Review of machine-vision based methodologies for displacement measurement in civil structures

Yan Xu[1], James M.W. Brownjohn[1]

[1]Vibration Engineering Section, College of Engineering, Mathematics and Physical Sciences, University of Exeter, UK EX4 4QF

Email: yx298@exeter.ac.uk

Abstract

Vision-based systems are promising tools for displacement measurement in civil structures, possessing advantages over traditional displacement sensors in instrumentation cost, installation efforts and measurement capacity in terms of frequency range and spatial resolution. Approximately one hundred papers to date have appeared on this subject, investigating topics like: system development and improvement, the viability on field applications and the potential for structural condition assessment. The main contribution of this paper is to present a literature review of vision-based displacement measurement, from the perspectives of methodologies and applications. Video processing procedures in this paper are summarised as a three-component framework, camera calibration, target tracking and structural displacement calculation. Methods for each component are presented in principle, with discussions about the relative advantages and limitations. Applications in the two most active fields: bridge deformation and cable vibration measurement are examined followed by a summary of field challenges observed in monitoring tests. Important gaps requiring further investigation are presented e.g. robust tracking methods, non-contact sensing and measurement accuracy evaluation in field conditions.

Keywords Vision-based system; Structural displacement; Camera calibration; Target tracking.

1 INTRODUCTION

Structural health monitoring (SHM) is aimed at providing valuable information for structural assessment and decision support for maintenance through relevant measures of structural response. Deformation is an important metric for structural condition and performance assessment for several reasons. In particular serviceability is reflected through deformation during normal operation, since extreme values and ranges indicate problems that may limit operational use, while time-varying deformation patterns constructed from discrete displacement measurements can provide a wealth of information about structure condition.

Conventional sensors like linear variable differential transformers (LVDT) require a stationary reference point for installation and direct access to monitoring structures that could be challenging on site. The global positioning systems (GPS) have the limitation of measurement accuracy (i.e. sub-centimetre [1] or centimetre level [2]) and are mostly applied for monitoring campaigns in flexible large-scale structures. Integration schemes from acceleration measurement are only feasible for short-time signals and might fail to capture the static or quasi-static components in displacement signals. Such limitations of more traditional displacement sensing technologies have driven research in non-contact optical sensing. Vision-based monitoring methods have promising features e.g. simple instrumentation and installation, operation remote from the structure and capacity for multi-point measurement using a single (camera) sensor.

Although there have been earlier optics-based methods used for monitoring civil structure deformation e.g. in the Tacoma Narrows Bridge [3] and the Tagus Bridge [4], among the earliest applications of opto-electronic vision-based continuous structural deformation monitoring using CCD (charge-coupled device) arrays was to Humber Bridge and Severn Bridge in the UK [5], [6]. Since then a number of systems have been developed and evaluated for structural deformation monitoring in high-rise buildings [7], short-span bridges [8], [9], [10], [11] and long-span bridges [12], [13], [14].

Vision-based systems offer significant potential for structural condition assessment, in particular for system identification [15–17]. In addition, deformation information has been used for finite element model calibration [18], damage detection [19] and contribution to bridge weigh-in-motion system with camera assistance for traffic monitoring [20].

Investigations have been made in system improvement in both video acquisition hardware and video processing software. The feasible video acquisition devices are expanded to include smartphone cameras [15][21], while artificial targets required in conventional systems were discarded in some recent applications under specific camera configurations [8, 15, 22]. Efficient target tracking techniques in the computer vision field have been validated in structural deformation monitoring [15], [22], [23] and the measurement results describing structural displacement have been expanded to three-dimensional [24, 25], [17], [26] and six degree of freedom (DOF) motions [11, 14].

This paper aims to present a summary of key work in the field of vision-based systems for structural displacement monitoring while highlighting the principles, advantages and shortcomings of these systems. Although previous reviews of vision-based structural monitoring exist [27–29], the contribution of this work is to provide an overview of system classifications, methodologies and applications in field monitoring.

The paper is organised as follows. The components of a vision-based system for displacement monitoring are introduced, followed by a comparison of several mature vision-based systems in application scopes in section 2. In section 3, vision-based systems are categorised based on methods of video processing, with three components in video processing procedures (i.e. camera calibration, target tracking and structural displacement calculation) reviewed in terms of principle, applications, advantages and shortcomings, respectively. In section 4, applications for bridge deformation and cable vibration measurement are reviewed followed by a discussion of measurement challenges in field applications. Finally, important gaps requiring further investigation are presented e.g. robust tracking methods, non-contact sensing and measurement accuracy evaluation in field conditions.

2 VISION-BASED DISPLACEMENT MONITORING SYSTEMS

Applying a vision-based system for structural displacement monitoring requires setting up one or more cameras in a stable location, looking at the ‘target’ contained in a structure and deriving the structural displacement through target tracking. Here the ‘target’ could be either artificial (e.g. pre-installed marker, LED lamp or planar panel with special patterns) or an existing structural feature (e.g. bolts or holes). As shown in Fig. 1, the hardware generally comprises camera, lens, laptop/portable computer with video-processing package and some accessories e.g. tripod. The video processing software is critical: its role is acquiring the video frames covering the target region, tracking the target locations in image sequences and finally transforming the target location in image to time history of structural displacement.

Systems for extracting metric information from images or videos exist in several fields as indicated in Table 1 e.g. digital image correlation (DIC) [9], [31], [32], photogrammetric techniques [33] and motion capture systems (MCS) [17, 34]. DIC is a measurement tool to extract full-field displacements or strains of a member surface in experimental solid mechanics [32, 35, 36]. Photogrammetry, originally in the production of topographic maps

[37], is expanded to include deflection monitoring of bridge structures [38]. Motion capture systems (MCS) are usually applied to capture the movements of a high degree-of-freedom skeleton structure with a number of joints (e.g. human bodies) [39].

A vision-based system for structural displacement monitoring owns its unique features, as indicated in the last row of Table 1. Researchers have performed several investigations into system development targeted at structural applications and these studies will be reviewed in terms of methodologies in the next section.

3 REVIEW OF VISION-BASED STRUCTURAL DISPLACEMENT MEASUREMENT

In this study, vision-based systems in literature are classified based on video-processing methodologies. A typical video processing software package could fit into a three-component framework shown in Fig. 2. The derived displacement data could be interpreted for bridge condition assessment.

If the monitoring campaign is only for system identification and exact vibration values [40, 41] are not required, target tracking may be the only part of the whole video processing procedure needed, but coordinate transformation might be necessary to align the image motion directions with the structural axes.

Next, the methods for camera calibration, target tracking and structural displacement calculation in literature are reviewed separately.

3.1 Camera calibration

Camera calibration concerns building the projection relationship between the 3D structural points in the structural coordinate system and the corresponding 2D points in the image plane. The determined projection transformation could be used to recover the real locations of targets in structure given the target locations in the image.

Three categories of projection transformation are reported in the literature including the full projection matrix, planar homography and scale factor as indicated in Table 2. In most cases, the projection transformation is following the full perspective model while it could be simplified to an affine camera model when cameras are equipped with large focal length lenses [25].

3.1.1 Full projection matrix

Principle

The full projection matrix is the general form of projection transformation from the 3D structural system to the 2D image plane under no constraint on camera orientation and structural movement directions and is usually used to reconstruct the target 3D structural displacement. The projection relationship is demonstrated in Fig. 3 with a point $\mathrm{P_S}(\mathbf{X}_\mathrm{W} = [\mathrm{X}, \mathrm{Y}, \mathrm{Z}, 1]^\mathrm{T})$ in the structural coordinate system mapping to a point $\mathrm{P_I}(\mathbf{u} = [u, v, 1]^\mathrm{T})$ in the 2D image plane,

$$\alpha\{\mathbf{u}\} = [\mathrm{H}]_{3\times4}\{\mathbf{X}_\mathrm{W}\} \tag{1}$$

where $[\mathrm{H}]_{3\times4}$ is a full projection matrix and α is an arbitrary coefficient.

The calibration process is shown in Fig. 4 with two main steps. The camera intrinsic matrix is usually estimated in the laboratory by analysing a set of images of a calibration object taken from different viewpoints [42]. The calibration object is typically a flat plane or 3D object with a dot or grid pattern of known spacing such as the chessboard pattern shown in Fig. 4. At least three views of the calibration object with four corner points are required, but it is suggested to use at least ten images to derive more robust estimates [43]. After laboratory calibration, any lens functions e.g. autofocus and automated image stabilisation that might lead to changes in camera internal parameters are disabled.

Consumer-grade cameras and smartphone cameras always employ wide angle lenses to increase the field-of-view [15], leading to distorted images particularly in the corner regions of the frame as shown in Fig. 5(a). The lens distortion parameters could also be estimated in laboratory calibration and applied to correct the image with the rectified one in Fig. 5(b). For cameras equipped with lenses producing no apparent lens distortion, the distortion correction step is not necessary. Naturally for the monitoring measurements, it is preferable to locate the target region in the central area of the field of view [10] which suffers less lens distortion, as shown in Fig. 5(a).

In the second step, the camera extrinsic matrix representing the camera position and orientation is estimated on site through point correspondences, i.e. 3D structural coordinates of control points and 2D image coordinates of their projections in an image. Given at least four sets of point correspondences, least-squares optimisation is used to find the best option of camera extrinsic matrix that minimises the total re-projection error between the detected image points and the calculated image projection points.

The calibration algorithms are available in the Vision System Toolbox of MATLAB and the open-source library OpenCV.

Application review

Camera calibration for full projection matrix estimation is commonly used to measure 3D structural displacement, with a few examples illustrating the method: The procedures of laboratory and site camera calibration are described by Kim et al. [44] in an application to structural displacement monitoring in a three-span bridge under truck-induced vibration. The viability of motion capture systems for the laboratory vibration measurement was verified [34] using a T-shaped calibration wand for the estimation of camera extrinsic parameters. In the case of a long span bridge, Martins et al. [14] applied the calibration method to measure the 3D structural displacement at mid-span with the assistance of a set of four active targets. The estimated camera parameters can be refined when multiple cameras with overlapped views are involved. For example, the methodology described by Chang and Ji [24] is based on the epipolar geometry principle of stereoscopic vision where five points including structural point $\mathrm{P_S}$, projection points in two image planes $\mathrm{P_I^1}$ and $\mathrm{P_I^2}$, and two camera optical centres should all be coplanar, as shown in Fig. 6.

Remarks

The full projection matrix is an accurate representation of the projection relationship and is thus applicable to any configuration of cameras on site. The lens distortion problems common for consumer-grade cameras do not prevent their use for such measurements, since corrections are readily made for distortion using laboratory camera calibration.

Camera calibration on site requires position information for some structural points. In existing studies this has been mainly acquired through the installation of artificial targets. Including artificial targets in laboratory tests is easy e.g. attaching a planar chessboard target [24, 44] or placing a planar T-shaped wand in the field of view [17, 34] while the installation efforts in field tests [14] are much greater. The existing examples of two field applications are summarised in Table 3, indicating the feasibility of this method for both short-range and long-range monitoring tests.

3.1.2 Planar homography

Principle

For the case where the target moves within a plane contained in the 3D structural system (e.g. the XY plane), the projection relationship could be simplified to a planar homography between a 2D structural plane $(\mathbf{X}_\mathrm{P} = [\mathrm{X}, \mathrm{Y}, 1]^\mathrm{T})$ and a 2D image plane $(\mathbf{u} = [u, v, 1]^\mathrm{T})$

$$\alpha\{\mathbf{u}\} = [\mathrm{P}]_{3\times 3}\{\mathbf{X}_\mathrm{P}\} \tag{2}$$

where $[\mathrm{P}]_{3\times 3}$ is the planar homography matrix and α is an arbitrary coefficient.

The reconstructed results using planar homography are usually the 2D structural displacement of targets.

The calibration process requires at least four sets of 2D-to-2D point correspondences [46], similar to the estimation process on site in full projection method.

Application review

The planar homography considers the geometric distortion in the projection process and thus has no constraint on camera positioning [47]. The 2D direct linear transform is effective for the planar homography estimation [48], for example the method was applied to monitor the oscillation of a laboratory steel frame with a dense array of markers glued to the surface [48] and the mid-span deformation of a long-span bridge with an attached planar artificial target [49].

Remarks

Planar homography applies no constraint on camera positioning and can be used to recover the target 2D structural displacements. In its application it is usual that the geometric information needed for calibration is provided by attaching artificial planar targets with known dimensions.

This calibration method is based on the assumption that the target moves within a structural plane with negligible motion along the third axis. Any motion not contained within this plane will lead to measurement error unless the motion is purely perpendicular to the camera optical axis.

3.1.3 Scale factor

Principle

Scale factor is the simplest projection transformation and assumes an equal depth-of-field for all projected points or a camera configuration where the optical axis is perpendicular to one structural plane [48]. With this assumption, the mapping process converts to a 1D-1D projection indicated in Fig. 7. The scale factor SF from the structural displacement to the image motion could be determined by one-dimensional correspondence or the camera-to-target distance,

$$\mathrm{SF} = \frac{\left|\mathrm{P_I Q_I}\right|}{\left|\mathrm{P_S Q_S}\right|} \tag{3}$$

or

$$\mathrm{SF} = \frac{f_{pix}}{D} \tag{4}$$

where $\left|\mathrm{P_S Q_S}\right|$ and $\left|\mathrm{P_I Q_I}\right|$ are the known physical dimension on the structural surface and the corresponding pixel length of the projection in image; f_{pix} is the camera lens focal length in terms of pixel units; and D denotes the distance from the camera optical centre to the structural surface plane.

For the system combining a camera with a total station, a projection coefficient called angular resolution [50, 51] is used to perform the transformation which represents the angle value (α in Fig. 7) from the camera optical axis to a projection line with the projection length ($\left|\mathrm{O_I P_I}\right|$) of one pixel. In principle, this projection transformation is similar to the scale factor estimated by camera-to-target distance in Equation (4) where the distance D is measured directly by electronic distance measurement (EDM) instrument and the focal length f_{pix} is related to the angular resolution θ by

$$\theta \approx \tan\theta = 1/f_{pix} \tag{5}$$

Application review

Scale factor has been widely used to transform image motion to structural displacement with the features summarised in Table 4. Mostly the scale factor is determined via a known dimensions in an artificial target attached to structure [5, 8–10, 12, 13, 15, 52–57] while the method using the camera-to-target distance [22] is less popular. For 2D structural displacement measurement, the scale factors for two axes within the target plane are calibrated separately according to dimension correspondences [53–56]. Error analysis indicates that the scale factor by dimension correspondence is less sensitive to the tilt of camera optical axis [8]. However, the scale factor using the camera-to-target distance has no dependence on artificial targets and thus is an easier way to realise completely non-contact monitoring [22].

Remarks

Scale factor is the simplest projection transformation, particularly when no artificial target is used [15, 22] and works when the camera optical axis is perpendicular to the structural surface. Camera positioning is less critical [8] when known structural dimensions are used for calibration. However, when applying the scale factor derived from the camera-to-target distance, the tilt angle of the camera optical axis is suggested to be less than 10° through laboratory validation tests in short distance (≤ 3.7 m) [58]. Care must be taken that different scale factors are applied to different axes to measure the 2D displacement. This simple method can also be used with cameras having apparent lens distortion, since the lens distortion correction method previously described can be used [15, 57].

3.2 Target tracking

Target tracking is the key part of a video processing software package. In this study, target tracking techniques are categorised into four types based on target characteristics shown in Table 5, partly referring to [59].

3.2.1 Template matching

Principle

Template matching is a classic technique for target tracking by searching in a new frame for an area most closely resembling a predefined template, following the procedures demonstrated in Fig. 8. A rectangular region that is a subset in the reference frame is first selected as the template, and could be either an artificial target [5] or a feature target on the structural surface [8]. A correlation criterion is required in order to evaluate the similarity level between the template and the new image subset. Robust criteria for matching are zero-mean normalised cross correlation (ZNCC) and zero-mean normalized sum of squared differences (ZNSSD) which are insensitive to offset and linear scale in illumination [35] while another similarity criterion based on orientation code is also reported to be effective [60]. The definition of the ZNCC criterion is provided as an example in Equation (6) while more details are given in [35].

$$C_{ZNCC} = \sum_{i=-M}^{M} \sum_{j=-N}^{N} \left[\frac{\left(f(x_i, y_j) - f_m \right)\left(g(x_i^{'}, y_j^{'}) - g_m \right)}{\Delta f \, \Delta g} \right] \tag{6}$$

where $f(x_i, y_j)$ and $g(x_i^{'}, y_j^{'})$ denote the image intensity values at the specified pixel locations in the template region and the new frame; f_m and g_m denote the average intensity values in the template region and the new frame; and Δf and Δg denote the standard deviations of intensity values in the template region and the new frame.

The location in the correlation map reaching the highest similarity is taken as the new image location of the target. The default resolution is at pixel level, so interpolation schemes [8] are used to refine the result to the subpixel level. The feasible interpolation methods include bi-cubic interpolation [56], second-order polynomial interpolation [57] in spatial domain and zero-padding interpolation in frequency domain [8]. If the selected target includes robust and identifiable features, Harris corner detection that identifies the edge intersection points through a score value related to the eigenvalues of image gradient matrix could be an alternative to refine the initial matching location [24].

Application review

Template matching is an established method widely applied in structural monitoring from the earliest work on the Humber and Severn Bridges in 1990s [5, 6]. Recent applications include displacement monitoring tests on a railway bridge [8], a long-span bridge [13] and a high-rise building [7].

Digital image correlation (DIC) is an extension of template matching mostly used in experimental mechanics [32, 35], with the difference that DIC considers the shape distortion under large deformation [61] i.e. Lucas-Kanade template matching [62]. As an example, a short-span railway bridge monitoring exercise [63] used normalised correlation-based matching and Lucas-Kanade matching and indicated high similarity in both time and frequency domain.

Remarks

Template matching is easy to use without user intervention apart from the initial selection of the template region. It does not have any special requirement for target patterns and has been validated to work well to track artificial planar targets with specific patterns [5, 6, 24], LED lamp targets [13] and feature targets on structural surfaces [8]. Template matching is not robust to changes in shading, illumination [30, 63] and background conditions [64] in field, although sensitivity to lighting changes might be reduced using camera auto-exposure settings [30]. The method is also not appropriate for tracking slender structural components, since the rectangular subset image used as a template might include background pixels with inconsistent motion.

3.2.2 Feature point matching

Principle

Instead of analysing all the locations within the target, feature point matching applies to sparse ‘special’ points within the target region, independently detecting these special points in two images and then finding point correspondences based on their local appearance. ‘Special’ points in an image, termed ‘interest points’ or ‘keypoints’ in computer vision, are those which are stable, distinctive and invariant to image transformation and illumination changes, such as building corners, connection bolts, or other patches with interesting shapes [65].

The procedures are indicated in Fig. 9. A popular keypoint detector in step (1) is the Harris corner detector [66] which is widely used in structural monitoring applications [11, 15, 22, 24, 57]. Instead of using the pixel values directly for similarity comparison, keypoints are often extracted and described by a more complex representation (i.e. feature descriptor) according to the shape and appearance of a small window around the keypoint [65]. The common descriptors and their matching criteria are indicated in Table 6. Float point based descriptors (e.g. scale-invariant feature transform (SIFT) [67] and speeded up robust features (SURF) [68]) are represented by float vectors, commonly reflecting various local intensity gradients of a pattern around the keypoint. Binary string based descriptors (e.g. Binary robust independent elementary features (BRIEF) [69], Oriented FAST and Rotated BRIEF (ORB) [70] and Fast retina keypoint descriptor (FREAK) [71]) are represented by binary vectors (with elements of 0 and 1) through pairwise comparisons of image intensities (i.e. whether the former is greater or less than the latter) over a special pattern around the keypoint. The matching criterion between two binary descriptors is usually the Hamming distance [69] equal to the number of elements which differ between the two vectors.

To verify the matched keypoint correspondences in step (3), geometric alignment is often used based on whether the keypoints in the first image could fit with the keypoints in the second image after a specific geometric transformation. The widely used approaches for discarding outliers are RANdom SAmple Consensus (RANSAC) [72] and least median of squares (LMS) [73]. The tracking output is the average motion of keypoints in an image that inherently has sub-pixel resolution and could be converted to the target location in the image.

Application review

Song et al. [74] proposed a target tracking method based on circular Hough transform for marker detection and coherent point drift algorithm for marker matching and the method was applied for system identification of a steel cantilever beam in the laboratory. Field applications include Khuc and Catbas [22, 75] who applied the FREAK and SIFT methods for deformation measurement in a stadium structure and a railway bridge and Ehrhart and Lienhart [59, 64] who applied the ORB method for deformation measurement in a short-span footbridge.

Remarks

Feature point matching is an efficient technique since it deals with sparse points instead of the whole region as in template matching. This technique uses local descriptors to represent keypoints instead of the raw image intensities and are less sensitive to illumination change, shape change and scale variation.

However, feature point matching requires the target region to have rich textures for distinctiveness during the whole recording period. Also several parameters need to be adjusted manually according to users' experience or judgement, e.g. contrast threshold for feature detector and distance threshold in matching criteria. These parameter adjustments might depend on environmental changes, e.g. the threshold for outlier removal might vary with the illumination condition [22].

Currently feature point matching technique has only been validated in several short-range measurement tests [22, 59, 64, 75]. However, the feasibility for long-range monitoring in terms of stability over several hours and how to choose the best feature descriptors are open questions.

3.2.3 Optical flow estimation

Principle

Instead of finding matching locations of a complete region or sparse keypoints, optical flow estimation detects motions or flows of all pixels within the target region. Optical flow is the apparent velocity of movement in an image resulting from brightness pattern shift [76]. The calculation imposes two constraints, one temporal constraint on image properties (e.g. image intensity constancy for the same pattern over time) and one spatial constraint that models the flow properties in an image (e.g. coherent motion in adjacent pixels) [77]. A function reflecting these two constraints is then defined and optimised to derive a dense estimation of velocity flow for each pixel. In structural monitoring applications, the output could be converted to image motion instead of velocity by replacing the temporal gradient of image properties in the function with the variation of image properties between two discrete frames. Outlier removal is used to retain only sensible image motions, and average image motion of the inlier pixels is converted to target location inherently having sub-pixel resolution.

Optical flow estimation is an established method with several variant techniques, such as ‘differential’, ‘spatiotemporal energy’ and ‘phase-based’. In this section only two methods, the differential technique of Lucas and Kanade (LK) [78] and the phase-based technique [79] are discussed.

LK optical flow estimation [78] is based on brightness constancy assumption, i.e. projection of the same point has the same image intensity in every frame. Since corner points or keypoints are good features mathematically for the computation of optical flows, LK method is usually applied for sparse estimation instead of computation at every pixel. With keypoints detected in the reference frame usually using the Shi-Tomasi corner detector [80], LK algorithm is applied to compute the image motion of each keypoint in the new frame from spatial-temporal image brightness variations,

$$\begin{bmatrix} \sum_i I_{xi}^2 & \sum_i I_{xi} I_{yi} \\ \sum_i I_{xi} I_{yi} & \sum_i I_{yi}^2 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = \begin{bmatrix} -\sum_i I_{xi} I_{ti} \\ -\sum_i I_{yi} I_{ti} \end{bmatrix} \tag{7}$$

where dx and dy denote the optical flows in the horizontal and vertical directions of the image plane; I_x, I_y and I_t represent the spatial and temporal gradients of image intensities; and i denotes the ith pixel location in a square patch (e.g. 3×3) around a feature point (x, y). The image motion is then estimated after discarding false motion estimates according to RANSAC or LMS, as with feature point matching.

Phase-based optical flow estimation is based on local phase constancy assumption. The method first proposed by Fleet and Jepson in 1990 [79], is receiving new attention together with the motion magnification technique [81] which visualises motions in image sequences that are not visible to the naked eye. The mathematical details of phase-based optical flow estimation are explained in [23] and the algorithm is briefly summarised here.

The Fourier shift theorem indicates that a delay of a signal in the time domain corresponds to a linear phase variation in the frequency domain. Similarly, the image motion in spatial domain is also reflected in phase changes in spatial frequency domain. The phase here is the local phase [82] corresponding to a specific spatial location instead of the whole image, usually derived by employing a quadrature pair of filters consisting of an even real part and an odd imaginary part [83] i.e. Gabor filters [84] and Gaussian derivative filters [23] (demonstrated in Fig. 10). The image motion at each pixel is then estimated from the spatial-temporal variations of the local phase for the filtered image.

Application review

LK optical flow estimation was applied in a laboratory test of a multi-storey metal tower [15] for system identification, and for field application in deformation measurement in a footbridge [59] and bridge stay-cable vibration measurement [85, 86].

Implementations of phase-based optical flow estimation were mostly focused on system identification, i.e. extracting modal frequencies and mode shapes in laboratory tests [23, 87] and identifying modal frequencies of high-rise tower buildings [88].

Remarks

Optical flow estimation enables tracking of features on a structural surface without the requirement for artificial targets. It provides fast and accurate results in controlled environmental conditions.

Like feature point matching, optical flow estimation prefers target patterns with distinct and robust features over the whole test period. Edges are not suitable for tracking due to the ‘aperture problem’ i.e. only the motion component perpendicular to the local edge direction could be detected instead of the true motion of the edge. If the structural motion along edges is one dimensional translation with known direction e.g. bridge stay cable vibration [85], optical flow estimation is viable.

Phase-based optical flow estimation is mostly applied for system identification in the laboratory but is harder to use in field conditions due to high signal noise [88]. Measurement of image motion is sensitive to the choice of pixel location [89], while a selection strategy to ensure satisfactory measurement has not yet been clearly reported. Changes of lighting and background conditions might lead to apparent measurement error [88].

3.2.4 Shape-based tracking

Other than general techniques, there are also some target tracking methods that depend on the special shapes of target patterns which could appear in custom-made artificial targets or structural components (e.g. line-like cables). Table 7 provides a summary of target patterns commonly used. With lack of generality, these methods have limitations for application.

3.2.5 Summary of target tracking performance

In terms of target tracking, the nominal algorithm resolution can be better than 0.01 pixel while the reported accuracy in practice varies from 0.5 to 0.01 pixel [95]. The real-time processing was realised in [8] [49] [63] using the template matching method, in [16] [86] using the optical flow estimation method and in [13] [51] [52] [53] using the shape-based tracking approaches. Although not reported in the existing applications, the feature point matching approach is capable of being used for real-time application [70]. Among the four tracking methods, template matching requires the least user intervention apart from the initial selection of template region while in the other three methods, some threshold values that might be environmentally dependent are required as user inputs.

Ehrhart and Lienhart [64] evaluated the performance of three techniques (optical flow, template matching and feature point matching) by tracking structural features of a footbridge and reported that feature point matching is robust to the changes of background condition (i.e. snowfall) whereas drift over time was observed in the measurement by the two other methods. Busca et al. [96] evaluated three techniques (template matching, edge detection and digital image correlation) on a steel truss railway bridge, concluding that the three techniques provide similar tracking performance while tracking accuracy is slightly poorer for natural targets. Khaloo and Lattanzi [97] investigated four optical flow estimation methods for dense displacement measurement. The study indicated that classic+NL method (i.e. introducing a weighted non-local term into the classical optical flow formulations [77]) provided the most consistent and accurate measurement. However, the coarse-to-fine schemes (i.e. building image pyramids for each frame and computing optical flows on each layer of pyramids to get rid of the small motion constraint) are necessary for Lucas–Kanade and Horn–Schunck methods to deal with large displacement.

3.3 Structural displacement calculation

Structural displacement could be easily derived from the change of structural coordinates given the image location of a target (output of target tracking) and the projection transformation relationship (output of camera calibration). In this case, the projection transformation is a fixed value or matrix without any update during the test.

Another less common method to derive structural displacement is based on the variation of real-time camera extrinsic matrix. The camera extrinsic matrix represents the camera pose i.e. position and orientation relative to the structural system. Since the camera is physically fixed during the recording, variation of camera extrinsic matrix is related to the change of target pose (position and orientation) and could be used to estimate the target motions in six degrees of freedom (6DOF).

3.3.1 Offline projection transformation

Principle

For single camera applications using scale factor or planar homography, the 2D structural coordinate/displacement is derived uniquely through transforming the target location/motion in an image to that in the structure via a projection transformation value or matrix.

When two or more cameras with overlapped views are used to monitor the same target, 3D structural displacement can be extracted based on triangulation method [46].

Application review

Applications of scale factor and planar homography for 2D structural displacement measurement have been reviewed in Section 3.1.

For stereoscopic view or multiple cameras, the triangulation method was used in [24], [33], [98] for 3D structural displacement measurement. A multi-camera arrangement provides more reliable results than a single view but the measurement quality has high dependency on the time-synchronisation of camera recordings.

3.3.2 Online pose estimation

Principle

For single camera applications, using a fixed projection transformation relationship only supports recovery of 2D structural displacement. Some researchers tried to extract more information about target motion (up to 6DOF) using a single camera by updating the real-time target pose in the structural system.

Estimation of camera extrinsic matrix is performed for every frame and the 3D translational and rotational target motions are extracted from the changes of camera extrinsic matrix compared to the initial frame. The calibration process requires at least four non-collinear points with known dimensions or spacing in structure that should have consistent motion.

Application review

Greenbaum et al. [99] applied the online pose estimation method for the laboratory 3D motion measurement of an oscillating rigid object with a few targets of known positions distributed on its surface. In field applications, Chang and Xiao [11] used a planar target with square grid patterns attached to a bridge surface for the measurement of 6DOF structural displacement while Martins et al. [14] tracked four non-coplanar LED targets together to reconstruct the 3D structural motion in a long span bridge.

Remarks

The greatest advantage of the method is the capacity to extract 6DOF structural motions from single camera, but it has a high requirement on the nature of tracked targets which should consist of at least four non-collinear points with precisely known geometry. The target or a set of target points should have rigid motions and be visible during the whole recording period e.g. artificial planar targets with salient corner points [11], distributed target points on structural surface [99] or a set of LED targets [14].

This technique cannot measure translation along the camera optical axis [11], thus the camera should be configured to avoid facing any motion direction of interest.

The measurement accuracy of this method might be poorer than offline projection transformation method. In a footbridge monitoring test by Chang and Xiao [11], using a 36.4 mm focal length camera placed about 5.2 m from mid-span generated measurement noise with standard deviations of 0.76 mm and 1.09 mm in two horizontal directions. This was much larger than would be achieved by offline projection transformation method in a similar test [100] (tracking 0.2 mm bridge vertical displacement with the 85 mm focal length and 27 m camera-to-target distance). Therefore this method is not recommended for field applications unless the target size is not negligible compared to the camera-to-target distance [96].

4 FIELD APPLICATIONS AND CHALLENGES

This section summarises the existing field applications of vision-based systems in two active fields, bridge deformation measurement and cable vibration measurement. A discussion about measurement challenges in field applications is also presented.

4.1 Application examples

Video acquisition devices are now expanded to include smartphone cameras, with numerous applications including vibration measurement of a laboratory multi-floor tower structure [15] and cable vibration measurement of a cable-stayed footbridge [21]. In these two applications, smartphones are only used as the data acquisition system with the recorded video files post-processing for data extraction. Smartphone applications for real-time video acquisition and processing are also viable [101] through experimental validations.

The existing applications of vision-based systems in field tests involve the deformation measurement of a wide range of structural types including: high-rise buildings [7, 88], bridges [5, 6, 8, 10–14, 18, 20, 22, 30, 44, 45, 49, 51–55, 59, 63, 64, 75, 98, 102–104] and stadium structures [22, 105]. Work in the two most active fields, i.e. bridge deformation measurement and cable vibration measurement are summarised in Table 8 and Table 9, respectively.

The viability of vision-based systems for bridge displacement measurement has been verified through comparison with traditional displacement sensors, e.g. LVDT [10, 55, 103], laser sensors [55] and potentiometers [44] for short-span bridge and GPS [13, 30, 49] for long-span bridges. The displacement data could be interpreted for system identification [8, 11, 12, 49, 54, 55, 63, 75], evaluation of load carrying capacity [53], model calibration [18] and estimation of vehicle weights [20]. Artificial targets are commonly used in existing applications to assist camera calibration, whereas recent investigations [51, 63, 75, 103, 104] overcome the dependence on artificial targets and realise completely non-contact sensing based on a simplified projection transformation i.e. scale factor. Another promising application of vision-based systems is to estimate cable tension forces based on vibration measurement. Measurement accuracy was verified through comparison work with traditional sensors e.g. accelerometers [40, 85, 106], velocimeters [41] and load cells [105]. Vision-based systems require no access to cables [16, 30, 40, 85, 86, 90, 105, 106] and are capable of measuring the vibrations of multiple cables using a single camera [16, 86, 105, 106] that is comparable to an array of accelerometers.

4.2 Measurement accuracy and challenges

Measurement accuracy of vision-based systems depends on several parameters, e.g. camera-to-target distance, target pattern features, lighting conditions, camera mounting stability and video processing methods. Khuc et al. [22] investigated the measurement accuracy of a vision-based system in a laboratory and suggested an accuracy of 0.04 mm in a short-range distance (< 14 m). Martins et al. [14] demonstrated the uncertainty evaluation of displacement measurement by a vision-based system on a long-span bridge monitoring test and illustrated a standard measurement accuracy of 1.7 mm in the vertical direction. The high noise range might limit the field application of vision-based systems for system identification on civil structures although high frame rate is taken for vision-based systems.

The achievable accuracy in field tests might be much poorer than that of controlled conditions. The authors investigated the field challenges through a series of monitoring tests in two short-span and two long-span bridges which have been reported in [30]. A summary of the main findings from the tests and the literature is presented here.

- Camera and support motion induced by wind [10] might lead to apparent measurement error. Except for improving camera mounting configurations [30], a common correction method is to additionally track the 'nominal' motion of a fixed target e.g. bridge towers or adjacent buildings. Recent work [97] indicates another promising approach for camera motion compensation through removing the averaged motion of background pixels based on dense optical flow estimation.
- Variation in lighting and background conditions is one of the critical challenges during field tests. The influence of lighting variations might be reduced by enabling camera auto-exposure settings [30].

Correlation-based template matching is not robust to this effect apart from testing during overcast weather, whilst the feature point matching method was reported to be less sensitive [64].

- Atmospheric refraction and turbulence of optical light propagating through the air are common error sources for any optical-based instrument, especially for long-range measurements. Refraction deviation could be minimal for short-term displacement measurement while the air turbulence movement has a larger influence [52]. Quantification of the induced error based on mathematical models is demonstrated in a vision-based measurement test of a long-span bridge [107].
- Observations from short-term tests (with duration less than twelve hours) do not find an apparent influence of temperature variations on measurement accuracy, while this effect is necessary to consider for long-term tests e.g. with duration a few months or more. A time-frequency approach indicates the potential for error compensation based on investigation of the correlation models linking measurement errors and temperatures [108].

5 SUMMARY AND PROSPECTS

As evidenced from the review, vision-based systems are promising tools for structural displacement measurement having advantages in cost, installation efforts and measurement capacities of frequency range and spatial resolution. Although the potential in field applications has been validated in many articles, there are a few aspects still to mature.

- Robust target tracking methods. Template matching and optical flow estimation are established methods widely used in short-range and long-range measurement tests, but they are not robust to lighting and background changes. Feature point matching is a relatively new and promising tracking method, but investigations regarding several aspects e.g. selection strategy of proper threshold parameters, sensitivity on environmental effects and field viability for long-range measurement are rare, and need to be expanded. It is still an open question about the most robust tracking method for vision-based systems to deal with changes in lighting conditions during field tests.
- Completely non-contact measurement. Artificial targets are commonly included to assist the camera calibration process, but dependence on artificial targets is eliminated in a few field applications [15, 22, 75, 104]. These studies were based on a simplified projection transformation i.e. scale factor that is not a general approach and imposes constraints on camera positioning. To develop a non-contact vision-based system for

the general case, requiring control points with known locations is the main obstacle which could possibly be resolved via the assistance of surveying instruments, such as total station.

- Distributed sensing of structural displacement. Vision-based systems have the capacity for simultaneous multi-point displacement measurement that is comparable or superior to an array of accelerometers for system identification. Currently, bridge applications primarily focus on the mid-span displacement measurement, while the potential of distributed sensing and system identification is not well investigated.
- Measurement uncertainty evaluation. Measurement accuracy and uncertainty are of great importance for a mature measurement system. Quantified descriptions about measurement accuracy haven been made in some references (e.g. [8, 22, 54, 64]) through comparisons with reference measurements. However, the quality of vision-based measurements could be time-varying, environmentally dependent and differ significantly with various test configurations. The influential factors include the test configurations (e.g. the camera-to-target distance and the target features), the video processing methods used and the environmental conditions (e.g. the lighting conditions, the atmospheric refraction and turbulence). A systematic evaluation of vision-based measurement methodologies will require extensive experimental effort by the research community with publication of case studies contributing to evolving guidance for field applications.

LIST OF ABBREVIATIONS

The following table describes the abbreviations and acronyms used throughout this article.

Abbreviation	Meaning
BRIEF	Binary robust independent elementary features
DIC	digital image correlation
DOF	degree of freedom
FREAK	Fast retina keypoint descriptor
LK	Lucas and Kanade optical flow
LMS	least median of squares
MCS	motion capture systems
ORB	Oriented FAST and Rotated BRIEF
RANSAC	RANdom SAmple Consensus
SF	scale factor
SIFT	scale-invariant feature transform
SURF	speeded up robust features
ZNCC	zero-mean normalised cross correlation
ZNSSD	zero-mean normalised sum of squared differences

Compliance with Ethical Standards:

Conflict of Interest:

The authors declare that they have no conflict of interest.

References

1. Casciati F, Fuggini C (2009) Engineering vibration monitoring by GPS: long duration records. Earthq Eng Eng Vib 8:459–467.

2. Nickitopoulou A, Protopsalti K, Stiros SC (2006) Monitoring dynamic and quasi-static deformations of large flexible engineering structures with GPS: Accuracy, limitations and promises. Eng Struct 28:1471–1482.

3. Farquharson FB, Vincent GS (1954) Aerodynamic stability of suspension bridges with special reference to the Tacoma Narrows bridge. No. 116. University of Washington Press

4. Marecos J, Castanheta M, Trigo JT (1969) Field observation of Tagus River suspension bridge. J Struct Eng 95:555–583.

5. Stephen GA, Brownjohn JMW, Taylor CA (1993) Measurements of static and dynamic displacement from visual monitoring of the Humber Bridge. Eng Struct 15:197–208.

6. Macdonald JHG, Dagless EL, Thomas BT, Taylor CA (1997) Dynamic measurements of the Second Severn Crossing. Proc Inst Civ Eng - Transp 123:241–248. doi: 10.1680/itran.1997.29978

7. Liao WY, Chen WH, Ni YQ, Xia Y (2010) Development of a vision-based real-time displacement measurement system for Guangzhou New TV Tower. In: Casciati F, Giordano M. (eds) Proc. 5th Eur. Work. Struct. Heal. Monit. Sorrento, Naples, Italy, pp 450–455

8. Feng D, Feng M, Ozer E, Fukuda Y (2015) A Vision-Based Sensor for Noncontact Structural Displacement Measurement. Sensors 15:16557–16575. doi: 10.3390/s150716557

9. Kim S-W, Kim N-S (2011) Multi-point Displacement Response Measurement of Civil Infrastructures Using Digital Image Processing. Procedia Eng 14:195–203. doi: 10.1016/j.proeng.2011.07.023

10. Ribeiro D, Calçada R, Ferreira J, Martins T (2014) Non-contact measurement of the dynamic displacement of railway bridges using an advanced video-based system. Eng Struct 75:164–180. doi: 10.1016/j.engstruct.2014.04.051

11. Chang CC, Xiao XH (2010) Three-Dimensional Structural Translation and Rotation Measurement Using Monocular Videogrammetry. J Eng Mech 136:840–848. doi: 10.1061/(ASCE)EM.1943-7889.0000127

12. Wahbeh AM, Caffrey JP, Masri SF (2003) A vision-based approach for the direct measurement of displacements in vibrating systems. Smart Mater Struct 12:785–794. doi: 10.1088/0964-1726/12/5/016

13. Ye XW, Ni YQ, Wai TT, et al (2013) A vision-based system for dynamic displacement measurement of long-span bridges: algorithm and verification. Smart Struct Syst 12:363–379. doi: 10.12989/sss.2013.12.3_4.363

14. Martins LL, Rebordão JM, Ribeiro AS (2015) Structural observation of long-span suspension bridges for safety assessment: implementation of an optical displacement measurement system. J Phys Conf Ser 588:12004. doi: 10.1088/1742-6596/588/1/012004

15. Yoon H, Elanwar H, Choi H, et al (2016) Target-free approach for vision-based structural system identification using consumer-grade cameras. Struct Control Heal Monit 23:1405–1416. doi: 10.1002/stc.1850

16. Caetano E, Silva S, Bateira J (2007) Application of a vision system to the monitoring of cable structures. In: Seventh Int. Symp. Cable Dyn. pp 225–236

17. Oh BK, Hwang JW, Kim Y, et al (2015) Vision-based system identification technique for building structures using a motion capture system. J Sound Vib 356:72–85. doi: 10.1016/j.jsv.2015.07.011

18. Feng D, Feng MQ (2015) Model Updating of Railway Bridge Using In Situ Dynamic Displacement Measurement under Trainloads. J Bridg Eng 20:4015019. doi: 10.1061/(ASCE)BE.1943-5592.0000765

19. Cha Y-J, Chen JG, Büyüköztürk O (2017) Output-only computer vision based damage detection using phase-based optical flow and unscented Kalman filters. Eng Struct 132:300–313. doi: 10.1016/j.engstruct.2016.11.038

20. Ojio T, Carey CH, OBrien EJ, et al (2016) Contactless Bridge Weigh-in-Motion. J Bridg Eng 21:4016032. doi: 10.1061/(ASCE)BE.1943-5592.0000776

21. Zhao X, Ri K, Wang N (2017) Experimental Verification for Cable Force Estimation Using Handheld Shooting of Smartphones. J Sensors 2017:1–13. doi: 10.1155/2017/5625396

22. Khuc T, Catbas FN (2017) Completely contactless structural health monitoring of real-life structures using cameras and computer vision. Struct Control Heal Monit 24:e1852. doi: 10.1002/stc.1852

23. Chen JG, Wadhwa N, Cha Y-J, et al (2015) Modal identification of simple structures with high-speed video using motion magnification. J Sound Vib 345:58–71. doi: 10.1016/j.jsv.2015.01.024

24. Chang CC, Ji YF (2007) Flexible Videogrammetric Technique for Three-Dimensional Structural Vibration Measurement. J Eng Mech 133:656–664. doi: 10.1061/(ASCE)0733-9399(2007)133:6(656)

25. Santos CA, Costa CO, Batista J (2016) A vision-based system for measuring the displacements of large structures: Simultaneous adaptive calibration and full motion estimation. Mech Syst Signal Process 72–73:678–694. doi: 10.1016/j.ymssp.2015.10.033

26. Schreier HW (2004) Advances in Light Microscope Stereo Vision. Exp Mech 44:278–288. doi: 10.1177/0014485104041546

27. Wu L, Casciati F (2014) Local positioning systems versus structural monitoring: a review. Struct Control Heal Monit 21:1209–1221. doi: 10.1002/stc.1643

28. Baqersad J, Poozesh P, Niezrecki C, Avitabile P (2017) Photogrammetry and optical methods in structural

dynamics – A review. Mech Syst Signal Process 86:17–34. doi: 10.1016/j.ymssp.2016.02.011

29. Ye XW, Dong CZ, Liu T (2016) A Review of Machine Vision-Based Structural Health Monitoring: Methodologies and Applications. J Sensors 2016:1–10. doi: 10.1155/2016/7103039

30. Brownjohn JMW, Xu Y, Hester D (2017) Vision-Based Bridge Deformation Monitoring. Front Built Environ 3:1–16. doi: 10.3389/fbuil.2017.00023

31. Yoneyama S, Kitagawa A, Iwata S, et al (2007) Bridge deflection measurement using digital image correlation. Exp Tech 31:34–40. doi: 10.1111/j.1747-1567.2007.00132.x

32. Schreier H, Orteu J-J, Sutton MA (2009) Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications. Springer Science & Business Media

33. Jáuregui D V., White KR, Woodward CB, Leitch KR (2003) Noncontact Photogrammetric Measurement of Vertical Bridge Deflection. J Bridg Eng 8:212–222. doi: 10.1061/(ASCE)1084-0702(2003)8:4(212)

34. Park SW, Park HS, Kim JH, Adeli H (2015) 3D displacement measurement model for health monitoring of structures using a motion capture system. Measurement 59:352–362. doi: 10.1016/j.measurement.2014.09.063

35. Pan B, Qian K, Xie H, Asundi A (2009) Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review. Meas Sci Technol 20:62001. doi: 10.1088/0957-0233/20/6/062001

36. Sutton MA, Yan JH, Tiwari V, et al (2008) The effect of out-of-plane motion on 2D and 3D digital image correlation measurements. Opt Lasers Eng 46:746–757. doi: 10.1016/j.optlaseng.2008.05.005

37. Hartley RI, Mundy JL (1993) Relationship between photogrammmetry and computer vision. In: Integr. Photogramm. Tech. with Scene Anal. Mach. Vis. pp 92–105

38. Jiang R, Jáuregui D V., White KR (2008) Close-range photogrammetry applications in bridge measurement: Literature review. Measurement 41:823–834. doi: 10.1016/j.measurement.2007.12.005

39. Moeslund G, Granum E (2001) A survey of computer vision-based human motion capture. Comput Vis Image Underst 81:231–268.

40. Kim S-W, Kim N-S (2013) Dynamic characteristics of suspension bridge hanger cables using digital image processing. NDT E Int 59:25–33. doi: 10.1016/j.ndteint.2013.05.002

41. Chen CC, Wu WH, Tseng HZ, et al (2015) Application of digital photogrammetry techniques in identifying the mode shape ratios of stay cables with multiple camcorders. Meas J Int Meas Confed 75:134–146. doi: 10.1016/j.measurement.2015.07.037

42. Zhang Z (2000) A flexible new technique for camera calibration. IEEE Trans Pattern Anal Mach Intell 22:1330–1334. doi: 10.1109/34.888718

43. Bradski G, Kaehler A (2008) Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Sebastopol, CA

44. Kim SC, Kim HK, Lee CG, Kim SB (2006) A vision system for identifying structural vibration in civil engineering constructions. In: SICE-ICASE Int. Jt. Conf. Bexco, Busan, Korea, pp 5813–5818

45. Martins LL, Rebordão JM, Ribeiro ÁS (2014) Optical Metrology applied to 3D displacement measurement of long-span suspension bridge dynamics. In: A. Cunha, E. Caetano, P. Ribeiro GM (ed) Proc. 9th Int. Conf. Struct. Dyn. Porto, Portugal, pp 2135–2142

46. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press

47. Lee J, Cho S, Sim S (2015) Monocular Vision-based Displacement Measurement System Robust to Angle and Distance Using Homography. 6th Int. Conf. Adv. Exp. Struct. Eng.

48. Wu L-J, Casciati F, Casciati S (2014) Dynamic testing of a laboratory model via vision-based sensing. Eng Struct 60:113–125. doi: 10.1016/j.engstruct.2013.12.002

49. Xu Y, Brownjohn J, Hester D, Koo KY (2016) Dynamic displacement measurement of a long span bridge using vision-based system. In: 8th Eur. Work. Struct. Heal. Monit. (EWSHM 2016). Bilbao, Spain, pp 5–8

50. Charalampous E, Psimoulis P, Guillaume S, et al (2015) Measuring sub-mm structural displacements using QDaedalus: a digital clip-on measuring system developed for total stations. Appl Geomatics 7:91–101. doi: 10.1007/s12518-014-0150-z

51. Ehrhart M, Lienhart W (2015) Image-based dynamic deformation monitoring of civil engineering structures from long ranges. In: Lam EY, Niel KS (eds) Proc. SPIE. International Society for Optics and Photonics, San Francisco, California, p 94050J

52. Olaszek P (1999) Investigation of the dynamic characteristic of bridge structures using a computer vision

method. Measurement 25:227–236. doi: 10.1016/S0263-2241(99)00006-8

53. Lee JJ, Cho S, Shinozuka M, et al (2006) Evaluation of Bridge Load Carrying Capacity Based on Dynamic Displacement Measurement Using Real-time Image Processing Techniques. Steel Struct 6:377–385.

54. Lee JJ, Shinozuka M (2006) A vision-based system for remote sensing of bridge displacement. NDT E Int 39:425–431. doi: 10.1016/j.ndteint.2005.12.003

55. Fukuda Y, Feng MQ, Shinozuka M (2010) Cost-effective vision-based system for monitoring dynamic response of civil engineering structures. Struct Control Heal Monit 17:918–936. doi: 10.1002/stc.360

56. Choi H-S, Cheung J-H, Kim S-H, Ahn J-H (2011) Structural dynamic displacement vision system using digital image processing. NDT E Int 44:597–608. doi: 10.1016/j.ndteint.2011.06.003

57. Sładek J, Ostrowska K, Kohut P, et al (2013) Development of a vision based deflection measurement system and its accuracy assessment. Measurement 46:1237–1249. doi: 10.1016/j.measurement.2012.10.021

58. Choi I, Kim J, Kim D (2016) A Target-Less Vision-Based Displacement Sensor Based on Image Convex Hull Optimization for Measuring the Dynamic Response of Building Structures. Sensors 16:2085. doi: 10.3390/s16122085

59. Ehrhart M, Lienhart W (2015) Development and evaluation of a long range image-based monitoring system for civil engineering structures. In: Shull PJ (ed) Proc. SPIE Struct. Heal. Monit. Insp. Adv. Mater. Aerospace, Civ. Infrastruct. San Diego, California, United States, p 94370K

60. Fukuda Y, Feng MQ, Narita Y, et al (2013) Vision-Based Displacement Sensor for Monitoring Dynamic Response Using Robust Object Search Algorithm. IEEE Sens J 13:4725–4732. doi: 10.1109/JSEN.2013.2273309

61. Baker S, Matthews I (2004) Lucas-Kanade 20 Years On: A Unifying Framework. Int J Comput Vis 56:221–255. doi: 10.1023/B:VISI.0000011205.11775.fd

62. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proc. 7th Int. Jt. Conf. Artif. Intell. Morgan Kaufmann Publishers, Vancouver, BC, Canada, pp 674–679

63. Guo J, Zhu C (2016) Dynamic displacement measurement of large-scale structures based on the Lucas–

Kanade template tracking algorithm. Mech Syst Signal Process 66–67:425–436. doi: 10.1016/j.ymssp.2015.06.004

64. Ehrhart M, Lienhart W (2015) Monitoring of Civil Engineering Structures using a State-of-the-art Image Assisted Total Station. J Appl Geod 9:174–182. doi: 10.1515/jag-2015-0005

65. Szeliski R (2011) Computer Vision: Algorithms and Applications. Springer Science & Business Media, London, p 832

66. Harris C, Stephens M (1988) A combined corner and edge detector. Proc. Fourth Alvey Vis. Conf. 15:

67. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110.

68. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) Speeded-up robust features (SURF). Comput Vis image Underst 110:346–359.

69. Calonder M, Lepetit V, Strecha C, Fua P (2010) Brief: Binary robust independent elementary features. In: Eur. Conf. Comput. Vis. Crete, Greece, pp 778–792

70. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. In: Proc. IEEE Int. Conf. Comput. Vis. IEEE, pp 2564–2571

71. Alahi A, Ortiz R, Vandergheynst P (2012) Freak: Fast retina keypoint. In: IEEE Conf. Comput. Vis. pattern Recognit. pp 510–517

72. Fischler MA, Bolles RC (1981) Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24:381–395.

73. Rousseeuw PJ (1984) Least median of squares regression. J Am Stat Assoc 79:871–880.

74. Song Y-Z, Bowen CR, Kim AH, et al (2014) Virtual visual sensors and their application in structural health monitoring. Struct Heal Monit An Int J 13:251–264. doi: 10.1177/1475921714522841

75. Khuc T, Catbas FN (2017) Computer vision-based displacement and vibration monitoring without using physical target on structures. Struct Infrastruct Eng 13:505–516. doi: 10.1080/15732479.2016.1164729

76. Beauchemin SS, Barron JL (1995) The computation of optical flow. ACM Comput Surv 27:433–466. doi: 10.1145/212094.212141

77. Sun D, Roth S, Black MJ (2010) Secrets of optical flow estimation and their principles. In: IEEE Conf. Comput. Vis. Pattern Recognit. pp 2432–2439

78. Tomasi C, Kanade T (1991) Detection and Tracking of Point Features. Int J Comput Vis. doi: 10.1016/S0031-3203(03)00234-6

79. Fleet DJ, Jepson AD (1990) Computation od Component Image Velocity from local phase information. Int J Comput Vis 5:77–104.

80. Jianbo Shi, Tomasi (1994) Good features to track. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR-94. IEEE Comput. Soc. Press, pp 593–600

81. Wadhwa N, Rubinstein M, Durand F, Freeman WT (2013) Phase-based video motion processing. ACM Trans Graph 32:1. doi: 10.1145/2461912.2461966

82. Bülow T (1999) Hypercomplex spectral signal representations for the processing and analysis of images. Christian-Albrechts-Universität zu Kiel

83. Boukerroui D, Noble JA, Brady M (2004) On the Choice of Band-Pass Quadrature Filters. J Math Imaging Vis 21:53–80. doi: 10.1023/B:JMIV.0000026557.50965.09

84. Gautama T, Van Hulle MA (2002) A phase-based approach to the estimation of the optical flow field using spatial filtering. IEEE Trans Neural Networks 13:1127–1136. doi: 10.1109/TNN.2002.1031944

85. Ji YF, Chang CC (2008) Nontarget Image-Based Technique for Small Cable Vibration Measurement. J Bridg Eng 13:34–42. doi: 10.1061/(ASCE)1084-0702(2008)13:1(34)

86. Caetano E, Silva S, Bateira J (2011) A vision system for vibration monitoring of civil engineering structures. Exp Tech 35:74–82. doi: 10.1111/j.1747-1567.2010.00653.x

87. Yang Y, Dorn C, Mancini T, et al (2017) Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification. Mech Syst Signal Process 85:567–590. doi: 10.1016/j.ymssp.2016.08.041

88. Chen JG, Davis A, Wadhwa N, et al (2017) Video Camera-Based Vibration Measurement for Civil Infrastructure Applications. J Infrastruct Syst 23:11. doi: 10.1061/(asce)is.1943-555x.0000348

89. Diamond DH, Heyns PS, Oberholster AJ (2017) Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence. Measurement 95:166–172. doi: 10.1016/j.measurement.2016.10.021

90. Ji YF, Chang CC (2008) Nontarget Stereo Vision Technique for Spatiotemporal Response Measurement

of Line-Like Structures. J Eng Mech 134:466–474. doi: 10.1061/(ASCE)0733-9399(2008)134:6(466)

91. Shan B, Zheng S, Ou J (2015) Free vibration monitoring experiment of a stayed-cable model based on stereovision. Measurement 76:228–239. doi: 10.1016/j.measurement.2015.08.025

92. Park C, Ho H-N, Jo B-W, Lee J-J (2014) An Efficient Vision-Based Three-Dimensional Motion Measurement System for Civil Infra-Structures. Exp Tech 40:1–9. doi: 10.1111/ext.12103

93. Ho H-N, Lee J-H, Park Y-S, Lee J-J (2012) A Synchronized Multipoint Vision-Based System for Displacement Measurement of Civil Infrastructures. Sci World J 2012:1–9. doi: 10.1100/2012/519146

94. Yi J-H, Kim J-H, Jeong WM, Chae J-W (2013) Field evaluation of optical-based three-dimensional dynamic motion measurement system with multiple targets for a floating structure. Ocean Eng 62:140–151. doi: 10.1016/j.oceaneng.2012.12.046

95. Bing P, Hui-min X, Bo-qin X, Fu-long D (2006) Performance of sub-pixel registration algorithms in digital image correlation. Meas Sci Technol 17:1615–1621. doi: 10.1088/0957-0233/17/6/045

96. Busca G, Cigada A, Mazzoleni P, Zappa E (2014) Vibration Monitoring of Multiple Bridge Points by Means of a Unique Vision-Based Measuring System. Exp Mech 54:255–271. doi: 10.1007/s11340-013-9784-8

97. Khaloo A, Lattanzi D (2017) Pixel-wise structural motion tracking from rectified repurposed videos. Struct Control Heal Monit 24:e2009. doi: 10.1002/stc.2009

98. Santos CA, Batista JP, Costa CO (2009) A Non-Contact Measurement System for Monitoring the Displacements of Long Deck Suspension Bridges. In: 16th Int. road Fed. World Road Meet. Lisbon, Portugal, pp 1–7

99. Greenbaum RJY, Smyth AW, Chatzis MN (2016) Monocular Computer Vision Method for the Experimental Study of Three-Dimensional Rocking Motion. J Eng Mech 142:4015062. doi: 10.1061/(ASCE)EM.1943-7889.0000972

100. Brownjohn JMW, Hester D, Xu Y, et al (2016) Viability of optical tracking systems for monitoring deformations of a long span bridge. 6th Eur. Conf. Struct. Control

101. Min J-H, Gelo NJ, Jo H (2016) Real-time image processing for non-contact monitoring of dynamic displacements using smartphone technologies. Sensors Smart Struct Technol Civil, Mech Aerosp Syst.

doi: 10.1117/12.2219418

102. Yoneyama S, Ueda H (2012) Bridge Deflection Measurement Using Digital Image Correlation with Camera Movement Correction. Mater Trans 53:285–290. doi: 10.2320/matertrans.I-M2011843

103. Feng MQ, Fukuda Y, Feng D, Mizuta M (2015) Nontarget Vision Sensor for Remote Measurement of Bridge Dynamic Response. J Bridg Eng 20:4015023. doi: 10.1061/(ASCE)BE.1943-5592.0000747

104. Feng D, Feng MQ (2017) Experimental validation of cost-effective vision-based structural health monitoring. Mech Syst Signal Process 88:199–211. doi: 10.1016/j.ymssp.2016.11.021

105. Feng D, Scarangello T, Feng MQ, Ye Q (2017) Cable tension force estimate using novel noncontact vision-based sensor. Measurement 99:44–52. doi: 10.1016/j.measurement.2016.12.020

106. Kim S-W, Jeon B-G, Kim N-S, Park J-C (2013) Vision-based monitoring system for evaluating cable tensile forces on a cable-stayed bridge. Struct Heal Monit An Int J 12:440–456. doi: 10.1177/1475921713500513

107. Lages Martins L, Nunes Vicente Rebordão JM, Silva Ribeiro Á (2014) Thermal Influence on Long-Distance Optical Measurement of Suspension Bridge Displacement. Int J Thermophys 35:693–711. doi: 10.1007/s10765-014-1607-3

108. Zhou HF, Zheng JF, Xie ZL, et al (2017) Temperature effects on vision measurement system in long-term continuous monitoring of displacement. Renew Energy. doi: 10.1016/j.renene.2017.07.104

TABLES

Table 1 Summary of vision-based systems

Vision-based systems	Main study objects	Measurement information	Features
Digital image correlation (DIC)	Small-scale experimental members under large distortional deformation	Full-field displacements or strains on member surface	• Laboratory application in controlled environment; • Fixed camera locations; • Dense measurement with high resolution; • Usually large deformation with shape distortion.
Motion capture systems (MCS)	Objects or human bodies with a high degree-of-freedom skeleton structure	3D locations of each joints in structure	• Laboratory application in controlled environment; • Fixed camera locations; • At least two cameras with overlapped views; • Markers and calibration object for calibration assistance.
Photogrammetry	Initially aerial and terrestrial applications; now bridges under live loads	3D geometry of objects and deflection measurement	• Field applications on structures mainly in stationary status; • Movable locations of camera; • Distributed control points for calibration assistance.
System for structural monitoring	Structures with small deformation compared with structure scale.	2D or 3D displacement with proper sample rate.	• Field applications and easy installation preferred; • High accuracy and also high calculation efficiency (for real-time dynamic measurement); • Small deformation compared with structure scale and camera-to-structure distance.

Table 2 Projection transformation from structure to image plane

Projection transformation		**Assumptions**	**Recovered localisation information of target**
(1)	Full projection matrix	--	3D structural coordinates
(2)	Planar homography	The motion along one axis in structural coordinate system is negligible	2D structural coordinates
(3)	Scale factor	The camera optical axis is perpendicular to one plane in the structural coordinate system (e.g. the target plane XY).	2D motions within the target plane

Table 3 Summary of two field applications in literature using the full projection matrix as projection transformation

References	**[11]**	**[14, 45]**
Focal length	36.4 mm	600 mm (composed by a 300 mm telephoto lens and a 2x tele-converter)
Camera-to-target distance	5.2 m	500 m
Artificial targets installed	A planar 3x3 chessboard	A 3D target set combined by distributed four LED targets with the whole dimensions of 250 mm, 350 mm and 250 mm along the three axes
Observed maximum displacement	6 mm	1.82 m
Measurement evaluation	Not commented about vertical measurement; Measurement noise along the two other directions with the standard deviations at 0.76 mm and 1.09 mm, respectively.	Uncertainty at 15 mm to 20 mm in the vertical and transverse directions.

Table 4 Features of two calibration methods for scale factor

Scale factor	**By camera-to-target distance**	**By dimension correspondences**
Target dependence	Target free	Artificial targets always required
Camera positioning constraint	Very sensitive to the tilt of camera optical axis	Less sensitive to the tilt of camera optical axis
Applications	Mostly used in the short-range measurement; The long-range measurement feasible for the vision-based systems assisted the total station.	Widely used in both the short-range and the long-range measurement

Table 5 Categories of target tracking methods

Tracking methods		**Regions or points tracked for matching**
(1)	Template matching	Rectangular subset of the frame as the target region
(2)	Feature point matching	Sparse ‘special’ points with salient features within the target region
(3)	Optical flow estimation	Every pixel location within the target region
(4)	Shape-based tracking	Line-type, circular-shaped or custom-made targets

Table 6 Categories of feature descriptors and corresponding matching criteria

Descriptor categories	**Descriptor names**	**Matching criteria**
Float point based	Scale-invariant feature transform (SIFT) [67] Speeded up robust features (SURF) [68]	Euclidean distances in feature space [65]
Binary string based	Binary robust independent elementary features (BRIEF) [69] Oriented FAST and Rotated BRIEF (ORB) [70] Fast retina keypoint descriptor (FREAK) [71]	Hamming distance [69]

Table 7 Examples of shape-based tracking

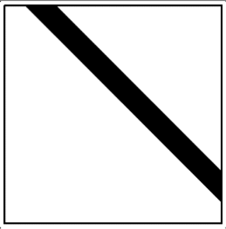
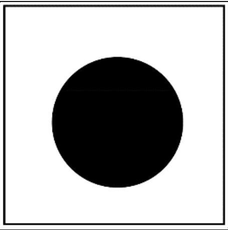
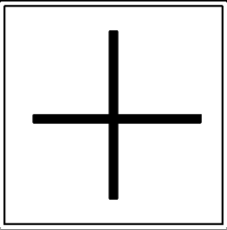
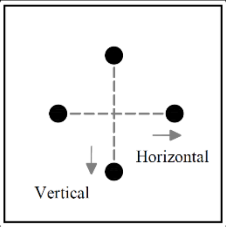
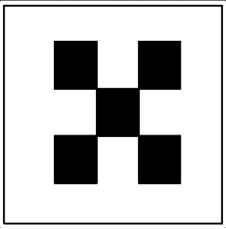
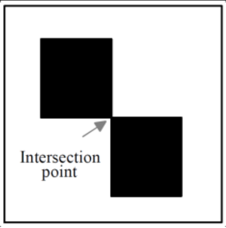
Ref.	Target patterns	Determination of target location
[30, 90]		Detecting the edges of line-shaped patterns and building image point correspondences among image sequences [90] or computing the cable motion from the distance between two identified edges [30]
[12, 41][91][92]		Detecting the edges of circular-shaped patterns through brightness thresholding or edge detection and computing the centroid coordinates for the circle
[52]		Detecting the edges of cross-shaped patterns through image gradient and computing the arithmetic mean of edge coordinates as the target location
[53–55][93][94]		Detecting four spots through brightness thresholding and computing the motions along the specified horizontal and vertical directions
[11]		Detecting grid dots by Harris corner detector and applying the image coordinates of grid dots for the estimation of camera extrinsic matrix
[10]		Detecting the edges of squares through brightness thresholding and computing the coordinates of the intersection point

Table 8 Review of studies about bridge displacement measurement using vision-based systems

Ref.	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
[5, 6]	Humber Bridge & Second Severn Crossing, UK	Scale factor	Correlation-based template matching	Planar target	2D displacement at mid-span	
[52]	A highway bridge & a railway viaduct	Scale factor	Shape-based tracking	Planar target	Vertical displacement at mid-span	
[12]	Vincent Thomas Bridge, USA	Scale factor	Shape-based tracking	LED targets	Vertical displacement at mid-span	• Extracting modal frequencies
[53–55]	Highway bridges	Scale factor	Shape-based tracking	Planar target	2D displacement at mid-span	• Estimating load carrying capacity for load test • Evaluating measurement by comparison with the reference sensors (LVDT & Laser)
[44]	A roadway bridge	Full projection matrix	Shape-based tracking	Planar target	2D displacement at mid-span	• Evaluating measurement by comparison with the reference sensor (potentiometer)
[11]	A cable-stayed footbridge	Online pose estimation	Shape-based tracking	Planar target	3D displacement at mid-span	• Extracting modal frequencies
[13]	Tsing Ma Bridge, Hong Kong	Scale factor	Correlation-based template matching	LED targets	Vertical displacement at mid-span	• Evaluating measurement by comparison with the reference sensor (GPS)
[63]	A railway viaduct	Scale factor	Lucas-Kanade template matching	Natural features	Vertical displacement of sound barrier	• Extracting modal frequencies
[10]	A railway bridge	Scale factor	Shape-based tracking	Planar target	Vertical displacement at mid-span	• Evaluating measurement by comparison with reference sensor (LVDT)
[8, 18, 103, 104]	A footbridge, a highway bridge & a railway bridge	Scale factor	Correlation-based template matching	Both planar target and natural features	Vertical displacement at mid-span	• Extracting modal frequencies • FE model calibration
[14, 45]	P25A bridge, Portugal	Online pose estimation	Shape-based tracking	LED targets	3D displacement at mid-span	• Evaluating measurement uncertainty
[51, 59, 64]	A footbridge	Scale factor	Correlation-based template matching	Both planar target and natural features	Vertical displacement at mid-span	• Extracting modal frequencies

Ref.	Application structures	Camera calibration	Target tracking method	Target type	Measured displacement	Data interpretation
			Feature point matching Optical flow estimation			• Evaluating measurement accuracy and stability of three tracking methods
[30, 49]	Humber Bridge, UK	Planar homography	Correlation-based template matching	Planar target	2D displacement at mid-span	• Extracting modal frequencies • Evaluating measurement by comparison with the reference sensor (GPS)
[20]	A roadway bridge	Scale factor	Lukas–Kanade method	Natural features	Vertical displacement at mid-span	• Estimating vehicle weights
[75]	A railway bridge	Scale factor	Feature point matching	Natural features	Vertical displacement at mid-span	• Extracting modal frequencies

Table 9 Review of studies about cable vibration measurement using vision-based systems

Ref.	**Application structures**	**Target tracking methods**	**Data interpretation**
[85]	A footbridge	Optical flow estimation	• Extracting modal frequencies
[90]	A footbridge	Shape-based tracking	• Extracting modal frequencies • Identifying mode shapes.
[86]	Guadiana Bridge, Portugal	Optical flow estimation	• Extracting modal frequencies
[40, 106]	Gwangan Bridge and a two-pylon cable-stayed bridge in Busan-Geoje Fixed Link, Korea	Correlation-based template matching	• Extracting modal frequencies • Estimating cable tension
[41]	Chi-Lu Bridge, Taiwan China	Shape-based tracking	• Extracting modal frequencies • Identifying the mode shape ratio of cables
[105]	Hard Rock Stadium, USA	Correlation-based template matching	• Extracting modal frequencies • Estimating cable tension
[30]	A footbridge	Shape-based tracking	• Extracting modal frequencies
[21]	A footbridge	Edge detection	• Extracting modal frequencies • Estimating cable tension

FIGURE CAPTIONS

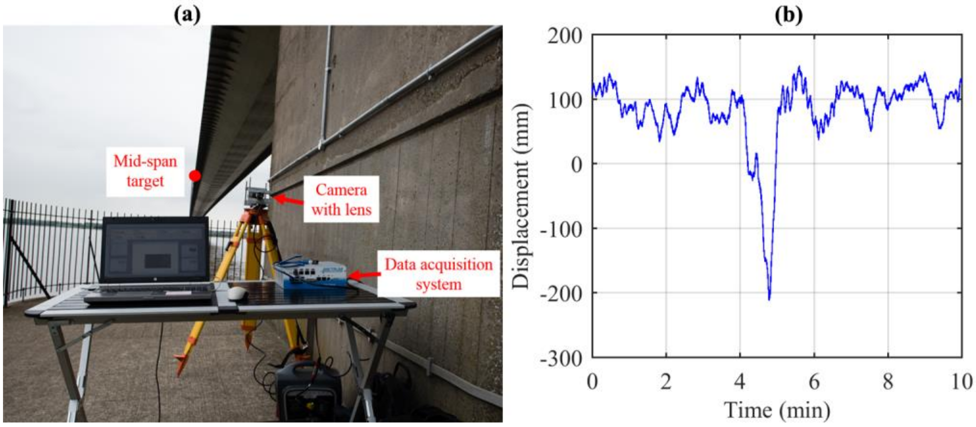


Fig. 1 Vision-based system for structural displacement monitoring of the Humber Bridge [30]: (a) site configuration of the vision-based monitoring system; and (b) 10-min time history signal of vertical displacement at the bridge mid-span measured by the vision-based monitoring system.

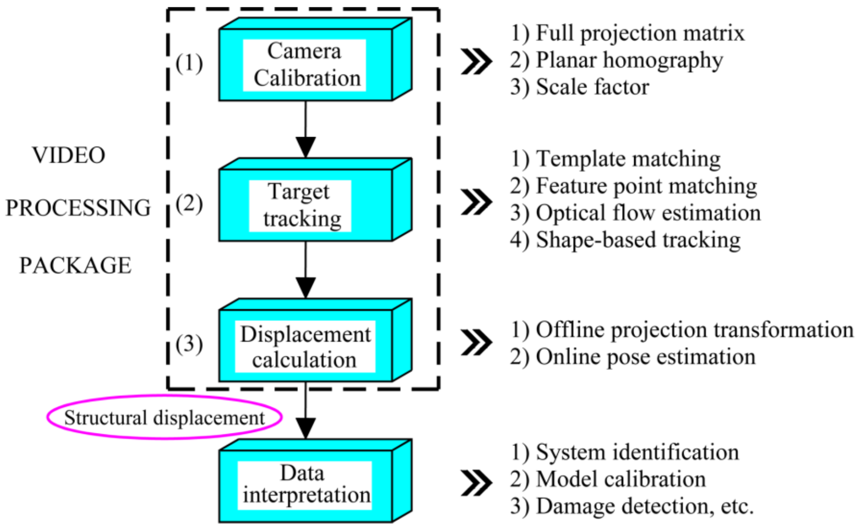
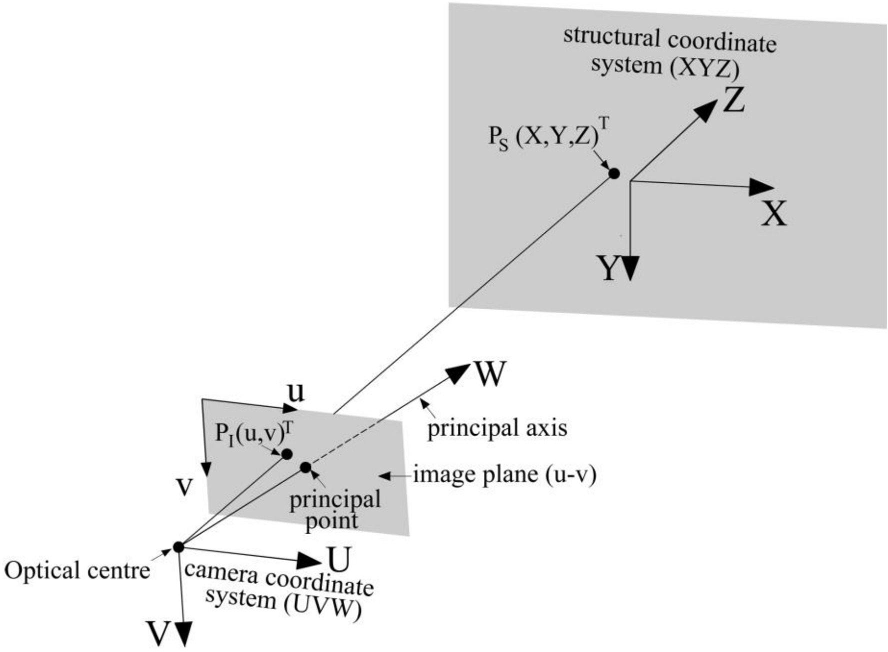


Fig. 2 Video processing procedures for structural displacement measurement and common methods in each step.

805 

806 Fig. 3 Camera projection model: central perspective projection.

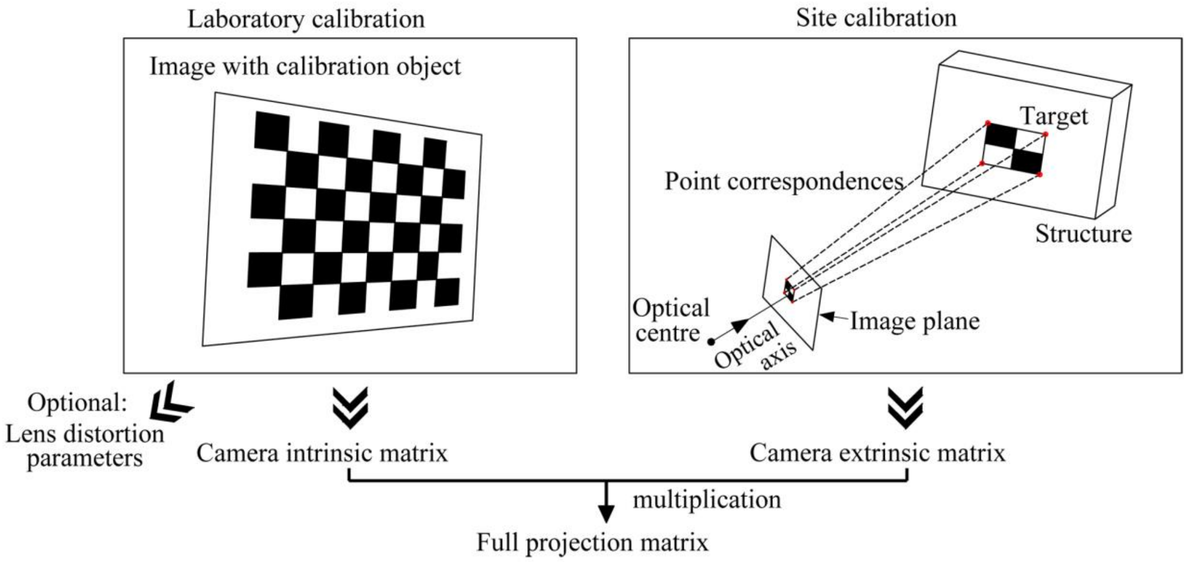


Fig. 4 Calibration steps for estimation of full projection matrix.

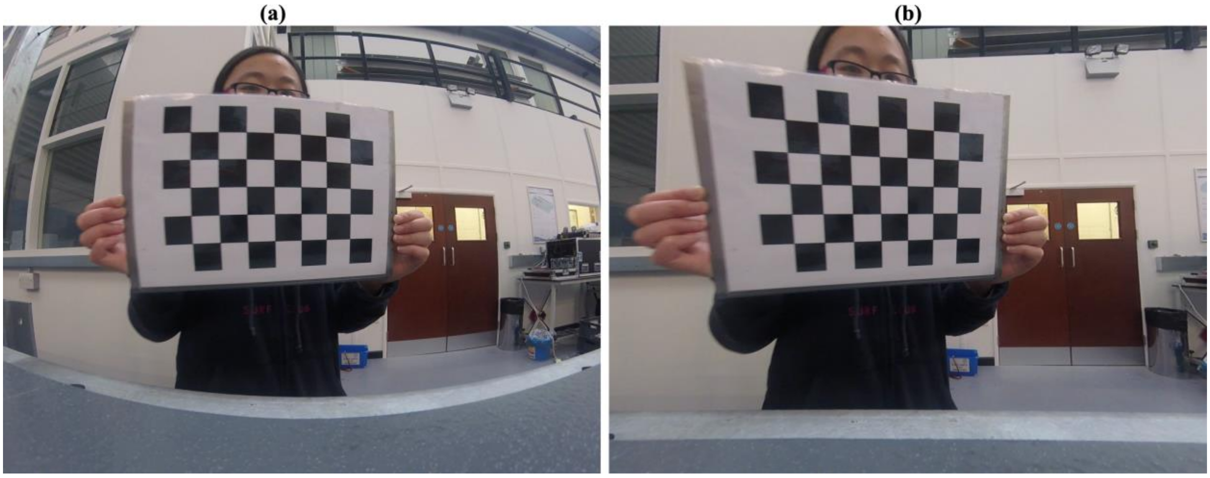


Fig. 5 Images of chessboard taken by GoPro Hero 4 Session camera: (a) raw image; and (b) image after distortion correction.

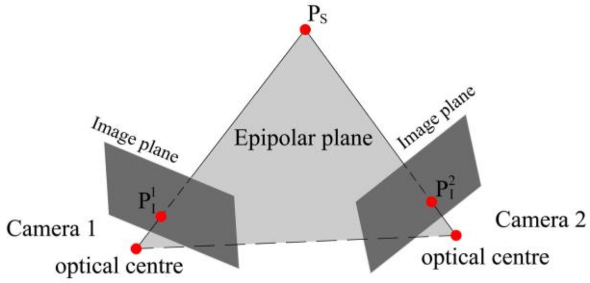


Fig. 6 Epipolar geometry principle of stereoscopic vision

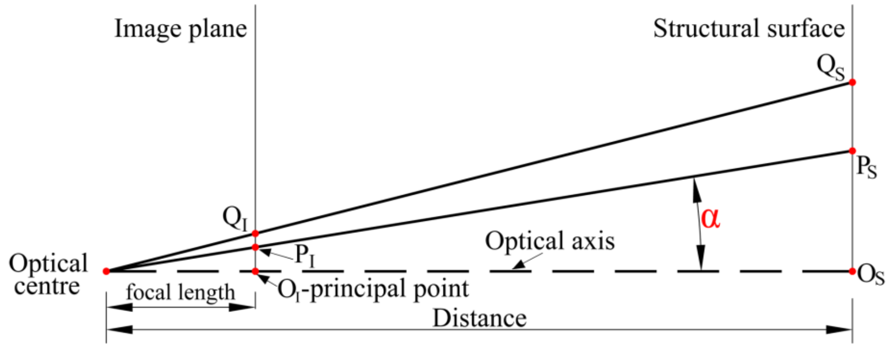
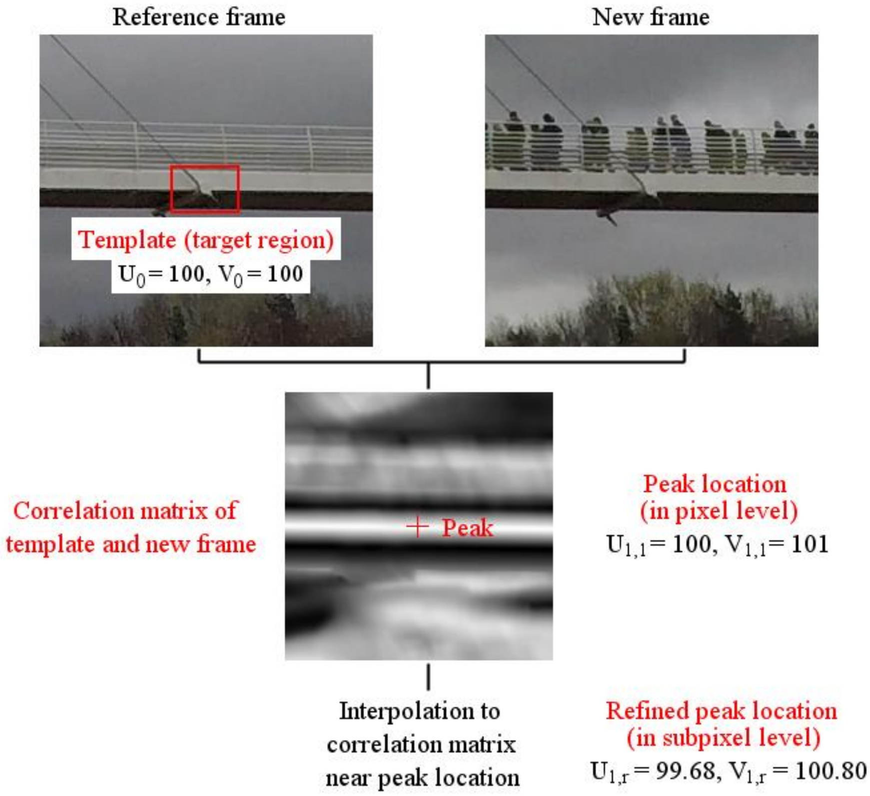


Fig. 7 Camera projection model when the optical axis of camera is perpendicular to the structural surface.

816 

817 Fig. 8 Procedures of template matching method for target tracking: The horizontal and vertical coordinates of the
818 target centre in the image plane are denoted as U and V, respectively; and the subscripts 0 and 1 represents the
819 image coordinates in the reference and new frames, respectively.

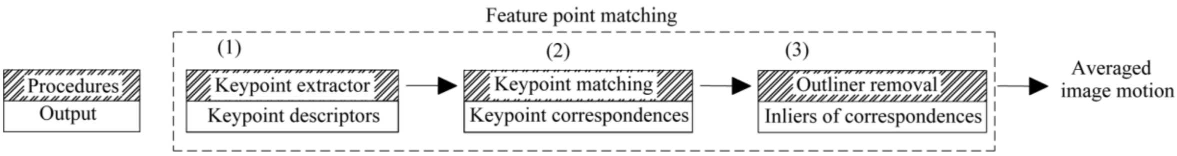


Fig. 9 Procedures of feature point matching for target tracking

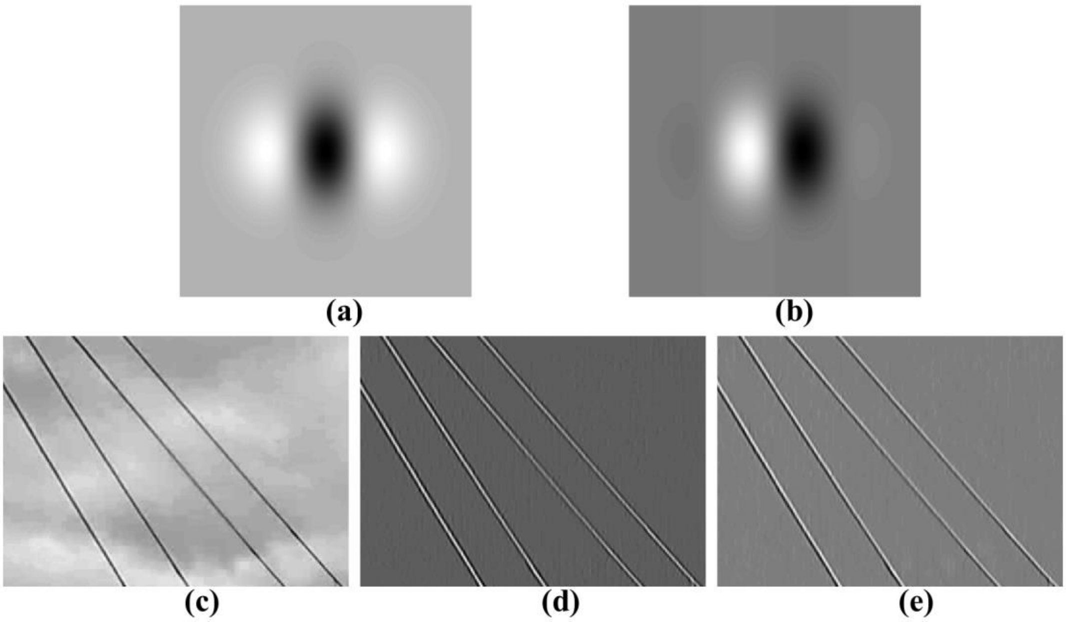


Fig. 10 Image after filtering by a quadrature pair of Gaussian derivative filters in the image width direction: (a) the real part of Gaussian derivative filters; (b) the imaginary part of Gaussian derivative filters; (c) the raw image of footbridge stay cables; (d) the real part of filtered image data; and (e) the imaginary part of filtered image data.