applications, the effective scale factor S is often obtained from known target dimensions or structural geometry at the measurement location [18,19]. Errors in these quantities, together with pixel localization errors in calibration images, lead to uncertainty in S that directly scales displacement estimates. Field implementations have shown that adopting measurement strategies that minimize camera-to-target distance, together with careful on-site calibration, can significantly reduce this component [25,41,44,48].

Intrinsic calibration (focal length, principal point, skew, pixel size, and distortion coefficients) is typically performed using planar calibration patterns and bundle adjustment. The covariance of estimated parameters, which reflects uncertainty due to image noise, pattern placement, and numerical conditioning, is a primary input to metrology-grade uncertainty propagation. Recent studies have begun to explicitly quantify calibration

uncertainty and propagate it through pinhole models to displacement, for example, by perturbing intrinsics within their covariance and evaluating the resulting variation in reconstructed motion [19,40,41]. For stereo and 3D setups, extrinsic calibration (relative pose between cameras) adds another layer of uncertainty that affects triangulation accuracy and becomes dominant at long standoff distances [39,40].

Lens distortion modeling contributes additional uncertainty. Standard polynomial models (radial and tangential distortion) may be insufficient for wide-angle or thermally varying optics. Residual reprojection error after calibration is often used as an empirical indicator of unmodeled distortion and is either folded into pixel-level localization uncertainty or treated as a separate component [18,37]. Temperature-induced changes in intrinsics and distortion were observed in long-term monitoring, suggesting that static calibration may be inadequate for multi-season deployments [25,41].

3.2. Image-Level and Subpixel Localization Uncertainties

Image-level uncertainties include sensor noise, quantization, blur, and interpolation effects that limit the precision of subpixel localization. Theoretical analyses in the DIC and image-registration literature show that the variance of estimated subpixel shifts depends on image noise, local intensity gradients, subset size, and the interpolation scheme used to refine correlation peaks [36–38]. In practice, displacement resolution on the order of 10^{-3} to 10^{-2} pixels can be achieved in high-quality laboratory conditions, but this degrades in field environments due to lower signal-to-noise ratios and motion blur [30,37].

For template-matching pipelines, upsampled cross-correlation or Taylor-approximation schemes are used to locate correlation peaks at subpixel precision. Approaches based on modified Taylor approximation and robust object search were shown to attain high subpixel accuracy in dynamic tests, but their residual errors still exhibit dependence on texture richness and template size [18,24]. In feature-tracking pipelines, KLT and other optical-flow algorithms approximate local motion via spatial gradients, and their localization uncertainty increases for low-texture regions, small displacements, and strong illumination variations [29,31]. Sparse- and dense-flow methods both accumulate interpolation error over time, producing drift if not regularly re-referenced.

Microscopic vision-based sensors and high-resolution close-range setups illustrate the lower bound of image-level uncertainty. In such systems, subpixel localization errors can be translated into nanometer- to micrometer-level displacement uncertainty, but only when imaging conditions (illumination, focus, stability) are tightly controlled [41]. The same theoretical tools are applicable to bridge-scale systems, but with larger image noise, lower spatial frequencies, and additional environmental perturbations.

3.3. Temporal Sampling and Rolling-Shutter Effects

Temporal uncertainties arise from finite frame rates and clock jitter, which affect sampling timing, synchronization, and temporal interpolation. In dynamic measurements, finite sampling can lead to aliasing and interpolation errors in displacement time series, especially when modal frequencies approach the Nyquist limit. By contrast, exposure time and rolling-shutter readout also affect image formation, contributing to motion blur and row-dependent geometric distortion, which in turn degrade subpixel localization accuracy and increase image-level uncertainty [19,25]. Frame timing jitter introduces random variations in sampling instants; when combined with high structural velocities, this produces displacement errors that can be modeled by sensitivity to time, $\partial d/\partial t$, multiplied by timing uncertainty. Time synchronization between cameras and reference sensors (e.g., accelerometers, LVDTs) is another source of temporal error that affects validation and

sensor fusion; dedicated time-synchronized vision sensor networks explicitly address this through hardware and protocol design [44].

Exposure time is a key image-formation parameter because it directly controls motion blur. For harmonic motion, long exposures integrate displacement over part of the vibration cycle and can bias subpixel estimates toward mid-exposure positions. Studies on long-span bridges and high-speed laboratory tests have shown that reducing exposure time relative to the vibration period, together with gain adjustment, can significantly reduce blur-induced bias and variance, although often at the expense of increased imaging noise [18,25,48]. Rolling-shutter readout also affects image-level localization by introducing row-wise acquisition time offsets within each frame, which can produce row-dependent geometric distortion when structural motion or camera motion is present. Although many SHM studies use global-shutter cameras, low-cost and embedded systems may rely on rolling-shutter sensors, making this effect non-negligible for high-frequency vibrations or moving-camera measurements. In such cases, rolling-shutter geometry can be modeled explicitly, and its impact on displacement can be treated as an additional image-formation and geometric uncertainty contribution, with temporal coupling through the readout timing [44,48].

3.4. Environmental and Operational Variability

Environmental and operational variabilities include illumination changes, adverse weather, temperature variations, and atmospheric turbulence. Illumination changes (sunlight, shadows, reflections, night-time lighting) affect image contrast and can induce apparent changes in texture, leading to tracking errors and decorrelation in DIC or template matching [25,30,31]. Algorithms that incorporate spatiotemporal context and illumination-robust features explicitly aim to mitigate this source of uncertainty by stabilizing tracking under varying light conditions [31,32]. Night-time monitoring introduces additional challenges; systems that switch to infrared imaging or employ retro-reflective markers rely on thermal or reflectivity contrast, with different noise characteristics and calibration requirements [4].

Weather conditions such as rain, fog, and snow reduce visibility and contrast, increasing image noise and causing partial occlusions. Temperature gradients in the air column between the camera and the structure produce optical turbulence (heat haze), which manifests as image jitter, here defined as apparent random feature-location fluctuations or image-domain displacements not caused by true structural motion, together with distortion of image features, particularly at long standoff distances over pavement or water surfaces. Heat-haze-induced errors were shown to significantly degrade displacement measurement accuracy for bridges; mitigation techniques based on dual-tree complex wavelet transforms and robust feature extraction can correct a substantial portion of this error, but residual uncertainty remains and must be accounted for in the measurement budget [19,49]. Recent work has also proposed mitigation methods for optical-turbulence-induced errors and optimal target design, demonstrating reductions in displacement error through appropriate marker patterns and image processing [50].

Operational factors such as traffic, construction activity, and crane operations introduce transient occlusions and motion in the background. These can cause temporary tracking failures, template drift, or incorrect feature associations, particularly in targetless pipelines. Robust monitoring strategies, therefore, include mechanisms for automatic detection and exclusion of corrupted frames, as well as adaptive ROI updates that respond to changing operational conditions [30,32,51].

3.5. Camera Platform Motion and Ego-Motion

Camera platform motion includes vibration of nominally fixed mounts and deliberate or uncontrolled movement in moving-camera systems. Even when mounted on bridge piers or abutments, cameras can experience vibrations and low-frequency drift due to wind, traffic, or structural creep. In markerless vision sensors for bridge monitoring, this effect is mitigated by referencing stationary background objects and subtracting their apparent motion from structural feature motion; any residual inconsistency contributes to uncertainty in camera-motion compensation [19]. The variance in the apparent motion of background features, under the assumption of structural rest, provides an empirical estimate of platform-motion-induced uncertainty.

For moving cameras on decks or UAVs, ego-motion is estimated using IMU, GNSS/RTK, and visual odometry. The covariance of pose estimates, together with uncertainties in camera-to-IMU extrinsics, propagate through projection models into structural displacement estimates. UAV-based displacement measurement systems, for example, must contend with platform vibration, gust-induced motion, and limited GNSS accuracy, all of which affect the stability of the reconstructed structural motion [34,45]. Reference-free ego-motion compensation methods, which directly estimate relative deformations between structural points while canceling rigid-body motion, reduce reliance on external references but still inherit uncertainty from image-based motion estimation [47].

Hybrid inertial–vision systems explicitly model camera motion as a state variable in a Kalman filter, with IMU data providing high-rate pose updates and vision measurements correcting drift [44,46]. In such systems, the uncertainty of displacement estimates is governed by both sensor noise characteristics and process-model assumptions, and can be extracted from the filter covariance.

3.6. Algorithmic and Model-Form Uncertainties

Algorithmic uncertainties arise from the limitations and failure modes of tracking, correlation, and learning algorithms. By model-form uncertainty, we mean the uncertainty associated with the assumptions and simplifications of the adopted measurement or system model, i.e., the mismatch between the real physical/measurement process and its mathematical representation. Feature detection and matching can fail due to repetitive patterns, low contrast, or occlusions; KLT trackers can drift when linear motion assumptions break down; and template matching can converge to local maxima when the appearance of the template changes gradually (template drift). Experimental studies comparing multiple tracking algorithms for structural displacement monitoring have shown that different algorithms produce varying sensitivities to motion amplitude, noise, and interference, even under the same image data [24,29]. Advanced matchers that incorporate robust loss functions and multi-scale strategies reduce, but do not eliminate, such uncertainties.

Model-form uncertainty includes errors due to simplified measurement models, such as assuming planar motion with a single scale factor when perspective effects and out-of-plane motion are non-negligible. In tall structures or inclined members, unmodeled out-of-plane displacement can project into the measured in-plane direction, introducing systematic bias [25,40]. Simplified dynamic models used in state-space fusion (e.g., single-degree-of-freedom assumptions) also contribute to model-form uncertainty when actual structural behavior deviates from these assumptions.

Deep-learning-based pipelines can introduce both random-like variability and systematic effects, particularly when feature extraction, enhancement, or compensation networks are trained on datasets that do not adequately represent field conditions. In practice, domain shift, training-set bias, and model mismatch may produce systematic displacement errors (bias) that should be identified and, where possible, compensated before

residual uncertainty is quantified in a metrology-oriented framework. Terms such as epistemic and aleatoric uncertainty are widely used in the machine-learning literature to describe model-related and data-related predictive uncertainty, respectively; however, these categories do not directly replace metrological uncertainty concepts. For traceable displacement measurement, probabilistic deep-learning outputs must therefore be calibrated and validated against physical reference measurements, with explicit treatment of bias correction, residual uncertainty, and reporting conventions compatible with measurement uncertainty frameworks [34,35].

3.7. Reference and Ground-Truth Sensor Uncertainties

Reference measurements are used not only to validate vision-based displacement estimates but also to calibrate or tune measurement models and, in some cases, to identify and compensate systematic effects (e.g., scale bias, drift, or model mismatch). Common references include LVDTs, laser displacement sensors, accelerometers (with displacement reconstructed via double integration or filtering), and total stations. Each of these devices has its own calibration and uncertainty budget. For example, LVDTs and laser sensors typically provide manufacturer-specified accuracy and linearity, which must be translated into standard uncertainty for comparison; accelerometer-based displacement reconstructions include additional uncertainty from bias drift, numerical integration, and model assumptions [4,44]. Consequently, when reference measurements are used for validation, calibration, or bias compensation, their uncertainty propagates into the overall uncertainty of the corrected or compared vision-based displacement estimate.

In many SHM studies, vision-based displacement accuracy is reported as RMSE relative to a reference sensor, without explicitly accounting for the reference sensor's uncertainty. From a metrological standpoint, the reference-sensor uncertainty should either be negligible in comparison or incorporated into the combined uncertainty when comparing or calibrating systems. Some recent metrology-oriented studies explicitly account for both the uncertainty of reference sensors and the uncertainties of calibration artifacts and processing algorithms when evaluating vision-based tracking systems for scaled masonry models [26,27]. This practice enables traceable performance claims and provides a template for future SHM-focused uncertainty budgets.

Overall, the taxonomy above shows that vision-based displacement measurement uncertainty is multifaceted and strongly context-dependent. Camera modeling and calibration, image-level noise and subpixel localization, temporal sampling, environmental variability, platform motion, algorithmic behavior, and reference-sensor characteristics all contribute to the combined uncertainty, but their relative importance changes with deployment conditions (e.g., laboratory vs. field, fixed vs. moving camera, short vs. long standoff distance) [52]. In practice, a metrology-oriented assessment requires not only identifying these components but also discussing their typical order of magnitude and their contribution to the overall uncertainty budget. Subsequent sections build on this taxonomy to synthesize UQ methods and uncertainty-budget formulations, while the discussion section provides a quantitative comparison of reported performance and uncertainty reporting practices in the literature. In practice, the dominant uncertainty contributors depend on the deployment regime. For short-range fixed-camera and laboratory setups, subpixel localization, calibration, and scale-factor uncertainty are often the leading contributors. For long-range bridge monitoring, environmental effects such as illumination variation, heat haze, optical turbulence, and visibility loss can exceed image-level noise and become dominant. In moving-camera or UAV-based measurements, pose estimation, ego-motion compensation, synchronization, and platform vibration are typically the limiting factors. Therefore, a practical uncertainty budget should not assign equal importance to all compo-

nents, but should prioritize the contributors that dominate under the specific measurement geometry, environmental condition, and SHM objective.

4. Uncertainty Quantification Methods

The previous section identified where uncertainty enters the vision-based measurement chain. The purpose of this section is different: it reviews how those uncertainty sources are quantified, propagated, and reported in the literature. The emphasis is therefore on UQ methodology rather than on the physical origin of the uncertainty. For this reason, the section is organized around analytical propagation, Monte Carlo propagation, DIC-specific uncertainty budgets, resampling methods, and probabilistic deep learning, each of which provides a different route for translating the uncertainty sources identified in Section 3 into reported uncertainty statements. The uncertainty taxonomy of Section 3 identifies numerous interacting sources of uncertainty in vision-based displacement measurement. This section focuses on methods used to quantify and propagate these uncertainties, with emphasis on approaches compatible with metrological standards and those demonstrated in SHM-oriented vision applications.

4.1. GUM-Aligned Analytical Propagation

The GUM provides a general framework for evaluating measurement uncertainty [53–55] through linearization of the measurement model and propagation of variances and covariances of input quantities. In a typical vision-based displacement application, the measurand d is modeled as a function of input quantities $\mathbf{x} = (S, \Delta p, \mathbf{K}, \boldsymbol{\theta}_{\text{dist}}, \ldots)$ as discussed in Sections 2 and 3. Assuming differentiability and small perturbations, the combined standard uncertainty is obtained as

$$u_c^2(d) = \sum_i c_i^2 u^2(x_i) + 2 \sum_{i \neq j} c_i c_j \operatorname{cov}(x_i, x_j),$$

where $c_i = \partial d / \partial x_i$ are sensitivity coefficients. In practice, the dominant GUM-compatible components [53] in vision-based SHM systems include the scale factor, intrinsic and extrinsic calibration parameters, and subpixel localization variance [19,40,41]. Sensitivity coefficients can be derived analytically for simple pinhole-plus-scale models, or numerically by finite differences for more complex geometric relationships.

Several works implicitly follow GUM principles by constructing uncertainty budgets that separate calibration errors, image registration uncertainties, and environmental effects. For example, close-range and microscopic vision-based displacement sensors analyze the contributions of calibration uncertainty, interpolation error, and mechanical alignment to overall resolution and accuracy, and report combined standard and expanded uncertainties for the measurand [41]. Photogrammetric 3D vibration measurements similarly propagate calibration and triangulation uncertainties to 3D displacement, sometimes including correlations among camera parameters estimated by bundle adjustment [39,40]. Metrology-oriented tracking systems for scaled masonry models explicitly treat reference artifact uncertainties, camera calibration uncertainties, and marker localization noise as Type B or Type A components in a structured budget [26,27].

The main advantages of GUM-aligned analytical propagation are transparency and computational efficiency. Once the measurement model and sensitivities are established, evaluating $u_c(d)$ and the expanded uncertainty $U = k\,u_c(d)$ is inexpensive and can be integrated in real-time displacement monitoring [19,48]. Limitations arise when measurement models are highly nonlinear, non-differentiable, or involve complex algorithms such as dense optical flow, phase-based motion magnification, or deep networks, where linearization may be inaccurate and sensitivities difficult to obtain.

4.2. Monte Carlo Propagation

For complex or strongly nonlinear measurement models, Monte Carlo (MC) methods provide a practical alternative to analytical propagation. MC-based uncertainty evaluation, as formalized in GUM Supplement 1 [56], proceeds by sampling input quantities from their joint probability distributions, propagating each realization through the full measurement model, and estimating the distribution of the measurand from the ensemble of outputs. In the context of vision-based displacement, this can be implemented by sampling calibration parameters from their covariance matrices, adding synthetic noise consistent with image-level variance, perturbing environmental parameters, and running the entire displacement pipeline (calibration, tracking, pixel-to-metric conversion) for each realization [19,46,49].

MC propagation was applied in several ways. In 3D hybrid inertial–vision systems, MC sampling of IMU biases, camera intrinsics, and extrinsics was used to assess how these uncertainties affect reconstructed 3D displacements, providing a more accurate picture than linearized GUM approaches when coupling between pose and displacement is strong [46]. For long-span bridge monitoring under heat haze and optical turbulence, MC-based virtual experiments were used to emulate different turbulence intensities, marker patterns, and imaging conditions, producing empirical distributions of displacement error under realistic field scenarios [49,50]. Studies in other metrology domains show that MC-derived uncertainties are often consistent with or slightly more conservative than GUM estimates, particularly when probability density functions are non-Gaussian or when input distributions are bounded or skewed [54].

The main drawback of MC propagation is computational cost. End-to-end vision pipelines, especially those using dense optical flow or deep networks, can be expensive to evaluate many times. Practical implementations, therefore, often use reduced-order models, surrogate approximations, or partial MC (sampling only the most influential inputs) to limit computational burden. Sequential Monte Carlo or particle-filter-based approaches have also been explored to capture time-varying displacement and its uncertainty in sensor networks, though mostly in the context of GNSS data [57]. For SHM, MC is best suited for offline design and calibration of vision-based systems and for establishing reference uncertainty budgets, while real-time monitoring relies on precomputed sensitivities and simpler online updates.

4.3. DIC-Specific Uncertainty Budgets

DIC has a long tradition of dedicated uncertainty analysis and uncertainty budgets, which, although often developed in materials-testing contexts, are directly relevant to structural applications. Analytical and experimental studies decompose DIC displacement uncertainty into contributions from image noise, speckle pattern characteristics, subset size, interpolation schemes, and out-of-plane motion [36–38]. For example, expressions for the standard deviation of displacement estimates as a function of signal-to-noise ratio, subset size, and speckle size are widely used to predict achievable resolution and to guide speckle and imaging design.

Global DIC and photometric DIC frameworks expand these budgets by including camera model uncertainties and regularization effects. In stereo DIC, uncertainties in intrinsics and extrinsics are propagated into 3D displacement and strain via the triangulation model, with calibration covariance matrices providing the necessary input statistics [39]. Experiments comparing measured displacements and strains with known analytical fields or with high-precision references validate these budgets and demonstrate consistency between predicted and observed uncertainties [37,38]. In civil-structural contexts, such error budgets were adapted to assess uncertainty in DIC-based strain and crack opening

measurements on concrete and masonry components, and to compare DIC with traditional mechanical gauges.

4.4. Bootstrap and Resampling Methods

Bootstrap and resampling methods offer a nonparametric route to UQ when the statistical structure of errors is complex or unknown. In vision-based displacement, bootstrap approaches were proposed for two main purposes: (i) quantifying uncertainty in estimated displacements by resampling features, image patches, or time segments, and (ii) assessing algorithmic robustness to outliers, occlusions, and template drift.

In feature-tracking pipelines, bootstrap can be implemented by resampling subsets of tracked feature points and recomputing aggregate displacement estimates or structural response measures (e.g., mode shapes, frequency response functions). The variability across resampled estimates yields empirical confidence intervals that implicitly incorporate algorithmic variability and feature-selection effects [29,32]. In dense optical-flow or DIC contexts, spatial block bootstrap of patches or subsets can be used to approximate uncertainty in full-field displacement, especially when analytical models are not available or are too simplistic [37]. Temporal bootstrap (e.g., moving-block bootstrap) over time windows can quantify uncertainty in dynamic parameters derived from displacement, such as damping ratios and modal frequencies.

Bootstrap methods have the advantage of minimal distributional assumptions, which is attractive when error distributions are heavy-tailed or when algorithmic failures produce occasional large outliers. However, they rely on approximate independence of resampled units and may underestimate uncertainty when strong spatial or temporal correlations exist. They also do not naturally incorporate Type B uncertainties, such as calibration errors or environmental effects, which must be modeled separately. In vision-based SHM, bootstrap is therefore most useful as a complement to GUM or MC-based approaches, providing empirical confidence estimates for algorithmic components of the measurement chain.

4.5. Probabilistic and Bayesian Deep Learning

Probabilistic deep learning provides tools to quantify predictive uncertainty in neural-network-based models, which are increasingly used in computer vision pipelines for SHM. Techniques such as Monte Carlo dropout, deep ensembles, and evidential networks can provide uncertainty estimates related to model uncertainty and uncertainty in the inputs and observations. In a metrology-oriented context, these outputs should be interpreted within an explicit measurement model, with a clear distinction between model-related uncertainty, input-quantity uncertainty, and any residual systematic effects after correction.

SHM-oriented work on deep-learning-based structural monitoring has explored model uncertainty to assess the reliability of damage detection and classification, but quantifying displacement uncertainty remains an emerging topic [34,35]. For example, when using deep networks to enhance resolution or suppress heat-haze distortions before classical tracking, predictive uncertainty estimates can be propagated as additional image-level uncertainty components. When networks directly output displacement or modal parameters, predictive distributions can be interpreted as probabilistic displacement estimates, provided that the training labels and calibration procedures are themselves traceable and accompanied by known uncertainty.

A key challenge is calibrating the relationship between network-predicted uncertainty and physical displacement uncertainty. Without careful calibration using reference sensors with known uncertainty, deep models may be over- or under-confident, making their uncertainty outputs difficult to integrate into metrology-grade budgets. Research in probabilistic deep learning suggests post-hoc calibration techniques (e.g., temperature scaling, isotonic

regression) and loss functions targeted at reliable uncertainty estimation, which could be adapted to displacement-related tasks. For vision-based SHM, probabilistic deep learning is therefore best viewed as a way to identify low-confidence conditions, flag potentially unreliable measurements, and augment classical UQ methods, rather than as a complete replacement for GUM- or MC-based budgets [58,59].

4.6. Comparative Summary for SHM Practice

This section does not introduce a new experimental method; rather, it proposes a reporting and uncertainty-budget template synthesized from the reviewed literature and from general metrological principles. Its purpose is to make explicit how the uncertainty sources and UQ methods discussed in Sections 3 and 4 can be assembled into a practical structure for SHM reporting. From the standpoint of SHM practice, GUM-aligned analytical propagation remains the most suitable framework for metrology-grade UQ in vision-based displacement measurement, especially for marker-based and simple geometric configurations where measurement models and sensitivities can be explicitly derived. MC methods are essential for complex, nonlinear pipelines, such as 3D hybrid inertial–vision systems or long-range measurements under optical turbulence, but are best used offline due to computational cost. DIC error budgets illustrate how detailed, domain-specific analyses can be embedded within a GUM-like framework and provide useful analogies for other image-based methods.

Bootstrap and resampling methods are valuable for capturing algorithmic uncertainty and robustness, particularly in feature tracking and optical flow, where analytical models are lacking [60]. Probabilistic deep learning brings complementary capabilities for uncertainty-aware feature extraction and enhancement, but its outputs must be carefully calibrated and integrated with classical metrology to support safety-critical decisions. The next section builds on these methods to propose a metrology-grade uncertainty budget template tailored to vision-based displacement measurement in civil SHM.

5. Metrology-Grade Uncertainty Budget Template

This section provides a metrology-oriented template for constructing an uncertainty budget for vision-based displacement measurements. The template is framed around a representative case: a monocular, fixed-camera, marker-based system measuring vertical displacement of a point on a bridge, but it can be adapted to targetless, DIC, stereo, and moving-camera pipelines.

5.1. Measurand and Measurement Model

Consider the vertical displacement d of a point on the structure, along a global axis, between an initial time t_0 and time t_1. The displacement is obtained by tracking a feature or marker in the image and converting the measured subpixel shift into metric units using calibrated geometry. A simplified measurement model is

$$d = S\,\Delta p + f(\Delta p; \mathbf{K}, \boldsymbol{\theta}_{\text{dist}}, \mathbf{R}, \mathbf{t}, \mathbf{x}) + \epsilon, \tag{8}$$

where S is an effective scale factor and $f(\cdot)$ denotes a geometric correction term accounting for perspective, distortion, and pose effects; it is not, in general, a constant offset, but a configuration-dependent correction to the scale-based approximation and lens distortion based on camera intrinsics $\mathbf{K}$, distortion parameters $\boldsymbol{\theta}_{\text{dist}}$, extrinsics $(\mathbf{R}, \mathbf{t})$, and nominal 3D coordinates $\mathbf{x}$ of the point, and ϵ denotes residual errors not captured by the model [19,40,41]. For many bridge applications with small out-of-plane motion and limited field of view, $f(\cdot)$ can be approximated as a bias term or neglected within a defined validity region, but for metrology-grade analysis, it is preferable to retain its dependence on calibration parameters and geometry.

Link Between the Generic Mapping, Level 2 Linearization, and the Practical Template

In general, the mapping stage is written as $d = g(\Delta\mathbf{p}; \boldsymbol{\theta}_g)$, where $\boldsymbol{\theta}_g$ collects the geometric parameters (camera model, pose/extrinsics, reference geometry, and calibration terms). The practical measurement model in Equation (8) can be interpreted as an implementation-oriented decomposition of this generic mapping into a dominant scale-based contribution, a geometry-dependent correction, and a residual term, namely

$$g(\Delta\mathbf{p}; \boldsymbol{\theta}_g) = S\Delta p + f(\Delta\mathbf{p}; K, \boldsymbol{\theta}_{dist}, R, t, \mathbf{x}) + \epsilon.$$

For small perturbations around a nominal configuration, the same mapping can be linearized consistently with the Level 2 formulation introduced in Equation (6), yielding

$$d \approx d_0 + \mathbf{J}(\boldsymbol{\theta}_g)\big|_{\boldsymbol{\xi}_0} \Delta\mathbf{p} + \epsilon, \tag{9}$$

where $\mathbf{J}(\boldsymbol{\theta}_g)\big|_{\boldsymbol{\xi}_0}$ is the local pixel-to-metric Jacobian evaluated at the nominal operating point $\boldsymbol{\xi}_0$ and converts pixel motion into metric motion along the measurand definition. Thus, Equation (8) and the linearized Level 2 form above are not separate models; rather, the latter is the first-order local approximation of the former, with the scale term, geometric correction, and residual components explicitly retained for subsequent uncertainty-budget construction.

The input quantities $\mathbf{x}$ in this model (with $\boldsymbol{\theta}_g$ denoting the geometric subset of inputs) typically include:

$$\mathbf{x} = \left(S, \Delta p, \mathbf{K}, \boldsymbol{\theta}_{\text{dist}}, \mathbf{R}, \mathbf{t}, \mathbf{X}, t_{\text{exp}}, t_{\text{frame}}, \ldots\right).$$

This set can be augmented by environmental parameters (temperature, turbulence indicators), platform-motion parameters, and reference-sensor characteristics when validation or fusion is involved [19,44,46].

5.2. Standard Uncertainty Components (Type A/Type B)

Table 1 summarizes standard uncertainty components for a typical vision-based displacement measurement. Each component is classified as Type A (statistical, from repeated observations) or Type B (from prior information, calibration, specifications, or models), with suggested estimation methods and typical distributions.

In deep learning-enhanced pipelines, additional components may be introduced to represent uncertainty associated with network outputs (e.g., super-resolved images, corrected heat-haze fields, or directly regressed displacements). These can be treated as Type B components if predictive uncertainty was calibrated against reference measurements, or as empirical Type A components if estimated via repeated inference under varying inputs or dropout patterns [34,35]. In our framework, deep learning (DL) and moving-camera setups do not replace the measurement model; they modify specific terms in the uncertainty that will be discussed in Section 6. DL used for enhancement (denoise/SR/dehaze) primarily affects the registration term $u(\Delta\mathbf{p})$ and the environmental term u_{env} (via reduced image degradation), while DL used for feature/ROI selection affects the algorithmic term u_{alg} (robustness and failure modes). End-to-end DL regression changes the mapping itself, hence it should be treated as a new model for $g(\cdot)$ with its own calibrated uncertainty, rather than as a drop-in replacement. For moving cameras (handheld/UAV), ego-motion, synchronization, and pose/extrinsic calibration uncertainties enter $\boldsymbol{\theta}_g$ and its covariance, and are propagated through the Level 2 Jacobian $\mathbf{J}(\boldsymbol{\theta}_g)$.

Table 1. Standard uncertainty components for a monocular vision-based displacement measurement.

Component	Type	Description	Estimation Method
Subpixel localization $u(\Delta p)$	A	Random error in subpixel shift from correlation or tracking (image noise, texture, interpolation) [36,37]	Standard deviation of repeated estimates under static or cyclic conditions; theoretical expressions from DIC/registration theory based on image SNR and gradient
Scale factor $u(S)$	B	Uncertainty in mm/pixel scale (target geometry, camera–object distance, homography) [19,41]	Propagate uncertainties of reference length and calibration pixel distances; use covariance of homography or calibration parameters
Intrinsics $u(\mathbf{K})$	B	Uncertainty in focal lengths and principal point	Use covariance matrix from calibration bundle adjustment; treat as multivariate normal
Distortion $u(\boldsymbol{\theta}_{\text{dist}})$	B	Model and parameter uncertainty in lens distortion	Residual reprojection errors; comparison of different distortion models; assign pixel-level equivalent uncertainty
Extrinsics $u(\mathbf{R}, \mathbf{t})$	B	Camera pose uncertainty relative to structure [40]	Covariance from calibration (multi-view or homography); survey instrument uncertainty if pose is measured geodetically
Out-of-plane motion u_{persp}	B	Effect of lateral/out-of-plane motion on measured in-plane displacement	Analytical sensitivity analysis using perspective geometry; bound out-of-plane motion using structural models; propagate worst-case or modeled distribution
Temporal sampling u_t	B/A	Frame period jitter, exposure time, synchronization errors [44]	Manufacturer specifications (frame timing, shutter) and measured jitter; propagate via displacement sensitivity to time (velocity)
Motion blur u_{blur}	B	Bias/variance due to finite exposure during motion	Model blur for harmonic response as function of exposure/period ratio; compare measurements at different exposures to estimate residual
Rolling shutter u_{RS}	B	Distortion due to row-wise readout	Use camera line time and expected motion to bound induced pixel error; include as additional pixel-level uncertainty
Platform motion u_{cam}	A/B	Camera vibration or ego-motion compensation residuals [19,46,47]	Apparent motion of background features under structural rest; IMU/GNSS pose covariance; variance of corrected background motion
Environmental u_{env}	B	Illumination, heat haze, weather, temperature effects [30,49,50]	Group repeated measurements by environmental condition; between-group variance; dedicated experiments varying illumination and turbulence to quantify effect
Algorithmic u_{alg}	A	Feature loss, mismatches, template drift, model-form error in tracking [29,32]	Bootstrap over features, patches, or time segments; repeated runs with varied parameters; spread of estimates under controlled conditions
Reference sensor u_{ref}	B	Uncertainty in ground-truth sensors (LVDT, LDV, accelerometers) [4,44]	Calibration certificates; GUM-style budgets; for reconstructed displacement from acceleration, include integration error and model uncertainty
Numerical u_{num}	B	Quantization and numerical precision effects	Analytical bounds from image quantization; comparison of single/double precision and interpolation schemes

5.3. Sensitivity Coefficients and Linear Propagation

For GUM-aligned analytical propagation, sensitivity coefficients $c_i = \partial d / \partial x_i$ must be determined for each input quantity x_i. For the simplified model in (8), key coefficients include

$$c_S = \frac{\partial d}{\partial S} = \Delta p, \qquad c_{\Delta p} = \frac{\partial d}{\partial \Delta p} = S,$$

and partial derivatives of $f(\cdot)$ with respect to intrinsics, extrinsics, and geometric parameters. These can often be obtained analytically for pinhole-plus-distortion models using standard photogrammetric derivatives [39,40]. For Level 0 mappings, the dominant sensitivity coefficients are usually associated with the scale factor and the measured pixel displacement. For Level 1 mappings, the sensitivity analysis must also include the parameters of the homography or affine transformation, because spatially varying scale and perspective effects can make the uncertainty position-dependent within the ROI. For Level 2 mappings, the sensitivity coefficients should be evaluated with respect to the full set of geometric parameters, including camera intrinsics, distortion parameters, pose, and reference geometry. In practical implementations, these derivatives may be obtained analytically for simple models or numerically by finite differences when the mapping function is implemented as a calibration or photogrammetric routine. This step is essential because the sensitivity coefficients determine whether calibration, image registration, reference geometry, or environmental terms dominate the combined uncertainty.

5.4. Monte Carlo Implementation of the Template

For highly nonlinear or black-box pipelines (e.g., dense optical flow with heat-haze correction, hybrid inertial–vision systems), Monte Carlo propagation offers a practical way to implement the uncertainty budget. In a GUM Supplement 1-compliant workflow, the steps are:

1. Specify probability distributions for each input quantity x_i based on Type A and Type B evaluations (e.g., normal distributions for calibration parameters, uniform distributions for bounded environmental variables).
2. Draw N independent samples $\mathbf{x}^{(k)}$ from the joint distribution of inputs, preserving known correlations (e.g., via Cholesky decomposition of covariance matrices for calibration parameters).
3. For each $\mathbf{x}^{(k)}$, run the complete vision-based measurement pipeline to obtain a displacement realization $d^{(k)}$ [46,49].
4. Estimate $u_c(d)$ as the empirical standard deviation of $\{d^{(k)}\}$ and derive coverage intervals from quantiles.

Virtual experiments and digital twins of the measurement process can be used to implement such MC-based uncertainty evaluations in a way that remains consistent with GUM and its supplements, especially when physical experiments are costly or limited [49,50,55]. For SHM, virtual experiments may include finite-element structural models coupled with synthetic image formation (including noise, blur, turbulence) and full processing chains, providing a powerful tool to validate analytical budgets and explore sensitivity to design choices.

MC results can be summarized in a table similar to GUM budgets, with MC-derived standard uncertainties replacing analytically computed ones. When the output distribution is noticeably non-Gaussian, coverage intervals should be reported explicitly (e.g., 2.5–97.5% quantiles) rather than using a fixed coverage factor k.

5.5. Expanded Uncertainty and Reporting Format

Once the combined standard uncertainty $u_c(d)$ is established, the expanded uncertainty is defined as

$$U = k\, u_c(d),$$

where k is a coverage factor chosen to achieve a desired level of confidence (e.g., $k \approx 2$ for approximately 95% coverage under near-Gaussian conditions). For non-Gaussian output distributions, k can be chosen based on the ratio of the half-width of the desired coverage interval to $u_c(d)$, or the interval itself can be reported.

For SHM applications, a recommended reporting format for a vision-based displacement measurement is:

- Measurand: definition of displacement component, spatial location, and time interval.
- Measurement model: explicit formula (e.g., (8)) and a brief description of the pipeline (calibration, tracking, pixel-to-metric conversion).
- Input quantities and uncertainty components: table of x_i, best estimates, standard uncertainties, type (A/B), distribution assumptions, and sensitivity coefficients.
- Combined uncertainty: reported $u_c(d)$ and method (analytical GUM or MC).
- Expanded uncertainty: coverage factor k and resulting U, with statement of confidence level and distributional assumptions.
- Reference comparison: if applicable, difference between vision-based and reference displacements, with both uncertainties stated.

Such structured reporting, inspired by calibration certificates and metrology standards, facilitates traceable comparison between different vision-based systems and between vision-based and traditional displacement sensors. It also supports integration of vision-based measurements into SHM decision-making frameworks that explicitly account for uncertainty, such as reliability-based assessment or risk-informed thresholding [4,27,44].

5.6. Illustrative Literature-Anchored Examples of LPU and Monte Carlo UQ

To make the uncertainty-budget template in Section 5 more operational, this subsection provides two short, literature-anchored examples: (i) a linearized uncertainty evaluation using the law of propagation of uncertainty (LPU) for a simple fixed-camera marker-based pipeline, and (ii) a Monte Carlo (MC) uncertainty evaluation for an environment-dominated case where nonlinearity and stochastic disturbances are more prominent. These examples are illustrative and are intended to show how the reporting framework can be instantiated using quantities reported in the reviewed literature [6,49,61].

A natural case for LPU is the marker-based monocular setting represented by the early robust object-search and subpixel tracking studies (e.g., [18,49]). This does not mean that LPU is limited to Level 0 formulations; rather, the scale-factor model is used here as the simplest illustrative case. The same first-order propagation framework can also be applied to Level 1 and Level 2 mappings by using the appropriate Jacobian of the adopted planar or full 3D pixel-to-metric transformation. In such cases, the displacement mapping is well approximated by a scalar scale-factor model,

$$d \approx S\,\Delta p. \tag{10}$$

This corresponds to the Level 0 mapping introduced in Section 2 when the ROI is small, camera pose is effectively fixed, and out-of-plane motion is negligible. In the reviewed summary of this study, a laboratory pixel scale of approximately 0.54 mm/pixel and subpixel-enhanced standard deviations on the order of 0.05 mm (down to 0.043 mm in one dynamic test) are reported, together with field calibration values for long-standoff measurements. These characteristics make the case suitable for a transparent LPU-style uncertainty evaluation in which the dominant inputs are S, Δp, and (when relevant) the reference-sensor contribution [18].

Under the simplified model $d = S\Delta p$, the combined standard uncertainty can be written as

$$u_c^2(d) = \Delta p^2\, u^2(S) + S^2\, u^2(\Delta p) + 2\, S\, \Delta p\, \mathrm{cov}(S, \Delta p), \tag{11}$$

with the covariance term omitted when independence is justified. In a practical reporting example based on this class of experiments, $u(\Delta p)$ may be estimated from repeated static or cyclic image measurements (Type A), while $u(S)$ may be obtained from the uncertainty of the calibration length and image-based calibration reading (Type B). If the result is compared against an LVDT or laser sensor, the reference uncertainty should be stated separately and, if non-negligible, included in the comparison uncertainty. This example illustrates why marker-based fixed-camera pipelines remain the most straightforward entry point for GUM-aligned reporting in SHM [18].

A representative case for MC-based uncertainty evaluation is the heat-haze mitigation study reviewed in this paper, where optical turbulence introduces strong apparent image jitter and the displacement error distribution is governed by environmental disturbance and mitigation performance rather than by geometric scale alone. The reviewed quantitative summary reports a baseline (no-haze) RMSE of 0.0847 mm, heat-haze RMSE values of 0.7564 mm (static) and 0.7248 mm (dynamic), and substantial RMSE reduction after the proposed mitigation method [49].

For such cases, an MC implementation is more appropriate than pure linear propagation because the effective measurement model includes image degradation, algorithmic correction, and environment-dependent residuals. A literature-anchored MC setup can therefore be formulated by treating the displacement estimate as

$$d = g\Big(\Delta p, \boldsymbol{\theta}_g, \eta_{\mathrm{env}}, \eta_{\mathrm{alg}}\Big), \tag{12}$$

where $\boldsymbol{\theta}_g$ collects geometric parameters, η_{env} represents turbulence-related image perturbations, and η_{alg} represents algorithmic residuals after correction. In a practical MC uncertainty evaluation, one samples (i) geometric parameters from calibration covariances, (ii) image-level perturbations consistent with the observed haze condition, and (iii) algorithm residual terms estimated from repeated tests or virtual experiments, then propagates these samples through the processing chain to obtain the empirical distribution of d. The standard uncertainty is then estimated from the sample standard deviation, while coverage intervals are reported from output quantiles when the distribution is non-Gaussian. This is consistent with the MC workflow described in Section 5.4 and with the use of virtual experiments for turbulence-related uncertainty studies discussed in Sections 4 and 6 [49,50].

Taken together, these two examples clarify the intended use of the template in Section 5: LPU is the preferred first choice when the mapping is simple and differentiable (e.g., marker-based fixed-camera measurements with stable geometry), whereas MC becomes necessary when environmental effects, algorithmic correction stages, or coupled moving-camera terms dominate the output uncertainty. The same reporting structure (measurand, model, inputs, uncertainty components, u_c, and coverage statement) can be retained in both cases, enabling consistent comparison across pipelines.

6. Discussion

This section synthesizes the preceding material along four dimensions: (i) measurement pipelines, (ii) uncertainty sources, (iii) UQ methods, and (iv) field robustness. Emphasis is placed on practical trade-offs for SHM and on how different choices affect metrology-grade performance.

6.1. Comparison of Measurement Pipelines

Table 2 compares the main method families discussed in Section 2 in terms of calibration demands, attainable resolution, scalability, and typical failure modes.

Table 2. Comparison of vision-based measurement pipelines for SHM.

Method Family	Strengths	Limitations	Typical SHM Use
Marker-based monocular	Simple geometry; clear scale S; robust tracking; well suited to GUM-style modeling [14,18,19,48]	Requires access for marker installation; perspective sensitivity at long standoff; marker visibility under weather/traffic	Mode shapes and deflections at key points on bridges and frames; validation experiments
Targetless feature tracking/optical flow	No markers; flexible feature choice; good for retrofitting to existing CCTV [29–32]	Sensitive to illumination, texture loss, occlusions; algorithmic failure modes harder to model; scale can be less transparent	Displacement monitoring when markers are impractical; operational bridge monitoring; multi-point response identification
DIC/stereo photogrammetry	Full-field displacement and strain; mature error budgets; strong link to metrology [36,37,39,40]	Requires surface preparation and optical access; heavy data; stereo setup complexity	Laboratory component tests; local strain and crack opening in RC/masonry; specialized field campaigns
Phase-based motion magnification + edges	Excellent for small vibrations; can visualize hidden modes; works with natural texture [42]	Calibration from phase to mm non-trivial; not ideal for large motions or strong nonlinearity; sensitive to noise	Modal identification and qualitative assessment; full-field vibration visualization; laboratory and near-field tests
Moving-camera/UAV	Flexible vantage points; can access otherwise unobservable locations; integrates with inspection tasks [19,34,45,47]	Ego-motion compensation and pose uncertainty dominate; higher operational complexity; limited long-term stability	Short-term surveys and campaign measurements on bridges; combined displacement and damage mapping
Hybrid vision–inertial/ accelerometer fusion	Combines spatial richness of vision and temporal richness of inertial sensors; enables continuous monitoring [4,44,46]	More complex modeling and calibration; requires careful fusion of covariance; sensor synchronization	Long-term bridge monitoring; 3D displacement recovery; drift-free dynamic displacement
Deep learning-enhanced	Improved robustness to illumination/heat haze; learned features and descriptors; potential for end-to-end displacement regression [31,34,35,49,50]	Black-box behavior; requires large, representative datasets; predictive uncertainty rarely calibrated to mm	Long-range bridge monitoring under adverse conditions; automated feature/ROI selection; integration with digital twins

A clear pattern emerges from the literature: pipelines with simpler geometry, such as the marker-based monocular systems described by [18] or the high-precision microscopic sensors of [41], are the most amenable to rigorous uncertainty modeling. For instance, Ref. [18] demonstrated that with controlled subpixel algorithms (OCM), standard deviations in the laboratory can be reduced to as little as 0.043 mm, approaching the quantization limits. In contrast, as pipelines increase in complexity to include targetless tracking, moving platforms, or deep learning components, the uncertainty budget becomes increasingly dominated by algorithmic and environmental terms that are harder to trace. While systems like the "Raspberry Vision" [48] show that cost-effective hardware can achieve reported field performance ($\approx$0.55 mm error), the reliance on speed-optimized template matching introduces trade-offs in resolution that must be carefully managed against environmental noise. Motion-magnification-based pipelines occupy an intermediate position in this

comparison. They are particularly valuable when the objective is not only pointwise displacement recovery but also visualization and extraction of small-amplitude vibration features over structural regions. Building-scale shaking-table applications demonstrate that, when video processing is validated against accelerometers and motion-capture systems, motion magnification can contribute to modal-parameter estimation and pre-/post-damage comparison. However, these pipelines require careful treatment of amplification factors, frequency-band selection, image noise, and the propagation of video-derived uncertainty into derived dynamic indicators [43].

6.2. Uncertainty Sources: Dominant Contributors and Interactions

Table 3 summarizes which uncertainty sources are typically dominant for different pipelines.

Table 3. Dominant uncertainty contributors by method family (qualitative assessment).

Method Family	Geometry-Related (Calibration, Scale)	Image/Temporal (Subpixel, Blur, RS)	Environment/Algorithm/ Platform
Marker-based monocular	Scale factor and camera pose; distortion modeling [19,41]	Subpixel localization; exposure; frame rate [18,37]	Moderate environmental impact; platform vibration manageable via background referencing [19]
Targetless tracking/optical flow	Scale and homography if natural features off reference plane	Subpixel flow accuracy; texture-dependent noise; temporal aliasing [29,30]	Illumination, occlusions, heat haze, algorithm failures often dominant [31,32,49]
DIC/stereo	Intrinsic/extrinsic calibration and triangulation geometry [39,40]	Well-characterized subpixel uncertainty; speckle quality; blur	Environment moderated by lab conditions; out-of-plane motion and speckle degradation in field
Phase-based + edges	Geometry secondary for relative motions; scale factor calibration critical for absolute displacements	Phase noise and filter design; blur; frame rate [42]	Highly sensitive to noise and illumination; algorithmic tuning crucial
Moving-camera/UAV	Pose calibration and ego-motion estimation dominate [46,47]	Blur and RS exacerbated by platform motion; timing errors	Wind-induced motion; GNSS quality; environment and algorithm interplay dominate [34,45]
Hybrid vision–inertial	Geometry and time alignment between camera and inertial sensors	Subpixel localization; inertial sensor noise and bias	Platform motion modeled explicitly; environment still affects vision [44,46]
Deep learning-enhanced	Same as underlying vision model; geometry rarely explicit in network	Image-level uncertainty partially reduced (denoising, SR)	Model-form and data-driven uncertainties; domain shift; environment-dependent performance [34,35,49]

As discussed in the source-specific uncertainty subsections, the dominant contributors and their typical order of magnitude depend strongly on the deployment scenario; the present subsection recalls and synthesizes those trends across method families. The experimental data reviewed here confirm that no single uncertainty source is universally dominant, and the hierarchy of errors shifts with geometry, standoff distance, platform configuration, and environmental conditions. In short-range, fixed-camera applications, geometric and subpixel terms are often primary: Refs. [36,37] report image-level noise floors on the order of 10^{-3} to 10^{-2} pixels under high-quality speckle and imaging conditions. By contrast, in long-range field monitoring, environmental effects can dominate the budget; for example, Ref. [49] reports heat-haze-induced apparent displacement errors of $\sim$0.76 mm

in static tests, much larger than the intrinsic baseline sensor noise (∼0.08 mm). Platform instability likewise becomes a major contributor in mobile systems: Ref. [46] reports an increase in in-plane RMSE from 0.35 mm (stationary camera) to 1.22 mm (moving camera with tilt compensation). These examples reinforce that field-oriented UQ should prioritize the contributors that dominate in the specific operating context, rather than assuming image-level or geometric terms are always limiting.

6.3. Uncertainty Quantification Methods in Practice

Table 4 compares the UQ methods discussed in Section 4 with respect to assumptions, computational cost, and suitability for different pipelines.

Table 4. Comparison of UQ methods for vision-based displacement.

UQ Method	Key Assumptions	Computational Cost	Best Suited for
GUM analytical propagation	Differentiable measurement model; small perturbations; approximate linearity; known distributions and covariances [19,37,39]	Low, once sensitivities known	Marker-based monocular; close-range DIC; simple stereo setups; on-line uncertainty reporting in field monitoring
Monte Carlo (GUM S1)	Input distributions known; ability to sample; full pipeline executable many times [46,49,55]	High to very high; depends on model complexity and sample size	Complex 3D hybrid systems; long-range measurements under turbulence; design and calibration studies; validation of analytical budgets
DIC-specific error budgets	DIC correlation model valid; controlled speckle and imaging conditions [36–38]	Low to moderate; formulas and limited experiments	Full-field strain and displacement in lab; component-level structural tests; reference for other pipelines
Bootstrap/resampling	Resampled units approximately representative; limited dependence	Moderate; multiple re-estimations on same data [29,32]	Algorithmic uncertainty in feature tracking/flow; confidence intervals in post-processing; robustness assessment under occlusions and noise
Probabilistic deep learning	Network outputs predictive distributions; training data representative; uncertainty calibration feasible [34,35]	Medium to high; multiple stochastic passes or ensembles	Flagging low-confidence conditions; pre-filtering or weighting measurements; integration in data-driven SHM and digital twins
Virtual experiments/digital twins	Valid simulator of measurement process; alignment between virtual and real experiments [49,50,55]	High offline; cheap online once budgets derived	Design of measurement setups; sensitivity studies; UQ for scenarios difficult to reproduce experimentally (extreme weather, long spans)

For day-to-day SHM practice, the most effective strategy is rarely a single method but rather a hybrid approach. GUM serves as the foundation for well-understood components like calibration and geometry, as employed in the microscopic budgets of [41]. When analytical models become intractable, such as when quantifying the impact of heat haze or complex platform motion, Monte Carlo simulations and virtual experiments become essential. This was demonstrated by [49], where virtual experiments allowed for the isolation and quantification of turbulence-induced errors that were otherwise inseparable from structural motion. For the algorithmic layer, particularly in feature tracking, bootstrap or resampling techniques provide a data-driven measure of variability. Finally, where

deep learning is incorporated, its predictive uncertainty should currently be treated as an auxiliary confidence indicator, calibrated empirically against reference sensors, until more rigorous metrological mappings are established.

6.4. Field Robustness and Validation: Lessons Learned

Field deployments across various bridges and full-scale structures point to a consistent set of robustness challenges that must be addressed to achieve reliable monitoring.

Illumination variability and adverse weather are perhaps the most pervasive disruptors. Studies such as [48] indicate that measurement error can nearly double (from 0.036 mm to 0.06 mm) under fluctuating light conditions, while [25] noted the difficulty of maintaining tracking integrity in outdoor environments. Although deep learning offers potential remedies, the method in [49] achieved a 95% reduction in dynamic error under heat haze using advanced filtering; however, these solutions introduce their own data-dependent uncertainties.

Controlled shaking-table studies provide an important bridge between laboratory calibration and field deployment, because they allow vision-derived responses to be compared against multiple independent measurement systems under known damage progression. In motion-magnification-based building tests, the availability of accelerometers, motion-capture references, and geometric benchmarks makes it possible to assess not only displacement agreement but also the reliability of derived modal parameters and damage-sensitive indicators [43]. Optical turbulence and heat haze represent a distinct physical limit for long-span measurements. Ref. [49] explicitly quantified the "shimmer" effect, showing that it can introduce millimeter-level jitter that mimics structural vibration. Consequently, uncertainty budgets for long-span bridges are incomplete without a specific term for refractive index variations.

Camera stability is another frequent point of failure. Even nominally fixed cameras on bridges suffer from ego-motion due to wind or traffic vibration. Background referencing, as used by [19], can mitigate this, but residual errors often remain. The challenge is exacerbated in moving-camera or UAV-based systems, where [45,46] show that rigorous pose modeling is the only way to separate platform motion from structural displacement.

Finally, the issue of long-term drift remains under-addressed. Over extended monitoring periods, mechanical creep in mounts or thermal shifts in optics can induce quasi-static apparent displacements. Most current validation studies, such as those in [44], focus on short-term dynamic agreement and often neglect the propagation of reference-sensor uncertainty, a practice that leads to overly optimistic performance claims compared to rigorous metrological benchmarks [27].

Closing the Loop: Robustness Factors as Uncertainty-Budget Inputs

To move beyond qualitative discussions, field robustness factors must be explicitly represented in the uncertainty budget. Illumination variability, haze, and platform motion should not be treated merely as operational caveats but as drivers of specific budget terms, primarily the environmental term u_{env}, the platform-motion term u_{cam}, and the algorithmic term u_{alg}.

To make these terms computable, the measurement system must log a minimal set of field indicators alongside the displacement estimates. Table 5 and Figure 10 outline how robustness factors map to budget terms and suggest practical indicators such as background jitter, RMS, or image contrast statistics that can be used to estimate variance components in real-time.

Table 5. Field robustness factors mapped to uncertainty-budget terms (Table 1) and recommended field indicators for logging and estimation.

Robustness Factor	Budget Terms Strengthened	What to Log (Minimal Practical Indicators)	How to Estimate
Illumination variability (day/night, shadows, glare)	u_{env}, u_{alg}, $u(\Delta\mathbf{p})$	ROI mean and std intensity (μ_L, σ_L); contrast (RMS); saturation/underexposure rate; gradient energy $\mathbb{E}[\|\nabla I\|]$ as trackability proxy; tracking failure rate (lost features and reinitializations)	Binning by $(\mu_L, \sigma_L, \mathbb{E}[\|\nabla I\|])$ and between-bin variance; bootstrap over frames/features
Haze/fog/rain/snow (visibility loss)	u_{env}, u_{alg}, $u(\Delta\mathbf{p})$	Visibility proxy from image (e.g., contrast attenuation, dark channel prior score); % corrupted frames; occlusion fraction; feature survival rate; SNR proxy	Binning by visibility proxy; MC with noise/contrast degradation; robustness curves (uncertainty vs. visibility)
Optical turbulence/heat haze	u_{env}, $u(\Delta\mathbf{p})$, u_{alg}	Background jitter RMS (apparent motion of static background); non-rigidity residual: $\|\Delta\mathbf{p} - \Delta\mathbf{p}_{rigid}\|$; temporal shimmer index (high-frequency intensity/flow fluctuations); optional: air temperature gradient, wind	Use static-region jitter to estimate u_{env}; binning by jitter RMS; MC using measured jitter distribution
Platform/camera motion (wind, vibration, moving camera)	u_{cam}, u_t, u_{RS}, u_{blur}, u_{alg}	Pre/post compensation background motion RMS; IMU/GNSS pose covariance (if available); rolling-shutter readout time; blur proxy (e.g., Laplacian variance); frame timestamp quality/dropped frames	Estimate u_{cam} from residual background motion; propagate pose covariance in Level 2 mapping; binning by blur/RS proxies
Occlusion/traffic/vegetation/ scene changes	u_{env}, u_{alg}, $u(\Delta\mathbf{p})$	Occlusion percentage (masking/segmentation-based); track survival rate; % inliers (RANSAC); re-detection frequency; ROI stability score	Binning by occlusion and inlier rate; bootstrap over tracks/inliers; conservative expansion of u_{alg} under high occlusion

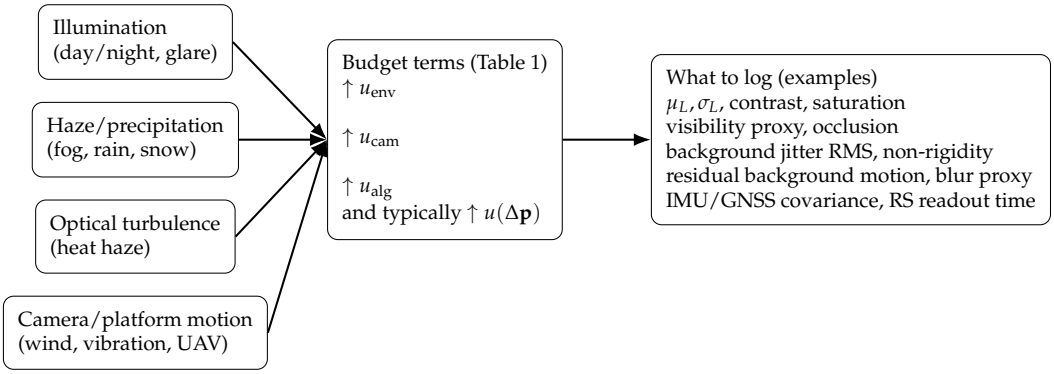


Figure 10. Robustness factors strengthen specific budget terms and require logging of measurable field indicators to make u_{env}, u_{cam}, and u_{alg} computable.

Several experimental and field-oriented studies already illustrate partial implementations of this logic. For example, time-synchronized vision sensor networks explicitly log

timing and synchronization information [44], heat haze and optical-turbulence mitigation studies use background or reference-region image fluctuations to characterize environmental disturbance [49,50], and shaking-table video-processing studies combine video measurements with accelerometers and motion-capture references to assess the reliability of dynamic indicators. Table 5 and Figure 10 generalize these practices into a compact set of indicators that can be logged routinely during field deployment.

6.5. Implications for SHM and Standardization

The synthesis provided above points toward several necessary evolutions in SHM research and practice. First, there is a critical need to balance complexity with traceability. While highly sophisticated pipelines such as those combining deep learning, optical flow, and moving platforms may improve apparent accuracy under challenging conditions, they often obscure the uncertainty budget. For safety-critical applications, simpler pipelines with well-understood error characteristics may be preferable, even if their nominal resolution is lower.

Second, the field requires standardized templates for UQ reporting. The current literature often reports simple RMSE values without context, making it difficult to separate measurement error from reference-sensor uncertainty. Community-agreed templates that explicitly list measurement models, input uncertainties, and environmental conditions modeled after the template in Section 5 would greatly enhance transparency and comparability.

Third, uncertainty bounds must become integral to SHM decision-making rather than serving as auxiliary statistics. Recent work on reliability assessment demonstrates that propagating displacement uncertainties into flexibility and safety metrics is both feasible and necessary for risk-informed management [41].

Finally, the role of virtual experiments and benchmark datasets cannot be overstated. As shown by the virtual turbulence studies in [49,50], high-fidelity simulations provide a controlled environment to validate UQ methods and benchmark algorithms against reference scenarios that are impossible to replicate perfectly in the field. These resources are essential for training the next generation of data-driven correction models.

When vision-derived displacement time histories are used for modal identification or damage assessment, uncertainty should not stop at the displacement level. The uncertainty associated with pixel localization, geometric mapping, synchronization, and environmental effects propagates into derived quantities such as natural frequencies, mode shapes, damping ratios, modal curvatures, and damage-sensitive indicators. This is consistent with vibration-based damage-identification studies showing that measurement noise and discretization can propagate into modal-curvature-based damage indicators and affect damage localization. In practice, this propagation may be evaluated through repeated time-window analysis, bootstrap resampling of displacement records or tracked points, Monte Carlo perturbation of displacement time histories within their uncertainty bounds, or covariance propagation inside the adopted modal-identification algorithm. Reporting only the agreement between identified modal parameters and reference sensors can therefore be incomplete unless the uncertainty of both the vision-derived displacement and the reference dynamic quantities is stated [62].

A critical implication of this synthesis is that improved visual tracking accuracy alone is insufficient for metrology-grade SHM. Many reported systems demonstrate strong agreement with reference sensors under specific test conditions, but they do not yet provide complete uncertainty statements, validity domains, reference-sensor uncertainty propagation, or long-term stability evidence. Consequently, the field should move from performance demonstration toward measurement qualification. This requires not only better algorithms, but also standardized calibration procedures, uncertainty-budget reporting, environmental

logging, and validation protocols that explicitly state where and under which conditions a vision-based displacement estimate can be trusted for SHM decision-making.

6.6. Quantitative Performance of Reported Methods

Table 6 summarizes the numerical performance of the specific methods reviewed in this paper in a unified format. Representative values (e.g., typical subpixel floors, environmental error magnitudes, and moving-platform penalties) are discussed earlier in the source-specific uncertainty discussion and in Section 6.2 when introducing the dominant contributors and their order of magnitude. The role of Table 6 is to consolidate these reported results study-by-study, by listing the test scenario, primary quantitative metrics (such as RMSE, maximum error, or resolution), and the extent of the uncertainty treatment provided. In this way, the table complements the qualitative taxonomy and discussion by enabling a direct cross-study comparison of performance claims and uncertainty-reporting practices.

Table 6. Quantitative performance summary for selected vision-based displacement methods.

Reference	Method/Pipeline Type	Scenario	Reported Metric(s)	Claimed Resolution/Range	Uncertainty Treatment
[18]	Marker-based monocular template matching (subpixel OCM)	Lab shaking-table tests; field test on long-span suspension bridge; fixed camera	Lab: pixel size 0.54 mm; theoretical max quantization error 0.27 mm; measured std. dev. without subpixel $\approx$ 0.14 mm, with subpixel $\approx$ 0.05 mm (earthquake test 0.043 mm).	Subpixel resolution 1/10 pixel $\Rightarrow$ $\approx$0.05 mm at 0.54 mm/px in lab; field calibration 3.01 mm/px at $\sim$300 m standoff.	Empirical standard deviations vs. LVDT; no formal propagation of calibration or LVDT uncertainty; no combined GUM budget.
[19]	Non-target monocular with background referencing (OCM)	Shaking table (1–10 mm, 1–20 Hz); railroad bridges under train loads (9–61 m, day/night)	Shaking table: "high accuracy" for 1–10 mm, 1–20 Hz; bridge tests: excellent agreement with LVDT time histories, but no tabulated RMSE; frequencies of train–bridge system identified correctly (first dominant at 1.19 Hz vs. theoretical 1.10 Hz).	Practical "millimeter-level" accuracy at tens of meters; amplitude range 1–10 mm (lab) and rail–bridge deflections under freight traffic.	Scale factor and camera-motion effects discussed qualitatively; performance evidenced via time-history and frequency comparisons; no explicit combined uncertainty budget.
[14]	Marker-based multi-point displacement (monocular; UCC/OCM)	Lab three-story frame; fixed camera	Reported good agreement with laser displacement sensors for floor displacements and mode shapes; detailed RMSE values vs. laser are not tabulated in the accessible text.	Pixel-to-metric scale from calibration; effective millimeter-level accuracy over story drifts of a few mm.	Empirical comparison with lasers and accelerometers; calibration and reference uncertainties not explicitly propagated.

Table 6. *Cont.*

Reference	Method/Pipeline Type	Scenario	Reported Metric(s)	Claimed Resolution/Range	Uncertainty Treatment
[25]	Marker-based multi-point cable-stayed footbridge monitoring	Full-scale cable-stayed footbridge; outdoor field	Deck and cable displacements of "order of a few centimeters"; modal frequencies of cables: e.g., 1.66 Hz vs. 1.63 Hz (accelerometer) and 2.10 Hz vs. 2.12 Hz; no explicit displacement RMSE table vs. accelerometers.	Centimeter-level dynamic deformation under pedestrian loading; cable frequency shifts up to ∼9% under heavy loads.	Field validation via frequency and time–frequency comparison with accelerometers; reference-sensor uncertainty and geometric mapping errors not decomposed.
[37]	2D DIC computational methods review (Zhao 2019)	Generic (materials testing, not SHM-specific)	Theoretical std. dev. of displacement σ_d on the order of 10^{-3}–10^{-2} px under good speckle and SNR; influence of subset size, interpolation, and speckle statistics quantified.	Achievable in-plane displacement resolution down to $\sim 10^{-3}$ px in high-SNR lab conditions; associated strain resolution in the $\sim 10^{-4}$ range for suitable setups.	Provides analytical and experimental error models; decomposes noise, interpolation, subset, and speckle effects; aligned with GUM-style thinking but not tied to one SHM test-bed.
[36]	DIC displacement accuracy estimation	Lab tests with imposed displacements	RMSE/std. dev. in px as functions of subset size, speckle quality, and SNR; typical noise levels of order 10^{-3}–10^{-2} px for well-designed patterns.	Micrometer-level resolution is achievable with appropriate optics and speckle for typical gauge lengths.	Detailed DIC error budget separating acquisition noise, interpolation, and algorithmic components; strong metrology orientation.
[38]	DIC accuracy: influence of speckle and setup	Similar to above	Displacement and strain error vs. speckle type and recording setup; identifies combinations that degrade accuracy.	Guidelines for speckle/imaging design to achieve target displacement and strain resolution (e.g., micrometer-level).	DIC-specific uncertainty formulations, focusing on speckle and imaging parameters, complement the above accuracy-estimation paper.
[46]	Hybrid inertial–vision 3D displacement	Indoor simulated bridge-like structure; moving camera compensated by calibration target and tilt sensor	At 9.75 m distance: with a stationary camera, mean in-plane RMSE $\approx$ 0.35 mm and out-of-plane RMSE $\approx$ 1.45 mm; with a moving camera + tilt sensor, in-plane RMSE increases to $\approx$1.22 mm and out-of-plane RMSE to $\approx$3.35 mm.	Static test displacements up to 50.8 mm in each axis; effective resolution better than 1 mm (in-plane) at ∼10 m range under ideal conditions.	Explicit tabulated RMSE in X/Y/Z for both stationary and moving camera; covariance of calibration and IMU/tilt used in the algorithm, but no fully GUM-formatted budget; Monte Carlo-style evaluation for some components.

Table 6. *Cont.*

Reference	Method/Pipeline Type	Scenario	Reported Metric(s)	Claimed Resolution/Range	Uncertainty Treatment
[41]	Microscopic vision displacement sensing (MISS, optical-disc pits)	Close-range microscopic lab setup; LVDT reference	Resolution improved from prior ~ 3 μm to ~ 20 nm using optical-disc pits and refined processing; measurements closely match LVDT across the available range, with sub-micrometer errors.	Resolution ≈ 20 nm; measuring range approaching the sensor's physical size (mm-scale travel) with linearity error kept small (few 0.01% level).	Metrology-style treatment: compensation for pixel drift and manufacturing errors; explicit separation of resolution, linearity, and repeatability; close to a GUM-consistent budget, though not framed in full GUM notation.
[48]	Real-time long-span bridge displacement with compact vision system (Raspberry Vision)	Lab monitor tests + two operational suspension bridges; low-cost Raspberry Pi-based hardware	Screen tests: RMSE ≤ 0.036 mm in ideal Case 1; $\leq$0.036 mm for 80% target completeness (Case 2) when $SRST \leq 0.7$; $\leq$0.06 mm under illumination variation and $\leq$0.07 mm with 30% salt-and-pepper noise. Field bridges: significantly improved accuracy vs. the previous system, with maximum error reduced to $\approx$0.55 mm in summarized comparison.	Edge-based template of 160×160 px; effective resolution of order 10^{-2}–10^{-1} mm in controlled lab conditions; bridge-scale resolution in the millimeter range at tens of meters.	Focus on RMSE vs. true motion (monitor) and vs. previous system; calibration and scale-factor effects discussed; no full GUM table, and reference-sensor uncertainties in the field are not decomposed.
[49]	Heat-haze error mitigation for bridge displacement	Lab bridge model with man-made heat haze; LDS reference	Baseline (no haze): RMSE vs. LDS 0.0847 mm. Under heat haze: static test RMSE 0.7564 mm, dynamic test RMSE 0.7248 mm. Proposed DT-CWT+SURF mitigation corrects 61.05% of static heat-haze-induced error and 95.31% of dynamic error (RMSE reduction).	Target ROI 250×250 px; millimeter-level apparent jitter from heat haze at lab standoff; residual RMSE after mitigation $\approx$ 0.29 mm (static) and $\approx$0.03 mm (dynamic) relative to LDS.	Uses RMSE relative to LDS to quantify error and mitigation ratio; heat haze is treated as an additional random disturbance; no explicit calibration or LDS uncertainty in a combined budget, but mitigation performance is numerically characterized.

Table 6. *Cont.*

Reference	Method/Pipeline Type	Scenario	Reported Metric(s)	Claimed Resolution/Range	Uncertainty Treatment
[42]	Video motion magnification + subpixel edge detection (PBMM + edges)	Lab beam tests; full-field displacement vs. plastic optical fiber (POF) reference sensors	Across three cases and three beam positions, RMSE between the proposed method and POF is typically $\lesssim$ 0.04–0.08 mm, while RMSE of "original video" (no MM/edge refinement) is larger (up to $\sim$0.16–0.20 mm).	Sub-mm resolution ($\sim$0.01–0.1 mm) over multi-mm dynamic displacements; full-field maps along beam length (0–800 mm).	RMSE vs. POF sensors reported per case and position; improvement due to PBMM+edge detection quantified; POF uncertainty and camera calibration errors not propagated in a formal budget.
[43]	Advanced video-based processing and motion magnification	Two-story reinforced-concrete frame under seismic loading in shaking-table tests; low-cost high-speed video; comparison with accelerometers, passive 3D motion-capture, and 3D laser-scan benchmarking	Modal parameters and structural deformations estimated before and after damage; video-derived results compared with independent vibration and geometric reference systems	Building-scale seismic testing; suitable for low-cost dynamic assessment and damage-oriented comparison	Multi-instrument validation is provided; however, a full GUM-style uncertainty budget and explicit propagation from video displacement to modal/damage indicators are not fully formalized.

6.7. Limitations of the Review

This review has several limitations. First, it is a targeted metrology-oriented review rather than a fully systematic review with bibliometric screening; therefore, the included studies are selected to represent measurement pipelines, uncertainty sources, and UQ practices rather than to provide an exhaustive census of all vision-based SHM publications. Second, the quantitative comparison is limited by the heterogeneity of reported metrics across the literature. Many studies report RMSE or correlation against a reference sensor, but do not provide complete uncertainty budgets, reference-sensor uncertainty, or environmental metadata, which limits direct cross-study comparison. Third, some method families, such as deep-learning-based displacement estimation, moving-camera/UAV systems, and phase-based motion magnification, are developing rapidly, but their metrology-grade validation remains incomplete. Fourth, the uncertainty-budget template proposed in this review is intended as a reporting and synthesis tool; it still requires case-specific instantiation, calibration data, and experimental validation before being used as a certified procedure. These limitations motivate the standardization and benchmark-development needs discussed in Section 7.

7. Open Research Gaps and Future Directions

This section distills the most consequential open problems identified in the preceding synthesis and outlines research directions needed to achieve metrology-grade, field-robust vision-based displacement measurement for SHM. Across the literature, progress is often constrained less by the availability of algorithms than by the absence of traceable benchmarks, comprehensive UQ, and long-term field evidence under realistic operating conditions. The gaps below can be grouped into two categories. Scientific and methodological gaps concern the formulation, propagation, and validation of uncertainty, including

benchmark datasets, Jacobian-based moving-camera UQ, turbulence modeling, subpixel limits, and probabilistic deep-learning calibration. Implementation gaps concern the translation of these methods into repeatable SHM practice, including long-term drift monitoring, standardized reporting, certification procedures, environmental logging, and decision-oriented use of uncertainty. This distinction is important because algorithmic progress alone does not guarantee metrological reliability unless it is accompanied by deployable calibration, validation, and reporting procedures.

7.1. Standardized Benchmarks and Datasets

A primary barrier to rigorous comparison is the lack of standardized, publicly available datasets that reflect realistic field conditions. Many studies rely on proprietary videos, use different reference sensors, and report performance under heterogeneous environmental settings, which makes cross-study claims difficult to verify. What is needed are coordinated, multi-site benchmark campaigns that provide synchronized video, traceable reference displacement measurements (e.g., LVDT, LDV, GNSS, accelerometers), and environmental metadata such as illumination, temperature, wind, and turbulence proxies. To be broadly useful, such benchmarks should span a representative range of structure classes (short- and long-span bridges, frames, towers), standoff distances, and camera configurations (fixed, deck-mounted, UAV), while explicitly covering challenging conditions including night operation, heat haze, and adverse weather. Equally important is documentation of calibration artifacts and reference-sensor uncertainty budgets so that reported errors can be interpreted metrologically rather than as context-free RMSE values. Datasets with these properties would enable objective pipeline comparison, support training and calibration of deep models, and provide common testbeds for robustness and UQ studies.

7.2. Traceable UQ for Moving Cameras and Ego-Motion Compensation

Moving-camera and UAV-based approaches are attractive for difficult-to-access locations and for combining inspection with monitoring, yet the uncertainty modeling needed for traceable ego-motion compensation remains underdeveloped. The core research gap is the absence of unified measurement models that simultaneously account for camera pose uncertainty (from IMU/GNSS/visual odometry), pixel-level localization uncertainty, and the structural motion being estimated. In practice, pose and calibration parameters are often correlated, and those correlations must be retained when propagating uncertainty through projection and mapping; otherwise, reported confidence bounds can be systematically over-optimistic. Progress in this area requires both methodological and experimental advances: pose covariance needs to be propagated through the full pipeline to displacement while preserving parameter coupling, and the resulting models must be validated on real structures using independent references and GUM- or Monte Carlo-based uncertainty budgets. Without this step, moving-camera systems are likely to remain valuable engineering instruments, but not traceable displacement sensors.

7.3. Long-Term Drift Characterization and Compensation

Most field demonstrations span hours to weeks, while SHM often targets months to years, making long-term stability a central unanswered question. Drift can arise from camera mount creep, thermal effects on optics and sensor geometry, focus changes, gradual changes in illumination patterns, and long-term structural settlement that may confound "instrument drift" with genuine response trends. A priority is to characterize these mechanisms quantitatively and model them as time-dependent stochastic processes rather than treating them as stationary noise. Complementary to this, systems need practical self-calibration and self-check routines, such as periodic imaging of reference targets or known structural configurations, to detect drift and trigger correction or recalibration. Finally, drift

must be integrated into uncertainty budgets in a way that clearly separates short-term random uncertainty from long-term bias, with confidence bounds that remain meaningful over extended monitoring horizons. Addressing these issues is essential for performance-trend analysis and reliability assessment based on vision-derived displacement.

7.4. Optical Turbulence, Heat Haze, and Adverse Weather

Heat haze and optical turbulence are increasingly recognized as dominant uncertainty sources for long-span bridges and long standoff distances, yet most treatments remain primarily algorithmic and are rarely embedded within measurement models. Future work should connect physically interpretable turbulence descriptors to observed image jitter and distortion and incorporate those relationships into uncertainty propagation, so that residual uncertainty after mitigation is quantified rather than implied. This will require controlled field experiments and high-fidelity virtual experiments capable of generating realistic turbulence effects with known ground truth. Similar modeling and validation are also needed for rain, fog, snow, and mixed illumination conditions, along with a clear characterization of operational envelopes beyond which vision-based displacement becomes unreliable. A systematic treatment would allow weather-related contributions to enter GUM- or Monte Carlo-based budgets as explicit terms rather than unmodeled performance degradation.

7.5. Subpixel Localization at Extreme Range

While image registration and DIC theory provide performance limits under idealized assumptions, systematic validation at extreme ranges typical of outdoor SHM (e.g., 100–300 m and beyond) remains scarce. The key gap is empirical evidence that maps achievable displacement resolution and bias as a function of range, optics, aperture, sensor characteristics, and environment. Carefully designed experiments that sweep these variables are needed to produce design curves that practitioners can use when selecting cameras, lenses, and targets. In parallel, DIC-style error models should be extended and validated for low-texture, long-range scenes common in large civil structures, where contrast, atmospheric effects, and perspective distortions alter classical assumptions. Delivering actionable selection rules that trade off field of view, resolution, and uncertainty would reduce over-optimistic accuracy claims and improve reproducibility in long-range deployments.

7.6. Multi-Sensor Fusion and Decision-Oriented Use of Uncertainty in SHM

Hybrid systems that combine vision with accelerometers, GNSS, strain gauges, or other sensing modalities are becoming increasingly common in SHM, but uncertainty is still often handled in a sensor-by-sensor manner rather than within a unified probabilistic framework. A major opportunity is therefore to develop fusion architectures, such as Kalman, particle, or ensemble filtering, that maintain and propagate full covariance across states and sensors while explicitly accounting for correlations among measurements and between process and measurement models. Such correlations are especially important when sensors are co-located or affected by shared disturbances, such as wind, temperature variation, or platform motion. Practical adoption would be accelerated by reference implementations and open-source tools tailored to SHM workflows, with standardized outputs that report fused displacement together with combined and expanded uncertainties.

At the same time, even when uncertainty bounds are available, they are often not used systematically in SHM decision-making, leading to ad hoc alarm thresholds and inconsistent maintenance actions. Future work should therefore connect uncertainty-aware fusion outputs to decision-theoretic frameworks that translate displacement and flexibility uncertainty into probabilities of exceeding serviceability or ultimate limit states, while explicitly quantifying the trade-off between false alarms and missed detections. Such

frameworks should also support integration with digital twins and prognosis tools, so that vision-based and fused measurements can update state estimates, reliability indicators, and maintenance priorities in near-real time. Making uncertainty both fusion-aware and decision-relevant is essential if metrology-grade displacement estimates are to become operational tools for asset management rather than remain descriptive performance metrics. In addition, recent studies on steel truss bridges highlight the importance of sensor placement, sensitivity analysis, and partial-model-based damage identification for translating measured structural responses into reliable damage indicators [63,64]. Although these studies are not vision-based displacement measurement methods, they reinforce the need to connect measurement uncertainty, sensor layout, and model-based damage-identification strategies in future SHM frameworks.

7.7. Integration of Probabilistic Deep Learning with Metrology

Deep learning has become influential for feature extraction, enhancement, denoising, and even direct displacement inference, yet probabilistic outputs are rarely linked to physically meaningful displacement uncertainty. A central challenge is calibration: uncertainties predicted by ensembles, MC dropout, or heteroscedastic models must be validated against errors measured by traceable reference sensors, not merely against internal consistency metrics. Training objectives and protocols should therefore encourage calibrated uncertainty rather than minimizing average error alone, and deep models should be integrated into measurement models so that their outputs are treated as input quantities with quantified uncertainty that is compatible with GUM/Monte Carlo propagation. Achieving this would move deep learning from an effective preprocessor to a metrologically accountable component of the measurement chain.

7.8. Standardized Reporting and Certification Practices

The literature exhibits wide variability in performance reporting, ranging from simple RMSE comparisons to partial error budgets, with full GUM-style uncertainty statements remaining rare. This creates barriers to both scientific comparison and engineering adoption. Progress requires minimal reporting standards that specify the measurand definition, the measurement model, key uncertainty contributors, combined and expanded uncertainties, and the environmental and operational conditions under which measurements were obtained. Beyond reporting, the community would benefit from recommended practices or codes of practice that define acceptable procedures for design verification, load testing, and continuous SHM using vision-based displacement. In parallel, exploring certification pathways analogous to those used for conventional metrological instruments could provide an institutional bridge from research prototypes to deployable monitoring systems.

Addressing these gaps will help move vision-based displacement measurement from a promising field technology to a traceable, robust, and decision-relevant component of modern SHM systems.

8. Conclusions

Vision-based displacement and deformation measurement has evolved from laboratory proof-of-concept to a mature technology capable of supporting SHM of civil infrastructure. This review has systematically addressed the four core research questions that define metrology-grade implementation.

The dominant measurement pipelines, including marker-based template matching, targetless feature tracking, optical flow, DIC, phase-based motion magnification, moving-camera systems with ego-motion compensation, hybrid vision–inertial fusion, and emerging deep-learning-enhanced approaches, can all be understood as instantiations of a unified

projection-plus-registration measurement model. Marker-based monocular and close-range DIC systems offer the clearest path to traceable UQ due to their simpler geometry and well-characterized error models, while targetless, moving-camera, and deep-learning-based pipelines provide greater flexibility at the cost of increased algorithmic and environmental uncertainties.

A comprehensive uncertainty taxonomy reveals that no single source dominates universally. Geometric uncertainties (calibration parameters, scale factor) are primary in controlled settings, while environmental effects (illumination variation, heat haze, optical turbulence), platform motion, and algorithmic failures become limiting factors in field deployments. Temporal effects (blur, rolling shutter, frame synchronization) and reference-sensor uncertainties must also be explicitly modeled for complete budgets.

GUM-aligned analytical propagation provides the most transparent and efficient framework for metrology-grade UQ when measurement models are reasonably simple and differentiable. Monte Carlo methods, DIC-specific error budgets, bootstrap resampling, and probabilistic deep learning offer complementary capabilities for complex pipelines and algorithmic robustness assessment. The proposed uncertainty-budget template in Section 5 provides a practical starting point for implementing and reporting such analyses in SHM practice.

Field robustness evidence from bridge deployments demonstrates that millimeter-level accuracy is routinely achievable under favorable conditions, but illumination changes, heat haze, camera motion, and long-term drift pose persistent challenges. Mitigation strategies exist, but they must be accompanied by explicit UQ rather than implicit algorithmic correction.

The review identifies eight actionable research gaps spanning standardized benchmarks and datasets, traceable UQ for moving-camera systems, long-term drift characterization, optical turbulence and adverse-weather modeling, subpixel localization at extreme range, multi-sensor fusion and decision-oriented use of uncertainty, probabilistic deep learning–metrology integration, and standardized reporting and certification practices. Addressing these gaps would help move vision-based displacement measurement toward more traceable, robust, and decision-relevant SHM practice. However, routine metrology-grade implementation will require further standardization of calibration procedures, uncertainty-budget reporting, reference-sensor comparison, environmental-condition logging, and validation protocols across laboratories and field deployments.

Future SHM systems will likely combine vision-based spatial richness with inertial temporal precision, deep-learning-based robustness, and rigorous UQ within standardized reporting frameworks. This synthesis provides both the conceptual foundation and practical guidance to realize such systems.

Author Contributions: A.N.: Acted as the principal contributor to this study, leading its conceptualization, methodological design, literature synthesis, analytical framework, manuscript drafting, coordination among co-authors, and critical revision. F.P.: Contributed to the conceptual refinement of the review, methodological validation, manuscript review and editing, and technical discussion of measurement and uncertainty-related aspects. I.T.: Contributed to the technical review of vision-based measurement approaches, manuscript revision, and refinement of the methodological and application-oriented sections. M.M.: Contributed to the civil and structural engineering interpretation of the reviewed studies, with particular input on SHM practice, structural testing, and manuscript review. L.D.V.: Provided senior metrological expertise, contributed to the conceptual framing of the uncertainty-oriented perspective, and supported the critical revision of the manuscript. M.D.: Supervised the project, provided scientific guidance and strategic direction, contributed to the overall intellectual framework, and critically reviewed the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the ISPEDIA project—Innovation and efficiency for infrastructure inspection through the use of drones, artificial intelligence, and smart devices, funded by the Italian Ministry of Enterprises and Made in Italy (CUP: B69J24001350005).

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors gratefully acknowledge CESMA (Advanced Metrological and Technological Services Center) of the University of Naples Federico II and LESIM (Laboratory of Signal and Measurement Information Processing), Department of Engineering, University of Sannio, for their support. The authors also acknowledge the use of ChatGPT (GPT-5.5, OpenAI) to assist with grammar checking and rephrasing during the preparation of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ponzo, F.C.; Iacovino, C.; Ditommaso, R.; Bonano, M.; Lanari, R.; Soldovieri, F.; Cuomo, V.; Bozzano, F.; Ciampi, P.; Rompato, M. Transport infrastructure SHM using integrated SAR data and on-site vibrational acquisitions: "Ponte Della Musica–Armando Trovajoli" case study. *Appl. Sci.* **2021**, *11*, 6504. [CrossRef]
2. Selvakumaran, S.; Rossi, C.; Marinoni, A.; Webb, G.; Bennetts, J.; Barton, E.; Plank, S.; Middleton, C. Combined InSAR and Terrestrial Structural Monitoring of Bridges. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 7141–7153. [CrossRef]
3. Ye, X.W.; Dong, C.Z.; Liu, T. A review of machine vision-based structural health monitoring: Methodologies and applications. *J. Sens.* **2016**, *2016*, 7103039. [CrossRef]
4. Xiu, C.; Weng, Y.; Shi, W. Vision and vibration data fusion-based structural dynamic displacement measurement with test validation. *Sensors* **2023**, *23*, 4547. [CrossRef] [PubMed]
5. Feng, D.; Feng, M.Q. Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection—A review. *Eng. Struct.* **2018**, *156*, 105–117. [CrossRef]
6. Feng, D.; Feng, M.Q.; Ozer, E.; Fukuda, Y. A vision-based sensor for noncontact structural displacement measurement. *Sensors* **2015**, *15*, 16557–16575. [CrossRef] [PubMed]
7. Lydon, D.; Lydon, M.; Taylor, S.; Del Rincon, J.M.; Hester, D.; Brownjohn, J. Development and field testing of a vision-based displacement system using a low cost wireless action camera. *Mech. Syst. Signal Process.* **2019**, *121*, 343–358. [CrossRef]
8. Dong, C.Z.; Celik, O.; Catbas, F.N.; OBrien, E.; Taylor, S. A robust vision-based method for displacement measurement under adverse environmental factors using spatio-temporal context learning and taylor approximation. *Sensors* **2019**, *19*, 3197. [PubMed]
9. Hong, J.; Sim, S.H.; Lee, J.; Kim, H.; Cho, S.; Lee, J. Long-term displacement measurement system for bridge bearing capable of camera reposition. *Measurement* **2024**, *235*, 114921. [CrossRef]
10. Hartlieb, S.; Zeller, A.; Haist, T.; Reichardt, A.; Tarín Sauer, C.; Reichelt, S. Advanced Imaging-Based Metrology for Precise Deformation Monitoring: Railway Bridge Case Study. *Struct. Control Health Monit.* **2025**, *2025*, 5603393. [CrossRef]
11. Wu, T.; Tang, L.; Zhang, X.; Liu, Y.; Li, X.; Zhou, Z. An Improved Structural Displacement Monitoring Approach by Acceleration-Aided Tilt Camera Measurement. *Struct. Control Health Monit.* **2023**, *2023*, 6247516. [CrossRef]
12. Tan, D.; Li, J.; Hao, H.; Nie, Z. Target-free vision-based approach for modal identification of a simply-supported bridge. *Eng. Struct.* **2023**, *279*, 115586.
13. Yu, S.; Zhang, J. Machine Learning-Based Distraction-Free Method for Measuring the Optical Displacement of Long-Span Bridge Structures. *Struct. Control Health Monit.* **2024**, *2024*, 1824559. [CrossRef]
14. Feng, D.; Feng, M.Q. Vision-based multipoint displacement measurement for structural health monitoring. *Struct. Control Health Monit.* **2016**, *23*, 876–890.
15. Gesualdo, A.; Fortunato, A.; Penta, F.; Monaco, M. Structural identification of tall buildings: A reinforced concrete structure as a case study. *Case Stud. Constr. Mater.* **2021**, *15*, e00701. [CrossRef]
16. Ni, Y.; Ma, Y.; Lin, H.; Zhang, H.; Zhuo, Y. Finite element model updating of existing bridge structures based on measured data and deep learning method. *Structures* **2025**, *79*, 109403. [CrossRef]
17. Juarez-Salazar, R.; Diaz-Ramirez, V.H. Distorted pinhole camera model for tangential distortion. In *Proceedings of the Optics and Photonics for Information Processing XV*; Iftekharuddin, K.M., Awwal, A.A.S., Diaz-Ramirez, V.H., Eds.; International Society for Optics and Photonics, SPIE: Philadelphia, PA, USA, 2021; Volume 11841, p. 118410D.
18. Fukuda, Y.; Feng, M.Q.; Narita, Y.; Kaneko, S.; Tanaka, T. Vision-based displacement sensor for monitoring dynamic response using robust object search algorithm. *IEEE Sens. J.* **2013**, *13*, 4725–4732. [CrossRef]
19. Feng, M.Q.; Fukuda, Y.; Feng, D.; Mizuta, M. Nontarget vision sensor for remote measurement of bridge dynamic response. *J. Bridge Eng.* **2015**, *20*, 04015023. [CrossRef]

20. Kromanis, R.; Kripakaran, P. A multiple camera position approach for accurate displacement measurement using computer vision. *J. Civ. Struct. Health Monit.* **2021**, *11*, 661–678. [CrossRef]
21. Weng, Y.; Shan, J.; Lu, Z.; Lu, X.; Spencer, B.F., Jr. Homography-based structural displacement measurement for large structures using unmanned aerial vehicles. *Comput. Aided Civ. Infrastruct. Eng.* **2021**, *36*, 1114–1128. [CrossRef]
22. Bräuer-Burchardt, C.; Ramm, R.; Kühmstedt, P.; Notni, G. The duality of ray-based and pinhole-camera modeling and 3D measurement improvements using the ray-based model. *Sensors* **2022**, *22*, 7540. [PubMed]
23. Fauster, E. Statistical Uncertainty Analysis for Image Processing Algorithms in Metric Vision Systems. Ph.D. Thesis, Montanuniversität Leoben, Leoben, Austria, 2008.
24. Liu, B.; Zhang, D.; Guo, J.; Zhu, C. Vision-based displacement measurement sensor using modified Taylor approximation approach. *Opt. Eng.* **2016**, *55*, 114103. [CrossRef]
25. Xu, Y.; Brownjohn, J.; Kong, D. A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge. *Struct. Control Health Monit.* **2018**, *25*, e2155. [CrossRef]
26. Daponte, P.; De Vito, L.; Tudosa, I.; Neyestani, A.; Picariello, F. Development and Evaluation of a Novel Marker-Based Tracking System for 3D-Scaled Masonry Models Using DeepTag. In Proceedings of the 2024 XXXIII International Scientific Conference Electronics (ET), Sozopol, Bulgaria, 17–19 September 2024; pp. 1–5. [CrossRef]
27. Daponte, P.; De Vito, L.; Iannuzzo, A.; Monaco, M.; Neyestani, A.; Picariello, F. Low-Cost Marked Tracking Monitoring System for 3D-Scaled Masonry Models. In Proceedings of the 2024 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv), Chania, Greece, 12–14 June 2024; pp. 172–177.
28. Mittal, S.; Sadhu, A. Fiducial marker-based decentralized computer vision for structural modal identification. *J. Sound Vib.* **2025**, *612*, 119152. [CrossRef]
29. Ye, X.; Dong, C.; Liu, T. Image-based structural dynamic displacement measurement using different multi-object tracking algorithms. *Smart Struct. Syst.* **2016**, *17*, 935–956.
30. Peng, Z.; Li, J.; Hao, H. Computer vision-based displacement identification and its application to bridge condition assessment under operational conditions. *Smart Constr.* **2024**, *1*. [CrossRef]
31. Chen, S.h.; Luo, Y.p.; Liao, F.y. Computer vision-based displacement measurement using spatio-temporal context and optical flow considering illumination variation. *J. Civ. Struct. Health Monit.* **2024**, *14*, 1765–1783. [CrossRef]
32. Zhang, S.; Han, Q.; Jiang, K.; Lu, X.; Wang, G. Robust and versatile vision-based dynamic displacement monitoring of natural feature targets in large-scale structures. *J. Civ. Struct. Health Monit.* **2024**, *14*, 1819–1837. [CrossRef]
33. Won, J.; Park, J.W.; Park, K.; Yoon, H.; Moon, D.S. Non-target structural displacement measurement using reference frame-based deepflow. *Sensors* **2019**, *19*, 2992. [PubMed]
34. Neyestani, A.; Ahmed, I.; Daponte, P.; De Vito, L. Concrete Crack Detection and Segmentation in Civil Infrastructures Using UAVs and Deep Learning. In Proceedings of the 2023 7th International Conference on Internet of Things and Applications (IoT), Isfahan, Iran, 25–26 October 2023; pp. 1–6. [CrossRef]
35. Neyestani, A.; Picariello, F.; Tudosa, I.; Daponte, P.; De Vito, L. Triplet Loss-Based Concrete Crack Verification for Structural Health Monitoring and Digital Twin Applications. In Proceedings of the 2024 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), London, UK, 21–23 October 2024; pp. 837–842. [CrossRef]
36. Dupré, J.C.; Bornert, M.; Robert, L.; Wattrisse, B. Digital image correlation: Displacement accuracy estimation. In *EPJ Web of Conferences*; EDP Sciences: Paris, France, 2010; Volume 6, p. 31006.
37. Zhao, J.; Sang, Y.; Duan, F. The state of the art of two-dimensional digital image correlation computational method. *Eng. Rep.* **2019**, *1*, e12038. [CrossRef]
38. Barranger, Y.; Doumalin, P.; Dupré, J.; Germaneau, A. Digital image correlation accuracy: Influence of kind of speckle and recording setup. In *EPJ Web of Conferences*; EDP Sciences: Paris, France, 2010; Volume 6, p. 31002.
39. Fouque, R.; Bouclier, R.; Passieux, J.C.; Périé, J.N. Photometric DIC: A unified framework for global Stereo Digital Image Correlation based on the construction of textured digital twins. *J. Theor. Comput. Appl. Mech.* **2022**. [CrossRef]
40. Yang, Z.; Wang, C.; Gao, J. 3-D structural vibration measurement using digital photogrammetry. In *Proceedings of the International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII 2015)*; Atlantis Press: Paris, France, 2015; pp. 1581–1586.
41. Wu, P.; Li, W.; Zhao, X. Displacement sensing based on microscopic vision with high resolution and large measuring range. *Comput. Aided Civ. Infrastruct. Eng.* **2024**, *39*, 2840–2858. [CrossRef]
42. Duan, D.Y.; Kuang, K.; Wang, Z.C.; Sun, X.T. Video Motion Magnification and Subpixel Edge Detection-Based Full-Field Dynamic Displacement Measurement. *Struct. Control Health Monit.* **2023**, *2023*, 7904198. [CrossRef]
43. Cataldo, A.; Roselli, I.; Fioriti, V.; Saitta, F.; Colucci, A.; Tatì, A.; Ponzo, F.C.; Ditommaso, R.; Mennuti, C.; Marzani, A. Advanced video-based processing for low-cost damage assessment of buildings under seismic loading in shaking table tests. *Sensors* **2023**, *23*, 5303. [CrossRef] [PubMed]

44. Lydon, D.; Lydon, M.; Del Rincon, J.M.; Taylor, S.E.; Robinson, D.; O'Brien, E.; Catbas, F.N. Development and field testing of a time-synchronized system for multi-point displacement calculation using low-cost wireless vision-based sensors. *IEEE Sens. J.* **2018**, *18*, 9744–9754. [CrossRef]
45. Pan, D.; Dai, W.; Xing, L.; Yu, Z.; Wu, J.; Zhang, Y. UAV Vision-Based Method for Multi-Point Displacement Measurement of Bridges. *Sensors* **2025**, *26*, 240. [PubMed]
46. Zhang, X.; Zeinali, Y.; Story, B.A.; Rajan, D. Measurement of three-dimensional structural displacement using a hybrid inertial vision-based system. *Sensors* **2019**, *19*, 4083. [PubMed]
47. Wang, S.; Fillmore, T.; Lee, J.; Eick, B.A.; Spencer, B.F., Jr. Reference-free ego-motion compensation for vision-based measurement of structural deformations of miter gates. *Struct. Health Monit.* **2025**, *24*, 2193–2208.
48. Wang, M.; Xu, F.; Koo, K.Y.; Wang, P. Real-time displacement measurement for long-span bridges using a compact vision-based system with speed-optimized template matching. *Comput. Aided Civ. Infrastruct. Eng.* **2024**, *39*, 1988–2009.
49. Kong, X.; Wang, B.; Feng, D.; Yuan, C.; Gu, R.; Ren, W.; Wei, K. A Novel Method for Heat Haze-Induced Error Mitigation in Vision-Based Bridge Displacement Measurement. *Sensors* **2024**, *24*, 5151. [PubMed]
50. Huang, X.; Dai, W.; Zhang, Y.; Xing, L.; Ye, Y. A mitigation method for optical-turbulence-induced errors and optimal target design in vision-based displacement measurement. *Sensors* **2023**, *23*, 1884. [PubMed]
51. Zhan, Y.; Huang, Y.; Fan, Z.; Li, B.; An, J.; Shao, J.; Tian, Y. Computer Vision-Based Pier Settlement Displacement Measurement of a Multispan Continuous Concrete Highway Bridge under Complex Construction Environments. *Struct. Control Health Monit.* **2024**, *2024*, 1866665.
52. Tomani, C.; Gruber, S.; Erdem, M.E.; Cremers, D.; Buettner, F. Post-hoc uncertainty calibration for domain drift scenarios. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10124–10132.
53. JCGM. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*; International Organization for Standardization: Geneva, Switzerland, 2008; Volume 50, p. 134.
54. Zhussupbekov, S.S.; Khan, S.G.; Ibrayeva, L.K. The analysis of approaches to measurement uncertainty evaluation for calibration. *Inform. Autom. Pomiary W Gospod. Ochr. Środowiska* **2019**, *9*, 26–29. [CrossRef]
55. Wübbeler, G.; Marschall, M.; Kniel, K.; Heißelmann, D.; Härtig, F.; Elster, C. GUM-compliant uncertainty evaluation using virtual experiments. *Metrology* **2022**, *2*, 114–127. [CrossRef]
56. JCGM. *Evaluation of Measurement Data–Supplement 1 to the Guide to the Expression of Uncertainty of Measurement-Propagation of Distributions Using a Monte Carlo Method*; Joint Committee for Guides in Metrology (JCGM): Paris, France, 2008. [CrossRef]
57. Piriyasatit, S.; Kuruoglu, E.E.; Ozeren, M.S. Time-Varying GPS Displacement Network Modeling by Sequential Monte Carlo. *Entropy* **2024**, *26*, 342. [CrossRef] [PubMed]
58. Wang, C. Calibration in Deep Learning: A Survey of the State-of-the-Art. *arXiv* **2023**, arXiv:2308.01222.
59. Zuo, C.; Qian, J.; Feng, S.; Yin, W.; Li, Y.; Fan, P.; Han, J.; Qian, K.; Chen, Q. Deep learning in optical metrology: A review. *Light Sci. Appl.* **2022**, *11*, 39. [CrossRef] [PubMed]
60. Wang, S.; Saida, T.; Nishio, M. Optical Flow-Based Structural Anomaly Detection in Seismic Events From Video Data Combined With Computational Cost Reduction Through Deep Learning. *Struct. Control Health Monit.* **2025**, *2025*, 4702519.
61. Wen, H.; Dong, R.; Dong, P. Structural displacement measurement using deep optical flow and uncertainty analysis. *Opt. Lasers Eng.* **2024**, *181*, 108364. [CrossRef]
62. Dessi, D.; Passacantilli, F.; Venturi, A. Analysis and mitigation of uncertainties in damage identification by modal-curvature based methods. *J. Sound Vib.* **2025**, *596*, 118769.
63. Xiao, F.; Mao, Y.; Tian, G.; Chen, G.S. Partial-model-based damage identification of long-span steel truss bridge based on stiffness separation method. *Struct. Control Health Monit.* **2024**, *2024*, 5530300.
64. Mao, Y.; Xiao, F.; Tian, G.; Xiang, Y. Sensitivity analysis and sensor placement for damage identification of steel truss bridge. *Structures* **2025**, *73*, 108310. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.