Vision-based dynamic monitoring of a steel footbridge

Journal of Physics: Conference Series **2647** (2024) 122010 doi:10.1088/1742-6596/2647/12/122010

Vision-based dynamic monitoring of a steel footbridge

E. Buoli[1], E. Bassoli[2], G. Eslami Varzaneh[3], F. Ponsi[4] and L. Vincenzi[5]

[1,2] Research assistant, Department of Engineering Enzo Ferrari, University of Modena and Reggio Emilia, Modena, Italy
[3] PhD Student, Department of Engineering Enzo Ferrari, University of Modena and Reggio Emilia, Modena, Italy
[4] Research assistant, Department of Civil, Enviromental and Materials Engineering, University of Bologna, Bologna, Italy
[5]Associate professor, Department of Engineering Enzo Ferrari, University of Modena and Reggio Emilia, Modena, Italy

E-mail: edoardo.buoli@unimore.it[1] elisa.bassoli@unimore.it[2] ghita.eslami@unimore.it[3] federico.ponsi@unibo.it[4] loris.vincenzi@unimore.it[5]

Abstract. The increasing slenderness and lightness of modern footbridges make the vibration serviceability assessment a key aspect for their design and maintenance. A promising approach to the characterization of footbridge dynamic behaviour is represented by computer vision-based techniques. In contrast to traditional monitoring systems relied on dense sensor networks, computer vision-based monitoring requires the installation of one or more cameras together with, if necessary, some targets on the monitored structure. The article presents some preliminary results of the short-term dynamic monitoring of a steel footbridge based on computer vision techniques. The structural deflection caused by a jumping pedestrian was recorded from three cameras in different measurement positions. The post-processing of the video frames is presented and discussed in the article. Special attention is paid to the use of circular targets placed on the footbridge, which allowed for the identification of deflections with sub-pixel resolution. A traditional accelerometer based monitoring system is also installed on the footbridge for validation purposes. Displacements evaluated through a double integration of the measured accelerations are compared to those obtained from the image processing. Results demonstrate the high potential of computer vision-based systems for the monitoring of structures and infrastructures.

1. Introduction

Structural monitoring applications such as health assessment and damage detection [1, 2], vibration serviceability assessment [3, 4], structural control [5] and model updating [6, 7] often require an in-depth knowledge of the structural dynamic behaviour. Traditional monitoring systems are composed of accelerometers that are able to record the response of a structure caused by different sources of excitation. Despite the large diffusion of these systems and the noticeable results obtained, some drawbacks can arise due to the necessity of an expensive and time-consuming installation of the sensors and of the acquisition system. Moreover, the spatial resolution of these systems is related to the number of sensors that can be installed on the structure. When a dense network is not employed, for example as for economic reasons, the

accuracy of the results is reduced, affecting, for instance, the damage localization process. On the other hand, displacement measurements can provide more useful and direct information about the structure under examination, since they can be used as an index of operational conditions and can allow a rapid detection of the permanent deformations.

However, displacement measurement is not an easy task for large scale structures. Already experimented contact-based (LVDT) or non-contact-based (high frequency GNSS stations) displacement sensors have a limited diffusion due to some drawbacks. The use of the former is limited because it is rare to have a stiff base where to install the sensors; the latter due to the lack of accuracy in the dynamic field and the cost of the instrumentation if several points need to be monitored.

A quite recent, innovative and promising approach is the use of non-contact-based computer vision techniques. The technological evolution of the video-cameras and the development of advanced image processing techniques allowed the application of the vision-based monitoring to different engineering fields as the monitoring of deformable structures. Indeed, recent cameras available on the market rely on high resolution and frame rates that are suitable for the monitoring purposes and the cost of instrumentation is considerably lower than that of the traditional systems. Moreover, only a camera and a series of targets applied on the structure in correspondence of the monitored positions are required [8–10]. In specific situations, it is also possible to exploit particular structural elements as targets [10–12].

The more recent applications of vision-based monitoring in the field of structural engineering are summarized in the review work of Zona [13]. Some researchers have tested such techniques in the laboratory on a steel frame placed on shaking tables [9]. Other researchers have performed monitoring tests on real structures such as steel pedestrian bridges [14–17], timber pedestrian bridges [18], cable suspension bridges [11], steel road bridges [19–21], reinforced concrete bridges [16, 22], steel structures for sports stadiums [23–25] and masonry structures [26, 27]. Vision-based techniques have also been employed to inspect structures for the detection of structural damages or the recognition of structural elements [10, 28] and to characterize the flow of a high-density crowd on footbridges [29]. However, some aspects of the vision-based monitoring are still challenging and need to be investigated in new applications.

The aim of this work is to develop and test a procedure for vision-based monitoring, The technique is tested on a steel footbridge. The paper reports the preliminary results obtained by the analysis of videos recorded by only one of the cameras installed to monitor the footbridge. To investigate potential and possible critical issues in a field experience of the procedure, the obtained displacements are compared to those obtained by a double integration of acceleration recorded by a traditional monitoring system installed on the footbridge.

The article is organized as follows. Section 2 describes the footbridge and the instrumentation employed for the monitoring. In section 3, the procedure for displacement detection is presented, while the corresponding results are highlighted in section 4, together with the comparison with a traditional system and some emerged critical issues. Finally, conclusions are drawn in section 5.

2. The case study

The case study is a cycling and pedestrian bridge (see Fig. 1) built in 2016 to allow the cross of the Panaro river between the localities of San Donnino and San Cesario sul Panaro, in the municipality of Modena (Italy). The structure is a 3D truss-girder steel footbridge with three spans and an overall length of 160 meters. The side spans are 45 meters long while the central span is 70 meters long. It has a box cross-section 3.00 m x 3.20 m, composed of truss girders with hollow tubular section elements (Fig. 1a). The whole structure is slender, lightweight and characterized by sensitive vibrations due to pedestrian, cyclist and wind effects. The instrumentation for vision-based monitoring consisted of three cameras (denoted with C1-

(a) (b)

Figure 1: (a) View along bridge axis and (b) Side view of the structure and its targets.

C3) and a number of targets composed of printed papers glued on rigid supports. The first camera (camera C1) is a Nikon D7500 and it acquired structural movements from the left riverside (Fig. 1a). The other two cameras (C2 and C3) were placed in the dry river bed (Fig. 1b).

Five cross-sections of the bridge were monitored and in each section a number of target was installed. Each target is composed of a single black circle on white paper in order to have a high contrast. Some targets were placed such as to be visible from the river side (as shown in Fig. 1a), while others were visible from the river bed (see Fig. 1b). All the targets have been positioned such as to ensure that all of them were clearly visible in the video frames.

Moreover, a traditional monitoring system composed of 6 biaxial MEMS accelerometers [30], was installed in three out of the five monitored footbridge sections for comparison purposes. The sampling frequency of the sensors was set to 40 Hz.

Footbridge vibrations were acquired by the cameras during the operational condition and in forced conditions. Results presented in the following involve the vibrations acquired during jumps of a pedestrian placed approximately in the midspan.

The jumping occurred with a cadence of 138 BPMs (Beats Per Minute) by following the beats generated by a metronome. The goal was to have an excitation frequency that was as close as possible to the natural frequency of the first vertical mode of the bridge equal to 2.32 Hz. This value was extracted from the acceleration response recorded during a previous ambient vibration test performed on the structure. The jumping session lasted about 15 seconds but the video recording continued for a few minutes before stopping the acquisition of the cameras, waiting for the motion attenuation. All the videos were acquired at 30 fps (frame per second) and using a 4K resolution.

3. Vision-based displacements detection

A video is a succession of frames visualized at high speed. Assuming that the camera is in a state of rest, a movement of the structure and thus of the targets installed on it can be perceived within each captured frame. The relative difference between the positions of the same target in different frames is related to the displacement of that point over time.

The identification of the structural movement needs a proper procedure that is able to find the displacements with the accuracy required for the monitoring purposes. In the following, the procedure used to obtain the vertical displacements from the acquired videos is illustrated.

The smallest element that composes a digital image is the pixel, so the image can be viewed as a pixel matrix. The position of each pixel that composes the image may be described by a couple of coordinates $[x_{\text{pixel}}, y_{\text{pixel}}]$ with reference to the global reference system, the latter having the origin in the left upper vertex of each frame. In a gray-scale image, an 8-bit pixel is associated with a unique color intensity value, within the range [0, 255].

Obviously, the pixel dimensions can be correlated with a physical dimension, for example knowing the size of an element in the image. A rough first consideration about the method accuracy may be that the minimum detectable displacement is equal to the minimal pixel dimension. However, for a reliable estimation of structural displacements, an effective procedure must detect movements with sub-pixel precision.

There exist several image processing techniques of edge and circle detection that are potentially able to reach a sub-pixel accuracy. The largest part of these techniques use the image intensity gradient as an indicator of the presence of an edge [31]. Focusing for example on the simple case of a black line on a white background in a 1D gray-scale image, if the line is not completely contained within a specific number of pixels, its edges are represented with an intensity value that assumes an intermediate value in the range [0, 255]. The progressive change of the intensity value moving from one pixel to another, quantified by the image gradient, gives information about the presence of an edge. In the case of a black circle on a white background, its radius and the position of its center can be estimated starting from the same consideration. The center position can be estimated with sub-pixel accuracy and, consequently, also its displacement between two subsequent frames.

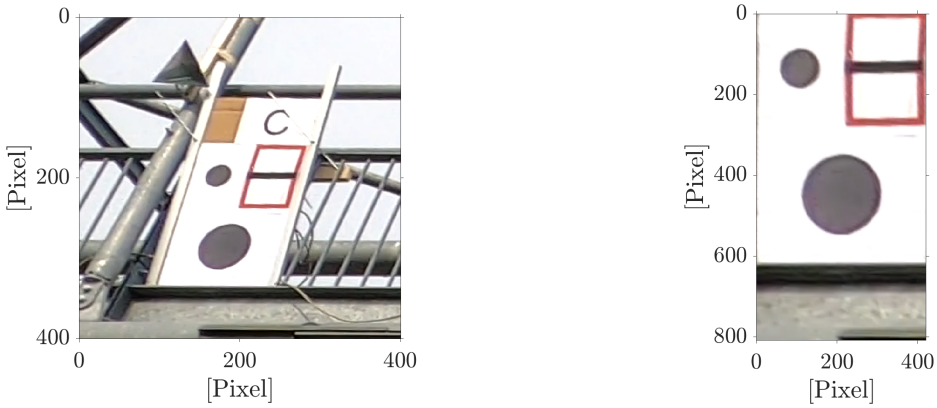
3.1. Pre-processing operations

The pre-processing procedure employs some of the tools that are available in the *Matlab Image processing and computer vision toolbox* [32] and that allow to import a video, extract frames and crop a rectangular region from a frame. The last operation is needed since the identification of circles printed on targets may be complex if operated over the entire frame. Indeed, different round contours, not related to the targets, could be present in the image. Each crop must be sufficiently wide to ensure the target visibility in all of the frames composing the video.

Depending on the position of the camera, different perspective deformations could be observed, resulting in an alteration of the circles in ellipses. The identification of distorted circles is generally problematic. For this reason, the procedure implements a correction of the perspective distortion of each crop. The correction is performed through the *Perspective Control/Correction* function [33]. An example of distortion removal can be seen in Figure 2. After these operations, each crop corrected from the prospective deformation has a new reference system that is centered in the new left upper corner.

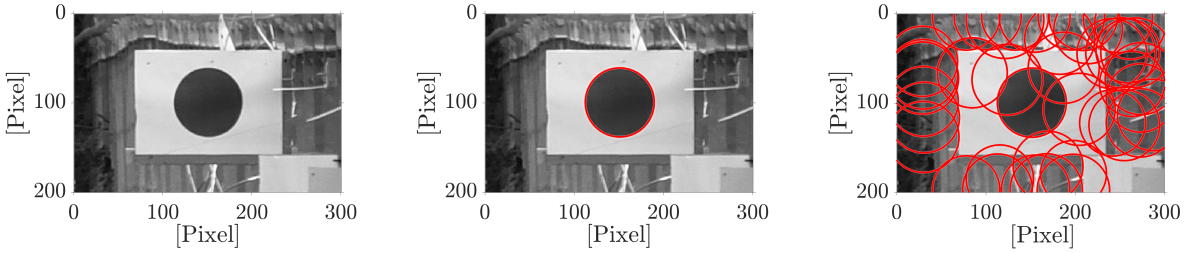
3.2. Circle identification and displacement computation

The identification of circles has been performed using the Circular Hough Transform (CHT) algorithm developed for finding circles in images [34, 35]. This method is used because of its robustness in the presence of noise, occlusion and varying illumination. The procedure identifies circles given a prescribed sensitivity. The latter is a scalar parameter whose value can range in [0, 1] and need to be tuned in order to find reliable results. In Fig. 3 the effect of the sensitivity value on the identification is shown. For a sensitivity value of 0.70 (and lower), no circle is identified (Fig. 3a). To increase the sensitivity up to 0.77 allows for a correct detection (Fig. 3b), while an higher value generates several false identifications (Fig. 3c). However, for frames with different conditions, the optimal sensitivity value can vary and be different from 0.77. For this reason, a procedure for an automatic sensitivity calibration has been defined. First, the number of circles to be detected is defined and the sensitivity is initialized to a low value. Then the sensitivity value is progressively increased with step-size of 0.01 until the



(a) (b)

Figure 2: (a) Perspective distortion of a circle and (b) applied correction.



(a) Sensitivity=0.70 (b) Sensitivity=0.77 (c) Sensitivity=0.99

Figure 3: Effect of changing the sensitivity value on the identification results.

correct number of circles is detected. The procedure produces a matrix containing the x and y coordinates of the centre of each circle and for each frame and a vector of the identified radius. The results obtained for each frame and crop are stored and post-processed only when all the frames have been analyzed. Indeed, assuming that the targets have more than one circle, the order of the identified circles is not guaranteed to remain the same for all frames. To solve the issue, a clustering operation is performed. The adopted clustering technique is based on the computation of the distances between the centres of the identified circles in a frame and those identified in a reference frame.

The obtained time series of displacements are expressed in pixels, but for monitoring purposes it is necessary to convert them into a metric scale. The conversion is performed by comparing the radius r of the circle printed on the target, measured in millimeters, and the mean values of the radius identified in all frames ($\bar{r}_{\text{id}}$) expressed in pixels. The conversion factor is thus computed as:

$$\eta = \frac{r}{\bar{r}_{\text{id}}} \tag{1}$$

4. Results

In this section some preliminary results are presented. In particular, displacement time series detected from the video acquired by the camera C1 with reference to the midspan section are analyzed. Fig. 4a reports the vertical displacement time history of a target located at the footbridge midspan. It is the result of the initial stage of the procedure and it clearly shows the presence of an unexpected trend. This behavior is attributed to possible vibrations of the

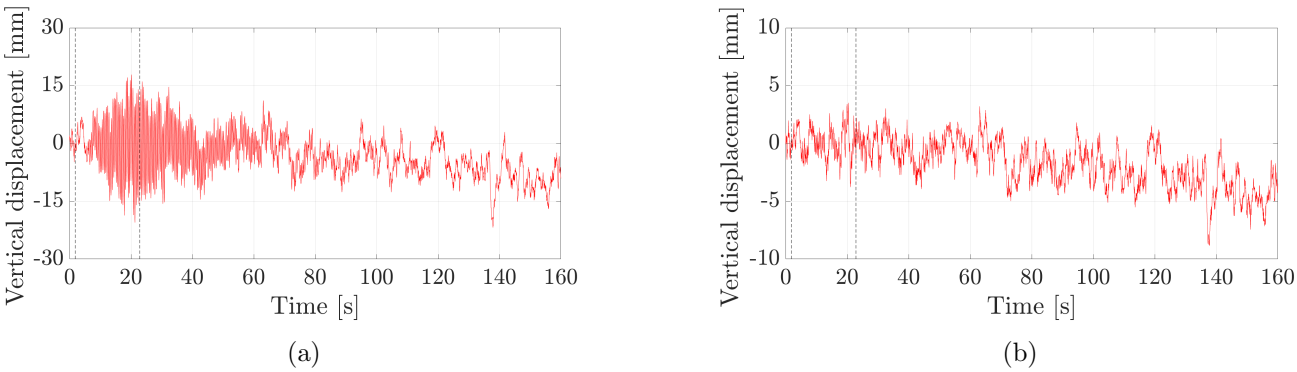
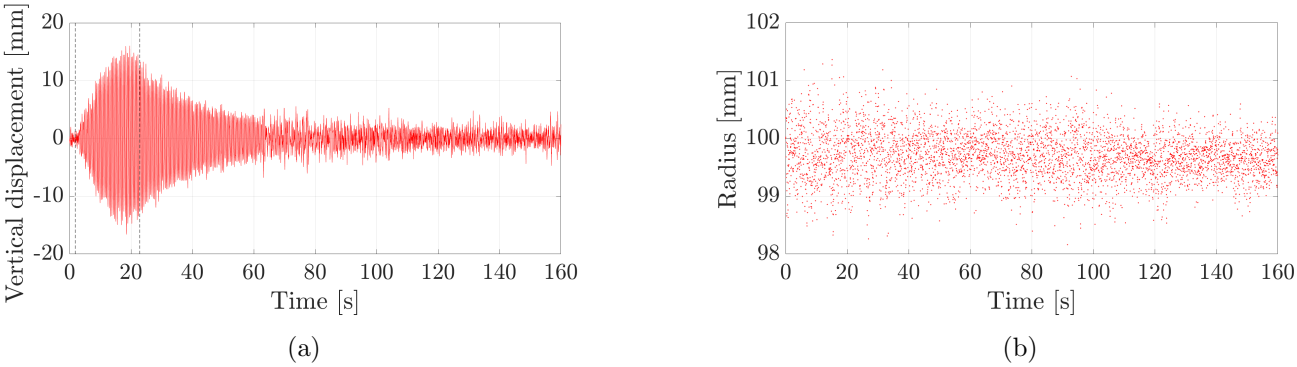


Figure 4: Detected (a) structural vertical displacement at the initial stage and (b) apparent movement of a fixed line at the riverside. Dashed lines: time instants corresponding to the start and to the stop of excitation.

camera that were not visible to the naked eye. Camera vibrations are confirmed by the fact that a fixed element, i.e. the line separating the terminal part of bridge deck from the vegetation at the riverside, slightly moves during the video. To obtain the final result show in Fig. 5a, camera vibrations have been identified and removed. The apparent vertical displacement of the fixed line has been evaluated through the detection of the line position in the frames and it is reported in Fig. 4b. To evaluate the line position, the frames are segmented on the basis of the Red Green Blue Vegetation Index (RGBVI), originally developed for the monitoring of agricultural biomass [36]. The RGBVI is the result of algebraic operations between R, G and B spectral bands of each acquired frame. The calibration of an appropriate RGBVI threshold allowed each frame to be binarized and consequently to separate the vegetation from other objects, obtaining a separation contour. In each frame, the separation contour is not always a straight line, because the location of the contour in each column of the digital image can have steps of a few pixels. For this reason, the line position is computed as the mean value among the contour locations in all the image columns. For each frame, the variation of the line position with respect to the value for first frame represents the apparent vertical displacement of the fixed line.

Once the camera movement of Fig. 4b is identified, the effect of this movement on the displacement measured in a specific section has to be properly evaluated depending on the camera-monitored section distance. Considering the midspan section, the measured displacement of Fig. 4a is adjusted accounting for the camera movement and, finally, a high-pass Butterworth filter is applied to remove the average trend from the time series. This allows obtaining the displacement time series of Fig. 5a.

From Fig. 5a, a non-negligible noise level is observed both before and after the forced vibrations. Noise can be partially attributed to the uncertainties in the circle identification and partially caused by other factors, including the procedure to adjust the perspective distortion and the uncertainties related to the procedure used to remove the camera motion. As for the circle identification, Fig. 5b shows the value of identified radius over time. It is shown that the radius oscillates between about 98.5 mm and 101.5 mm. Similar uncertainties can affect the determination of the circle center and thus the obtained displacements. Perspective correction can introduce some errors but it is necessary because the identification of circles is more complex if operated on frames affected by perspective effects. The more distorted the circle, the higher the sensitivity value required to identify it, with a consequent high risk to incur in false detection or add uncertainties in the radius and position identification. The perspective correction operated on each crop and frame [33] allows to identify the circles more easily, using lower sensitivity values.



(a) (b)

Figure 5: (a) Structural vertical displacement corrected by removing the movement of camera and (b) radius values of the target in the bridge middle section. Dashed lines: time instants corresponding to the start and to the stop of excitation.

4.1. Comparison with a traditional monitoring system

The results of the proposed vision-based monitoring technique are compared in this section with those obtained by a traditional monitoring system composed of accelerometers. The displacement time series has been reconstructed from the measured accelerations by adopting a method that combines Tikhonov regularization and an overlapping time window [37]. This method has proven to be efficient and accurate for the reconstruction of dynamic displacements in low frequency dominant structures. The comparison is shown in Fig. 6a. There is a good correspondence between the two time series but also remarkable differences. All the peaks are aligned in time but the maximum displacement does not occur at the same time instant. During the excitation phase of the structure, the mean absolute difference among displacement peaks, obtained by the two monitoring systems is 1.20 mm, with a maximum absolute difference of 3.40 mm (16.60 mm for the vision-based monitoring versus 13.20 mm for the reconstruction from accelerations). As concerns the noise level, the standard deviation of displacements obtained before the structure excitation by the vision-based technique is 0.81 mm, while that obtained by reconstruction from accelerations is 0.07 mm.

The comparison in the frequency domain is shown in Fig. 6b. The predominant peaks of the two series have approximately the same Power Spectral Density (PSD) function amplitude and frequency (2.31 Hz). The frequency of 2.31 Hz corresponds to the natural frequency of the first vertical bending mode. For the remaining peaks, there are differences in the PSD values. Moreover, some peaks of the PSD are hidden by the high value of noise and away from the peaks the PSD values show differences of about two order of magnitude.

5. Conclusions

In this article, a vision-based technique for the structural monitoring of deformable structures is presented. This technique is tested by performing a short-term monitoring on a steel truss footbridge. The monitoring systems consisted of three video-cameras installed close to the bridge and several circular targets attached to the structure. Displacement detection is based on a Circular Hough Transform algorithm, that enables the determination of the coordinates of the circle centre of the targets and their radii. The estimated displacement resulted affected by camera movements, that are removed by subtracting the apparent motion of a fixed line at the riverside. The structural displacements are then compared to those calculated from the accelerations measured by a traditional accelerometer-based monitoring system installed for comparison purposes. The comparison has revealed a general good matching but also some differences. In particular, a higher noise level is observed during ambient vibrations.

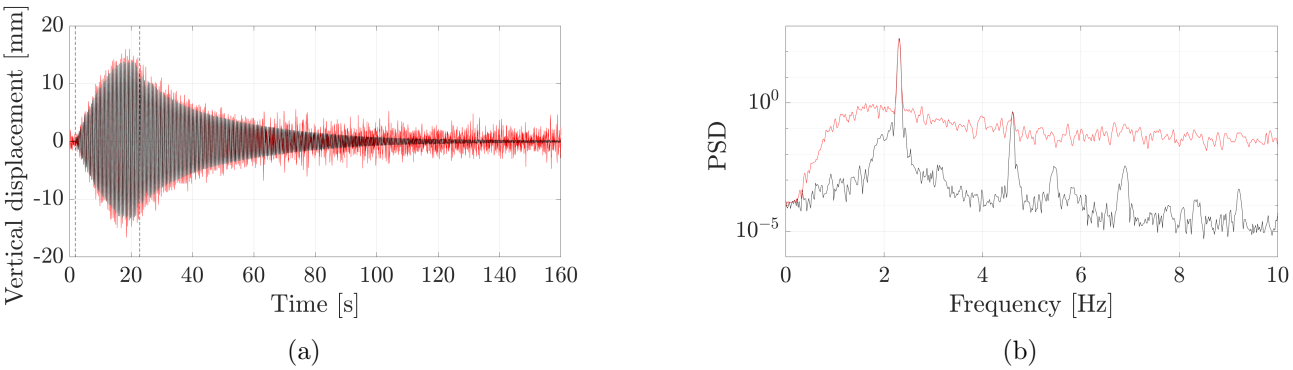


Figure 6: Comparison between monitoring systems in (a) time and (b) frequency domain. Red line: vision-based monitoring; black line: reconstruction from measured accelerations. Dashed lines: time instants corresponding to the start and to the stop of excitation.

Despite the underlined drawback, this result can be assessed as satisfactory considering the very low cost of the monitoring set-up. The presented research activity allows to find out the main factors influencing the accuracy of the vision-based technique. They are the size of the targets respect to the camera resolution, the presence of perspective distortions, the influence of the camera optical stabilizer and the need to subtract the effect of the vibration of the camera. The presence of perspective distortions and camera vibrations, both addressed with a specific approaches, are showed to be the most important issues to obtain reliable results. Future works will investigate different solutions, as the use of more effective designs and geometries for the targets and the use of proper designed target to be placed out of the footbridge to remove with more accuracy the possible camera vibrations.

To conclude, vision-based methodology for displacement detection in structural engineering is currently under development, but after some improvement it can represent an excellent resource for low cost structural monitoring, to be used in complement with traditional monitoring techniques.

6. Acknowledgments

This research was supported by the FARD-2022 Project for the structural monitoring with innovative techniques. The financial support of the University of Modena and Reggio Emilia and the Department of Engineering "Enzo Ferrari" is gratefully acknowledged.

References

[1] Comanducci G, Magalhães F, Ubertini F and Álvaro Cunha 2016 *Structural Health Monitoring* **15** 505–524
[2] Ponsi F, Bassoli E and Vincenzi L 2023 *Frontiers in Built Environment* **8**
[3] Van Nimmen K, Verbeke P, Lombaert G, De Roeck G and Van Den Broeck P 2016 *Journal of Bridge Engineering* **21**
[4] Bassoli E, Van Nimmen K, Vincenzi L and Van den Broeck P 2018 *Engineering Structures* **156** 537–547 ISSN 0141-0296
[5] Caetano E, Cunha A, Moutinho C and Magalhaes F 2010 *Engineering Structures* **32** 1082 – 1091
[6] Behmanesh I, Moaveni B, Lombaert G and Papadimitriou C 2015 *Mechanical Systems and Signal Processing* **64-65** 360 – 376
[7] Ponsi F, Bassoli E and Vincenzi L 2022 *Journal of Civil Structural Health Monitoring* **12** 1469–1492
[8] Kim S C, Kim H K, Lee C G and Kim S 2006
[9] Lee J J and Shinozuka M 2006 *NDT & E International* **39** 425–431
[10] Spencer B, Hoskere V and Narazaki Y 2019 *Engineering* **5**
[11] Caetano E, Bateira J and Silva S 2011 *Experimental Techniques* **35**

[12] Kromanis R 2021 *Characterizing Footbridge Response from Cyclist Crossings with Computer Vision-Based Monitoring* pp 83–95 ISBN 978-3-030-74257-7
[13] Zona A 2020 *Infrastructures* **6** 4
[14] Morlier J, Salom P and Bos F 2007 *Key Engineering Materials* **347**
[15] Xu Y, Brownjohn J and Kong D 2018 *Structural Control and Health Monitoring* **25** e2155
[16] Lydon D, Lydon M, Taylor S, Del Rincon J M, Hester D and Brownjohn J 2019 *Mechanical Systems and Signal Processing* **121** 343–358 ISSN 0888-3270
[17] Dong C Z, Bas S and Catbas F N 2020 *Engineering Structures* **224** 111224 ISSN 0141-0296
[18] Fradelos Y, Thalla O and Stiros S 2020 *Sensors* **20** 3217
[19] Feng D and Feng M Q 2017 *Mechanical Systems and Signal Processing* **88** 199–211 ISSN 0888-3270
[20] Chen J, Adams T, Sun H, Bell E and Büyüköztürk O 2018 *Journal of Structural Engineering (United States)* **144**
[21] Xu Y, Brownjohn J and Huseynov F 2019 *Journal of Bridge Engineering* **24** 05018014
[22] Harvey Jr P and Elisha G 2018 *Structural Control and Health Monitoring* **25**
[23] Khuc T and Catbas N 2016 *Structural Control and Health Monitoring* **24**
[24] Feng D, Scarangello T, Feng M Q and Ye Q 2017 *Measurement* **99** 44–52 ISSN 0263-2241
[25] Dong C, Celik O and Catbas N 2018 *Structural Health Monitoring* **18** 147592171880689
[26] Fioriti V, Roselli I, Tatì A, Romano R and De Canio G 2018 *Measurement* **129** 375–380 ISSN 0263-2241
[27] Acikgoz S, DeJong M J, Kechavarzi C and Soga K 2018 *Engineering Structures* **168** 544–558 ISSN 0141-0296
[28] Xu S, Wang J, Wang X and Shou W 2019
[29] van Hauwermeiren J, van Nimmen K, van den Broeck P and Vergauwen M 2020 *Infrastructures* **5**
[30] Guidorzi R, Diversi R, Vincenzi L and Simioli V 2010 *Proceedings of the 5th European Workshop - Structural Health Monitoring* pp 901–906
[31] Shrivakshan G and Chandrasekar C 2012 *International Journal of Computer Science Issues (IJCSI)* **9** 269
[32] The Mathworks, Inc. Natick, Massachusetts 2022 *MATLAB version 9.13.0.2080170 (R2022b)*
[33] Chan M 2023 Perspective control/correction matlab central file exchange URL `https://www.mathworks.com/matlabcentral/fileexchange/35531-perspective-control-correction`
[34] Yuen H, Princen J, Illingworth J and Kittler J 1990 *Image and Vision Computing* **8** 71–77 ISSN 0262-8856
[35] Atherton T and Kerbyson D 1999 *Image and Vision Computing* **17** 795–803 ISSN 0262-8856
[36] Bendig J, Yu K, Aasen H, Bolten A, Bennertz S, Broscheit J, Gnyp M L and Bareth G 2015 *International Journal of Applied Earth Observation and Geoinformation* **39** 79–87 ISSN 1569-8432
[37] Lee H S, Hong Y H and Park H W 2010 *International Journal for Numerical Methods in Engineering* **82** 403–434