Bridge Vibration Virtual Sensor Based on Eulerian Video Magnification and Gray Mean Difference

Jiayan Zheng1, 2, *, Qiumo Li2, Ruoyu Mao2, Yuzhou Wang2 and Zhixiang Zhou1, 2
1 State Key Laboratory of Mountain Bridge and Tunnel Engineering, Chongqing Jiaotong University, Chongqing, China
2 School of Civil Engineering, Chongqing Jiaotong University, Chongqing, China

*Corresponding author email: jiayanzheng@cqjtu.edu.cn

Abstract. A non-contact measurement method is proposed to overcome the shortcomings of traditional contact sensors in this paper. The camera is used as a non-contact virtual sensor of the whole field vibration. The motion in raw video was amplified by Eulerian Video Magnification technology and the gray mean of the selected region from processed video is obtained. The gray value difference between frames represents the dynamic response signal of the local area of the bridge structure under the excitation. The actual displacement amplitude in the time domain of the target structure can be obtained by conversion. A model bridge experiment was carried on, the results show that good coincidence in time domain and good accuracy in frequency domain, the error of the proposed method is 2.2% compared with the traditional measurement, indicating that the method without manual mark has the advantages of accurate, easy to obtain, low cost and high efficiency.

Keywords: Vibration; Bridge Structure; Eulerian Video Magnification; Gray mean difference; Virtual sensor.

1. Introduction
Vibration information is the key dynamic data of bridge structure and bridge health monitoring[1]. At present, the most conventional method is to obtain the relevant parameters (such as structural displacement, strain, curvature, vibration frequency, etc.) through the sensor network arranged on the bridge for analysis. The sensor has the advantages of mature technology, reasonable price and accurate data. However, the traditional sensor life is generally shorter than the service life of civil structure, and it is inconvenient to replace, and it will lead to the interruption of monitoring data. The traditional sensor is usually arranged in the key position of the structure, which can only realize single point measurement, and the sensor points in the whole field measurement are wide and complex in distribution. During the measurement process, it produces load effect due to contacting with the tested object directly[2]. For this, some new non-contact measurement methods have emerged, such as laser scanner, radar, vision measurement and so on. They can measure the position, size, shape and other parameters of the target without touching the measured object.

With the rapid development of photographic equipment and computer vision technology, the measurement based on machine vision becomes one of the most popular technologies because it does not need laser light source, magnetic field interference, complex interference light path and other auxiliary devices. There are a lot of researches on non-contact measurement of engineering structures. Wahbeh[3]measured the vibration displacement of large-scale bridges by arranging two laser points with fixed relative positions on the civil structure for tracking. Lee[4]introduced an advanced dynamic real-time displacement measurement system based on vision and the displacement of multi-layer
frame structure during vibration was analyzed in real time. Feng[5][6] studied the vibration signal of the bridge structure when the train passes through the railway bridge by finite element analysis and visual measurement, showing that the visual method has good robustness. Yang[7] used the high-speed video to extract the vibration of the cantilever structure under the impulse excitation, to obtain the displacement signals and estimated the modal parameter. Rao[8] analyzed the impact of underwater environment on the structure vibration and calculated the natural frequency and damping of the structure by extracting the amplitude information of underwater structure under the earthquake excitation. Zhou[9] proposed method based on SIFT point matching without artificial target to obtain the displacement time history of the target. Zheng[10] used edge operator to extract the contour of the vibration bridge in the magnified video, which proves the accuracy of the method to obtain the structural holographic shape data.

Most of these vibration measurements based on computer vision need to be manually marked on the tested structure in advance, such as arranging and pasting target points, LED point light source or laser point light source. Besides, it is hard to obtain structural deformation information directly from videos or images due to the tiny deformation of bridge structure, and the accuracy and reliability of identifying displacement only by pixel points on the edge contour line are low. For problems above, this paper proposed a method to extract structural displacement, frequency and other information without manual marking for a micro motion. Eulerian video magnification and image processing technology (marginal gray value method) were applied to quantify the vibration information of structure, and greatly reduced the error caused by noise and other reasons through pixel gray mean difference method. The validity and accuracy of the method are verified by comparing with the data from dynamic displacement sensor and vibration sensor in a self-anchored suspension bridge modal. This method provides a new non-contact measurement technology for NDT (Non-Destructive Testing) and SHM (Structural Health Monitoring).

2. Related Work

2.1. Eulerian Video Magnification (EVM)
Researchers from the computer science and Artificial Intelligence Laboratory (CSAIL) of Massachusetts Institute of Technology put forward a series of algorithms to amplify the weak change signals in the video, which can exaggerate the weak brightness changes and small object motion information in the original video. In 2005, Liu[11] proposed a method to enlarge the motion changes in video objects. This method tracks the motion track of feature points and enlarges their motion amplitude. It is a Lagrangian algorithm, and needs to fill the background, so the algorithm is very complex and time-consuming. Wu[12] proposed a technique called Eulerian video magnification, different from the Lagrangian perspective, the Euler perspective focus on the signal change of fixed pixel coordinates. In 2013, on the basis of Eulerian magnification algorithm, Wadhwa[13] proposed a phase-based image motion processing technology, which can translate the noise and achieve better amplification effect. The first step of EVM is to use image pyramid to filter the video sequence image in order to get the baseband of different spatial frequencies. This process helps to reduce noise and facilitate the approximation of image signals. The phase-based motion amplification algorithm uses the pyramid to decompose the video signal into the amplitude and phase of the local space group, the local space phase signal is decomposed into a series of spatial phase signals by Fourier transform, and the sine wave represents the harmonic motion. Then, the band-pass filter in the time domain is applied to each baseband to extract the part of change signal that we are interested in. Then amplify and recombine the changed signal, and finally synthesize the amplified video. The result is that the motion in the video is amplified within the specified band-pass range, while the signals in the rest band-pass range remain unchanged and are combined with the amplified signals into the amplified video. The result is that the motion in the video is amplified within the specified band-pass range, while the signals in the rest band-pass range remain unchanged and are combined with the amplified signals into the amplified video. The flow is shown in Figure 1.
Spatial filtering  

Time filtering  

Amplification And Composite video  

Deflection  

Input  

Output  

Time  

Deflection  

Figure 1. EVM processing flow.

2.2. Gray-scale Theory of Image

Digital images are often stored in RGB (red-green-blue) color mode measured by camera image sensor. Each of the three channels has 256 levels of brightness, each pixel can combine 16.77 million colors through 256 levels of brightness of each of the three channels.

In the amplification methods mentioned in the previous section, they are all completed in YIQ color space. Firstly, RGB is transformed into YIQ, then amplified, and then presented as RGB. Y component represents brightness (intensity) information of image, I component represents In-phase, color from orange to cyan, Q component represents Quadrature-phase, color from purple to yellow green (as shown in Figure 2). YIQ color space has the following linear relationship with RGB color space

\[
\begin{bmatrix}
Y \\
I \\
Q
\end{bmatrix} =
\begin{bmatrix}
0.299 & 0.587 & 0.114 \\
0.596 & -0.275 & -0.321 \\
0.212 & -0.528 & 0.311
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

(1)

![YIQ space](image)

Figure 2. YIQ colour space.

The pixel information contained in the micro motion is mainly concentrated in the Y channel in most cases, namely brightness information, and the colour has little impact on the results. Li[14] put forward the method of only amplifying the signal of Y channel, not amplifying the I and Q channels, and then combining the amplified Y channel signal with the original I and Q channel signal, which greatly improves the operation speed and hardly affects the amplification effect. He[15] converts the original video to grayscale and then enlarges the original video, the speed is significantly improved and the effect is good. In the gray value, 0 and 255 represent black and white respectively. This process can eliminate the hue and saturation information while maintaining brightness. For the amplification of motion, the color has almost no effect, so gray image signal can improve the efficiency and not affect the correctness of the results.

3. Method of Box Selection and Subtraction

Under the background of EVM and gray-scale image, a method based on the subtraction of gray mean of edge regions was proposed in this paper to obtain the required engineering physical quantities such as displacement and frequency from the video.
3.1. Gray Mean Obtaining in Selection Area
Firstly, the RGB image of video is converted to gray image. The part of first frame of video image was selected to calculate the gray mean within the selection range, and then the vibration parameters are obtained by the change of gray mean within the same frame selection region, as shown in Figure 3. The assumptions of this method are as follows:

- there are obvious differences between the tested object and the background color in the video.
- the vibration range of the tested object in the video should be within the selection range.
- assuming that the noise value generated by the image is the same in the short time, that is, the noise value in each frame after video decomposition is approximately equal, then the noise value can be offset by subtracting the gray mean of the front and back frames, and only the relative change of the gray value generated by the image due to the upper and lower vibration is retained.

Taking the model experiment in section 4 as an example. Assuming that the background is dark while the interested part is light, the observed structure moves only in the vertical direction. The gray mean $I_p(t)$ of selected areas change with the up and down vibration of the target structure. The gray mean of edge region is to average the intensity of a selection region from image edge, which can not only solve the problems of low recognition accuracy and large error caused by the blurry image edge contour, but also improve the signal-to-noise ratio of images.

$$I(t) = \frac{I_{\text{max}} - I_{\text{min}}}{L_p} x(t) + n(t)$$

(2)

where $I(t)$ is the gray value of selected area, $L_p$ is the length of the selected box, $n(t)$ is the noise of the selected area. $x(t)$ represents the position of $I$ on the x-axis over time. $I_{\text{max}}$ and $I_{\text{min}}$ are maximum and minimum pixel intensity of the box selection area. The length $W_p$ should be large enough to ensure that the maximum displacement amplitude be included in selected area, namely, the movement range of the main beam edge shall not exceed the selection area to avoid nonlinear.

3.2. Displacement Transformation
Due to the actual vibration has been amplified and there is an inclination angle between the camera (in figure 5) and the measured object in the measurement, there are errors in the test results, so it is necessary to consider the influence of amplification and geometric factors.

In order to establish the relationship between the gray value and the deflection of structures, the constant $B$ (mm / pixel) is introduced, which is realized by dividing the actual physical length $h$ (mm) of the frame selected part in the image by the counterpart of length $L_i$ (pixel) in video frame.
When the shooting direction is not perpendicular to the displacement direction, i.e. $\alpha \neq 0$, the geometric correction factor $C$ (dimensionless) is applied:

$$C = \frac{1}{\cos \left( \tan^{-1} \left( \frac{b}{a} \right) \right)}$$

where $a$ and $b$ are the horizontal and vertical distance between the camera and the target center respectively.

The vertical axis of the camera should be aligned with the vertical coordinate axis, and the angle $\beta$ (the horizontal angle between the camera and the structure center) can be ignored due to minimal effects on the vertical vibration test.

Considering that the Eulerian amplification technology has an amplification effect on the original video displacement, a reduction factor $K$ is adopted, whose value is the reciprocal of the Eulerian amplification factor. The target dynamic displacement $u_{act}(t)$ (mm) can be obtained:

$$u_{act}(t) = B \cdot C \cdot K \cdot I_p(t)$$

### 3.3. Noise Processing

The existence of image noise increases the error of the test results and reduces the test accuracy. In this paper, subtraction difference theory and pixel gray mean are used to reduce the influence of noise.

We introduced $n_p(t)$ to represent the noise value of selected area in this paper:

$$n_p(t) = \frac{1}{N} \sum_{i=1}^{N} n_i(t)$$

where $N$ is the total number of pixels in the selection area, and $n_i$ is the noise value in the pixel $i$, $t$ represents the video image at time $t$.

The gray value in equation (2) is also averaged, and its expression is

$$I_p(t) = \frac{1}{N} \sum_{i=1}^{N} I_i(t)$$

where $I_i$ is the gray value of pixel $i$.

Assuming that the noise is independent and the signal can be represented by a static process, the expression of the standard deviation of the pixel gray value can be obtained:

$$\sigma^2(I_p) = \frac{(I_{\text{max}} - I_{\text{min}})^2}{L_p^2} \sigma^2(x) + \frac{\sigma^2(n_p)}{N}$$

from the second item on the right side, it can be seen that with the increase of the number of pixels $N$, the noise decreases; that is to say, the smaller the standard deviation is, the more concentrated the numerical results are, the closer they are to the average value, indicating that the smaller the error caused by noise is.

We can replace $N$ with $W \times L_p$ (width and length of image selection area), the expression of SNR(signal to noise ratio) is as follows:

$$\text{SNR} = \frac{\frac{(I_{\text{max}} - I_{\text{min}})^2}{L_p^2} \sigma^2(x)}{\frac{\sigma^2(n_p)}{W \times L_p}} = \frac{W \times L_p}{I_p} (I_{\text{max}} - I_{\text{min}})^2 \frac{\sigma^2(x)}{\sigma^2(n_p)}$$
Equation 9 shows that in order to achieve the maximum SNR, it is necessary to maximize the \( W_p \) to \( L_p \) ratio, that is to say to keep the length as short as possible and the width as wide as possible, meanwhile that \( L_p \) must cover the maximum displacement amplitude. For a fixed camera and a short-time stable illumination condition, it can be assumed that the pixel noise power is constant, and an appropriate selection size can be specified. Another factor of SNR is difference between \( I_{\text{max}} \) and \( I_{\text{min}} \), which has a more significant effect on SNR. We can get a higher the SNR with greater the brightness difference between the target and the background.

4. Test on Suspension Model Bridge

In order to further explore and demonstrate the feasibility and applicability of non-contact detective method proposed in the last few section, a modal of self-anchored suspension bridge was chosen to test.

4.1. Test Overview

The test bridge is a scaled model bridge of Taohuayu Yellow River Bridge on Wuxi expressway, with a proportion of 1:30. The main span of the model bridge is 13.534m and the side span is 5.3m. The main girder is steel box girder, and the deck is consisted of prestressed concrete. Acceleration dial indicator are arranged in 1#-5# measuring points, as shown in Figure 6. The instrument of this test includes Sony FDR-AX700 camera, dial indicator, DH3818Y static acquisition instrument, DH5902 dynamic acquisition instrument. The layout of each instrument is shown in Figure 7.

![Figure 6. Laboratory suspension bridge. (mm).](image)

![Figure 7. Experimental instrument layout.](image)

In order to make the experiment more effective and less difficult, the good color contrast between the girder structure and the background, i.e. the \( (I_{\text{max}}-I_{\text{min}}) \) value is large, is benefit to capture the vibration of the structure. Meanwhile, most of the experiments are carried out under ideal conditions, which reduces the difficulty. Considering the characteristics of small stiffness and large flexibility of the suspension bridge, only a vertical impact is applied to the side span as the excitation to produce free vibration until the structure tends to be stable under the damping effect.

In this test, dial indicator and DH3818Y catch the deflection of main beam, and DH5902 collected the fundamental frequency information. Sony FDR-AX700 camera was used to obtain the image data, the frame rate is 100 fps (frames per second), which can meet the requirements of the vibration signal acquisition of the structure. The shooting direction is orthogonal to the side of the structure, mainly
shooting the vibration of the main beam in the mid span area (around measuring point 3#), so as to reduce the error caused by the angle effect, and an appropriate shooting distance was selected to avoid major lens distortion.

4.2. Data Pre-processing

4.2.1. Gray processing and amplification processing. The video is processed by Eulerian amplification and the clear holographic deformation is obtained. Then the video of suspension bridge is decomposed into frames, after that, the RGB images of these frames were converted into gray image. One of the gray frames is shown in Figure 8. (a) is raw image before amplification, and (b) is after amplification. The obvious effect of magnification can be seen from the dotted line in (b).

![Figure 8. Before and after magnifying in the same frame.](image)

4.2.2. Box selection and mean value acquisition. The experimental image data is processed (selection-decomposition-mean processing) automatically by the program written in MATLAB. During the vibration of the main beam, the variation range of beam is within the range of selection(figure 9), so as to avoid introducing non-linear data and signal interception. Different selection region can be set to extract displacement of different parts of bridge. Each selection area can be considered as a "virtual vibration sensor” mounted on the main beam to provide dynamic signals for the structure. The function is equivalent to the traditional sensor, which can obtain the dynamic response of the structure, and its acquired signals are applied to time-frequency analysis to get the basic frequency. Frame 1, frame 5, frame 10, frame 15 and frame 20 are selected for display.

![Figure 9. Box selection and decomposition.](image)

4.2.3. Acquisition of actual displacement. For connecting the gray value with the actual displacement, two steps are taken: (a) signal synchronization in the time domain, (b) the parameter $B$, $K$, $C$ was taken into account to calculate the true displacement. The vertical displacement is a sensitive index under the excitation, firstly, the variation can be obtained by subtracting the gray mean of two adjacent frames. Because it is a preliminary exploration, the error caused by the shooting angle should be minimized as much as possible. At this time, the position of the experimental camera was orthogonal to the shooting surface, the lens was aimed at the center of the target, and the height of the lens and the shooting target were the same, the geometric factor $C=1$. Then the displacement signal is transformed into displacement by equation 5. In the last, the deflection data from the dynamic dial indicator was acquired as standard. Taking the peak value of two sets of data as a unified index (figure 10), the time delay between two sets of data is calculated, and the time axis coordinate is unified accordingly. The correlation between the two groups of data was analyzed.
4.3. Analysis and Result

4.3.1. Displacement analysis. The displacement measured and gray value after unified coordinates are shown in Figure 11. It can be seen that the displacement curves from camera and dial indicator(3#) match well in the larger amplitude stage. With the attenuation of the amplitude, the error increases gradually, which may be caused by the image amplification. The signals details of the camera are less than the displacement data of the dial indicator, and had more noise. The result of correlation analysis of the two groups is shown in the right side of Figure 11. The figure illustrates that the correlation between the camera data and the dial indicator data is 0.9797(close to 1), indicating the high correlation between the two groups of data.

4.3.2. Frequency acquisition. The dynamic parameters (fundamental frequency) can be obtained by the time history data of acceleration and displacement through Fourier transform. The following content (figure 12) shows the structure frequency value after FFT of the displacement or acceleration time history data obtained in three ways: gray mean from video (a), dial indicator (b) and acceleration sensor (c). The fundamental frequency obtained by the three methods are summarized in Table 1. The fundamental frequency of the structure obtained by the camera method is 1.80Hz, the value of the dial indicator is 1.76Hz after Fourier transform, and the measured value of the vibration sensor is 1.76Hz. The difference is 0.04Hz, the relative error is only 2.2%. The results demonstrate that the edge area gray mean processing method proposed in this paper is able to obtain the displacement and fundamental frequency with good accuracy and feasibility. The reason of the error may be the image accuracy, and noise introduced by amplification processing. Later, the higher resolution camera and more appropriate magnification factor will be adopted to increase accuracy.
5. Discussion and Limitations

According to the low-frequency vibration characteristics of bridge structure, the specified band-pass range signal is amplified by EVM algorithm, and the method of subtracting of pixel gray value in edge area is proposed. The displacement of structure is obtained by the way of difference of gray value in adjacent frames, which can effectively obtain the change information of structure and avoid random error from single pixel, and also reduce the impact of noise in video image information. Compared with the traditional measurement, the high-speed camera is cheaper and more convenient. At the same time, the measurement results also have good accuracy, which meets the requirements of engineering measurement.

The experiment is carried out in the ideal indoor environment. The indoor light is stable, the camera almost has no jitter, the shooting direction is orthogonal to the beam elevation and the horizontal height is the same. Based on the three assumptions proposed in the third section, the frame selection subtraction method is adopted. In the actual outdoor measurement, due to the change of sun position or cloud illumination and background conditions, the wrong signal may be introduced. Secondly, in this experiment, only a part of the beam is photographed, and an area is selected as a single sensor for analysis. To obtain the overall information of the structure, the shooting distance needs to be larger, and the area selected by the frame needs to be more. In addition, sampling rate, resolution, noise, image sensor quality, size, lens type and other factors will affect the accuracy of measurement. In this case, we must pay attention to avoid too many changes of video sequence. In the actual outdoor measurement, the influence of shooting angle and light is great, it is necessary to carry out further research for applying this method to the long-term monitoring of structures.

6. Conclusion

In this paper, a new method is proposed to obtain bridge vibration information by Eulerian micro motion amplification technology and gray mean difference method. A high-speed camera is used as a non-contact virtual sensor of full field vibration. The displacement time history curve of the structure
is obtained by the virtual sensor, and the fundamental frequency of any point of the structure is calculated by FFT. The results of image data are compared with the measured results from traditional sensor, including the transformed displacement curve and the foundation frequency of the structure. The results show that the displacement curves are in good agreement and the accuracy of fundamental frequency is high. Therefore, this work provides a new non-contact measurement method for the use of camera as a remote virtual sensor to measure the relevant parameters, and makes a preliminary exploration and preparation for further structural state evaluation.

Acknowledgments
The authors acknowledge the support provided by NSFC (51778094; 51708068). We also acknowledge State Key Laboratory of Mountain Bridge and Tunnel Engineering for providing experimental equipment.

References
[1] Ji Y T, Lei Z, Ran Z H and Jia Y 2018 Research on dynamic analysis theory for bridge structure based on experimental modal Highway no.03 p78-83
[2] Zhang D S 2017 Research and applications on vision-based structural motion extraction algorithms University of Science and Technology of China
[3] Wahbeh A M, Caffrey J P and Masri S F 2003 A vision-based approach for the direct measurement of displacements in vibrating systems Smart Materials and Structures vol 12 p785-794
[4] Lee J H, Ho H N, Shinozuka M and Lee J J 2012 An advanced vision-based system for real-time displacement measurement Smart Materials Structures vol 21 125019.
[5] Feng D M, Sun H and Feng M Q 2015 Simultaneous identification of bridge structural parameters and vehicle loads Computers and Structures vol 157 p76-88
[6] Feng D M and Feng M Q 2015 Vision-based multipoint displacement measurement for structural health monitoring Structural Control and Health Monitoring vol 23 p876-890
[7] Yang H, Takaki T and Ishii I 2011 Simultaneous dynamics-based visual inspection using modal parameter estimation Journal of Robotics and Mechatronics vol 23 p180-195
[8] Rao R G V, Sreekala R, Kumar S K, Gopalakrishnan N, Muthumani K, Iyer N R, Lakshmanan N and Reddy G R 2016 Seismic response measurement of an under-water model through high speed camera and feature tracking Experimental Techniques vol 40 p83-90
[9] Zhou Y, Zhang L X, Liu T and Gong S M 2018 Structural system identification based on computer vision China Civil Engineering Journal vol 51 p17-23
[10] Zheng J Y, Mao R Y, Wu T, Zhou Z X and Tang L 2019 Eulerian-based micro motion amplification method to measure holographic deformation form of bridge Science Technology and Engineering vol 19 p347-354
[11] Liu C, Torralba A, Freeman W T, Durand F and Adelson E H 2005 Motion magnification (Los Angeles: SIGGRAPH) p519-526
[12] Wu H Y, Rubinstein M, Shih E, Guttag J, Durand F and Freeman W 2012 Eulerian video magnification for revealing subtle changes in the world ACM Transactions on Graphics vol 31(4) p1-8
[13] Wadhwa N, Rubinstein M, Durand F and Freeman W T 2013 Phase-based video motion processing ACM Transactions on Graphics vol 32 p1-10
[14] Li L P, Lin L, Sun S F, YIN H and Dong F M 2015 Improved video small motion magnification processing Computer Engineering and Applications vol 51 p195-200
[15] He Y M and Chen X Y 2019 Gray image magnification method for amplifying small motions in the video Computer and Digital Engineering vol 47 p2022-2026