Measurement 253 (2025) 117440

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

The optimal scale level of complex steerable pyramid for phase-based motion estimation under different motion ranges and target sizes

Sida Ai [a], Chuanzhi Dong [b], Qipei Mei [a,*]

[a] Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta T6G 2W2, Canada
[b] Department of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA

ARTICLE INFO

Keywords:
Phase-based motion estimation
Complex steerable pyramid
Motion range
Target size
Optimal scale level
Measurement strategy

ABSTRACT

In recent years, the phase-based motion estimation method has received significant attention in the field of vision-based motion estimation due to its robustness under illumination variation and high subpixel accuracy. The complex steerable pyramid (CSP) is widely adopted to generate phase of frames for motion estimation. The scale level of CSP is a significant influential parameter that affects the displacement measurement performance. However, systematic study regarding the optimal scale level of CSP in relation to motion video's properties (motion range and target size) has not been carried out. Understanding this relationship could be helpful to identify the optimal scale level based on these properties, rather than simply comparing measurement results with the ground truth. In our work, a series of numerical motion videos with different properties are employed to find the optimal scale levels of CSP for motion estimation. Based on the phase-based motion estimation results, we found that the optimal scale level has clearly positive relationships with both the motion range and target size of motion video. Leveraging these relationships, a measurement strategy is proposed to automatically select the optimal scale level without the need to know ground truth of motion. The proposed strategy is further verified through a series of laboratory experiments, including shaker tests on a column and impact tests on a four-floor frame structure, using phased-based method alongside conventional sensors such as laser displacement sensor and accelerometers.

1. Introduction

In the field of structure health monitoring, the displacement of structure is a very important signal. It can be used for studying a structure's dynamic properties, assessing load levels, and evaluating its health conditions [1]. The traditional displacement measurement strategy includes two types: contact and non-contact. The contact type methods for displacement measurement are linear variable differentia transformers (LVDT) and accelerometers [2,3]. For these methods, the major limitation is that additional mass or force is added to the structure, which may cause unreal structure dynamic properties and load assessment. In contrast, the non-contact type methods can avoid limitations from contact type methods. The non-contact type methods for displacement measurement include the Global Position System (GPS) based method. However, it cannot provide enough accuracy, with the level of accuracy of centimeters [4]. Laser Doppler Vibrometer (LDV) is an accurate non-contact measurement strategy; however, the cost of such measurement is high [5].

Recently, with the development of computer vision techniques and computation capabilities, the vision-based methods for displacement estimation have received increasing attention [1]. The vision-based methods, a non-contact approach, offers low cost and high measurement accuracy and robustness when paired with suitable computer vision algorithms [6,7]. In general, the algorithm for vision-based motion estimation can be divided into two groups, the intensity-based method and the phase-based method [8]. The intensity-based method includes template matching, feature point matching, digital image correlation (DIC) and intensity-based optical flow [1].

Template matching slides an region of interest (ROI)-sized window across the image, computing a similarity metric at each position; the location with the highest score is deemed the target. Brownjohn et al. [9,10] employed normalized correlation for displacement measurement on large bridges. Dong et al. [11] used an edge-based normalized cross-correlation for improved robustness. Upsampled cross correlation (UCC) further refines subpixel accuracy [12]. Nevertheless, template matching, originally a pixel-level method, struggles at high subpixel precision.

* Corresponding author.
E-mail address: qipei.mei@ualberta.ca (Q. Mei).

https://doi.org/10.1016/j.measurement.2025.117440
Received 22 October 2024; Received in revised form 18 March 2025; Accepted 28 March 2025
Available online 30 March 2025

Feature point matching locates a target by detecting and describing feature points within the region of interest (ROI), then matching these points frame by frame [1]. Huang et al. [13] proposed a method that incorporates spatial smoothness constraints to enhance the matching accuracy of traditional descriptors like Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF). Khuc and Catbas [14] also employed the SIFT feature detector and descriptor to do feature matching and monitored the displacement of a pre-stressed transit guideway under train loads. Hu et al. [15] implemented an oriented brief (ORB) detector and descriptor feature matching to monitor the displacement of a viaduct. However, the performance of the feature point matching method is highly dependent on the number of detected features, extraction thresholds, feature visibility, and lighting.

DIC is an advanced template-matching technique in which full-field measurement is achieved by using a high-contrast speckle pattern on the surface [16]. It uses a speckle pattern on the surface to track movement and calculate displacement and strain across the object. This non-contact technique provides precise, full-field data for various materials and structures under different conditions. Helfrick et al. [17] used a 3D DIC method to measure the full-field vibration of a structure in the laboratory. Hagara et al. [18] applied the DIC to identify the nature frequency and corresponding modal shapes of thin metal sheets. For effective use of DIC, the application of a speckle pattern on the material's surface is required, but this preparation process can be time-consuming and labor-intensive.

The intensity-based optical flow calculates full-field velocity based on constant intensity and small-motion assumptions, but encounters the aperture problem [19]. The Lucas–Kanade (LK) method [20] addresses this by assuming uniform motion in small neighborhoods, and its extension Kanade–Lucas–Tomasi (KLT) enables feature point tracking [21]. Bhowmick et al. [22] computed the optical flow based on the edges information to calculate displacements and modal parameters of a frame structure in laboratory efficiently. Lv et al. [23] conducted vibration measurements with the optical flow under a large rotational motion. Hoskere et al. [24] implemented LK optical flow and Shi-Tomasi corner to process the video captured from a UAV and measured the displacement of fiducial markers installed on a suspension bridge. The intensity-based optical flow method is highly sensitive to changes in illumination during displacement measurement, primarily because it operates under the assumption of constant intensity.

Besides intensity, phase constitutes another significant property of images, representing the positions of wave components within their cycles. The characteristic of phase is essential for feature detection and motion estimation, as it exhibits reduced sensitivity to lighting variations and adeptly captures the structural details of an image [25,26].

The importance of the phase information of the image for feature detection begins from knowing phase changes and the Mach band transition relationship [27]. Then, researchers began to use phase congruence for edge detection [25,27]. In 1990, Fleet et al. [26] firstly proposed phase-based pixels velocity estimation for video. Compared to traditional intensity-based motion estimation methods, phase-based method has better subpixel accuracy and robustness against illumination changes [26,28]. In 2002, Gautama et al. [29] used a group of 2D Gabor filters to compute the full field velocity of an image sequence, instead of using the spatial–temporal Gabor filter proposed by Fleet et al. [26]. This practical use of phase-based motion estimation had not been developed until Wadhwa et al. [30] used the phase information and complex steerable pyramid (CSP) for motion magnification. Since then, phase-based motion estimation has been used in the field of structure displacement monitoring with the phase generated by the CSP method [31,32]. The subpixel accuracy of phase-based optical flow was further studied and verified in laboratory experiments, which confirmed the high accuracy of phase-based motion estimation [33,34]. Researchers also used the phase-based motion magnification method for structure modal identification and damage detection [35,36]. Miao et al. [37] used the 2D Gabor filter to generate the phase for the displacement measurement and modal identification of a frame structure and found that the nonlinearity of phase has a strong relationship with the accuracy of motion estimation. Some new ways for phase generation were also explored. Wang et al. [38] proposed a log Gabor filter to generate the phase for motion estimation, which can avoid the direct current (DC) component of Gabor when it has large bandwidth. Liu et al. and Li et al. [39,40] proposed the Hilbert transform method for phase-based motion estimation and Riesz transforms for two-dimensional phase-based motion estimation.

Among all those phase generation methods, the CSP method is widely used for phase-based structural displacement estimation and motion magnification in practice, attributable to its properties of translation invariance and self-inverting [41,42]. To be specific, the phase-based optical flow operated by CSP has been widely used for displacement estimation in field experiment in civil engineering field [43–45]. Moreover, the accuracy estimations of phase-based optical flow were also conducted based on CSP [33,34]. The phase-based motion magnification based on CSP was used for modal shape identification of structures [31,35]. Peng et al. [46] have investigated the impact of filter types and parameters of CSP on the measurement accuracy of phase-based optical flow. However, parameter selection of the CSP for better motion estimation performance by considering the properties of motion video has not been systematically studied. CSP method functions by decomposing the image into different orientations and scale levels. When employing the CSP method for image feature detection, the scale level is chosen based on the scale of the target features intended for extraction from the image. However, for video motion estimation, the important properties of a video include target size and motion range. Analogous to the significance of selecting the appropriate scale level of CSP for image feature detection, whether the scale level of CSP should be carefully selected to suit the important properties of motion video for optimal motion estimation still needs detailed investigation.

Therefore, in this work, a series of numerical motion videos with different properties are employed to test the performance of the scale level of CSP for motion measurement. The optimal scale level exists for the measurement task of a certain motion video. Strong trends of optimal scale level for motion estimation with respect to both the target size and motion range of videos are found. To be specific, the optimal scale level is positively correlated with both the target size and motion range. Based on the numerical test results, a 2D contour map of the optimal scale level with respect to the target size and motion range is generated, leading to the proposal of a motion measurement strategy for achieving optimal phase-based optical flow measurement performance using CSP. This strategy optimizes measurement performance by considering the motion range and target size of motion video, even in the absence of ground truth motion data. Finally, this measurement strategy is systematically verified in a series of vision-based laboratory experiments.

The paper is organized as follows: Section 2 focuses on introducing the decomposition process of the CSP method, the conception and computation process of phase-based optical flow, and the role of scale level within the CSP method. Section 3 presents the setup and results of motion estimation experiments based on numerical motion videos to explore how the optimal scale level of CSP changes with target size and motion range. This section also proposes a 2D contour map to visualize the relationship between the optimal scale level, target size, and motion range, leading to the development of a measurement strategy that aims to automatically optimize the measurement performance. Section 4 validates this measurement strategy through laboratory experiments involving eight motion videos, confirming its effectiveness by comparing the deduced optimal scale levels from measurement strategy with actual optimal scale levels. Finally, Section 5 further verifies the efficiency of the proposed strategy in the dynamic property estimation of a four-story frame structure.

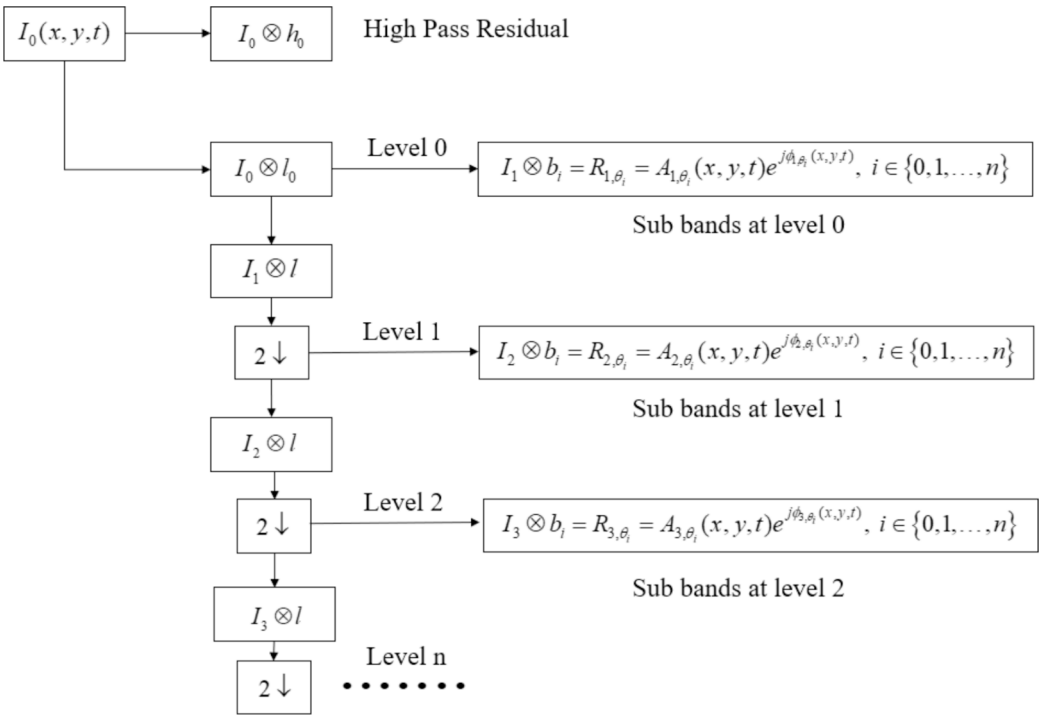


Fig. 1. The decomposition process of CSP.

2. Methodology

2.1. The decomposition process of CSP method

The CSP is an extension of the steerable pyramid and utilizes the Hilbert transform pair of real filters from the steerable pyramid as the complex component of the CSP filters [42,47]. The structure of the steerable pyramid resembles the discrete wavelet transform (DWT) method to some extent, as both employ high pass filters, low pass filters, and a downsampling process to decompose the original signal [48]. However, a unique aspect of the steerable pyramid method is its incorporation of steerable filters, which are applied to the lowpass sub-bands to capture the directional details of the image. The steerable pyramid is designed to be free of aliasing in the frequency domain, ensuring comprehensive coverage across all directions and frequencies without any frequency aliasing [42]. It should be noted that the translation invariance property of CSP with respect to any scale levels and directions naturally makes it a suitable tool for motion estimation[41].

Fig. 1 illustrates the process of utilizing the CSP to decompose a video and extract phase information for motion estimation. In this figure, h_0 is a high-pass filter, l_0, l are low-pass filters, and each b_i is a steerable filter. Users can decide the number of orientations for the steerable filters. Those orientations divide 360 degree evenly. However, the over-completeness of the CSP is linear with the number of orientations [42].

The variables $R_{i,\theta_i} = A_{i,\theta_i}(x,y,t)e^{j\phi_{i,\theta_i}(x,y,t)}$ represent the decomposition results of the CSP across various scales and orientations. As R_{i,θ_i} is a complex number, the phase angle $\phi_{i,\theta_i}(x,y,t)$ derived from R_{i,θ_i} can be utilized to compute the phase information necessary for motion estimation.

In the initial stage of decomposing a frame $I_0(x,y,t)$ from a motion video, the frame undergoes convolution with a high-pass filter h_0 to extract the high-frequency residuals. Concurrently, I_0 is convolved with a low-pass filter l_0 to derive the low-frequency band I_1 at scale level 1. Subsequently, I_1 is subjected to convolution with a set of steerable filters b_i, yielding the orientation decomposition results R_{1,θ_i} of CSP in scale level 1. The process continues by convolving I_1 with another low-pass filter l followed by downsampling to obtain the low-frequency band I_2 at scale level 2. Convolving with the same group of steerable filters b_i, the low frequency band I_2 is turned into the orientation decomposition results R_{2,θ_i} in scale level 2. Repeating the same computation process from I_1 to R_{2,θ_i} in the following scale level, the orientation decomposition results at all scale levels can be obtained.

2.2. Phase-based optical flow

In order to convert the phase information into the motion of pixels, the method of phase-based optical flow should be adopted. The phase-based optical flow method is computed based on the phase information from the complex convolution results between the complex steerable filters and low-pass sub-bands at different scale levels in the CSP, referred to as R_{j,θ_i} in the previous section. In Eq. (1), A_{j,θ_i} represents the amplitude of the complex result, while ϕ_{j,θ_i} denotes the phase.

$$I_j \otimes b_i = R_{j,\theta_i} = A_{j,\theta_i}(x,y,t)e^{j\phi_{j,\theta_i}(x,y,t)}, i,j \in \{0,1,\ldots,n\} \quad (1)$$

As proposed by Fleets et al. [26], the phase information can have a good representation of the motion pattern of frames. They adopted the traditional intensity-based optical flow method to compute pixel velocity, employing both the phase constant and small motion assumption to determine pixel velocity based on phase information.

In traditional intensity-based optical flow method, pixel intensity can convey motion information in two directions, corresponding to two unknown components. However, the intensity constant assumption of optical flow method imposes only a single constraint. Consequently, this leads to the well-known aperture problem in the optical flow method [21].

However, for the computation of the component velocity of the phase-based optical flow method, the aperture problem does not exist. This is because the phase information only contains one-directional information in the image, which is the nature of the phase. Consequently, in phase-based optical flow, there are no extraneous unknown parameters. The phase constant assumption can be employed to resolve the velocity in a singular direction based on phase information, aligning with the direction of the phase gradient.

Here is the process of computing phase-based optical flow. Once the

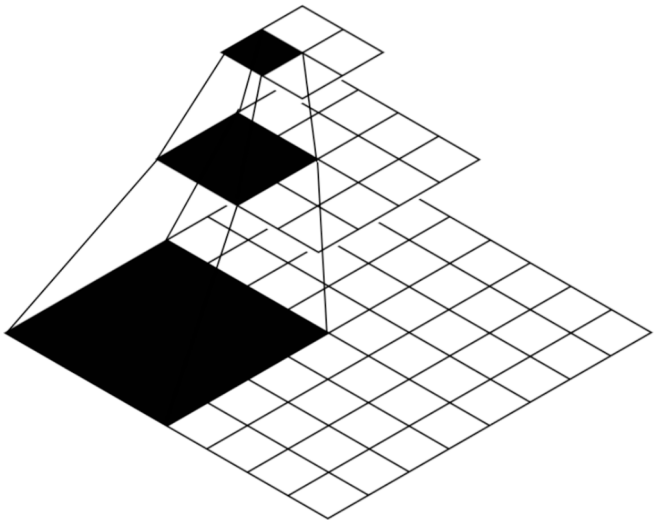


Fig. 2. The pixel correspondence between scale levels of CSP.

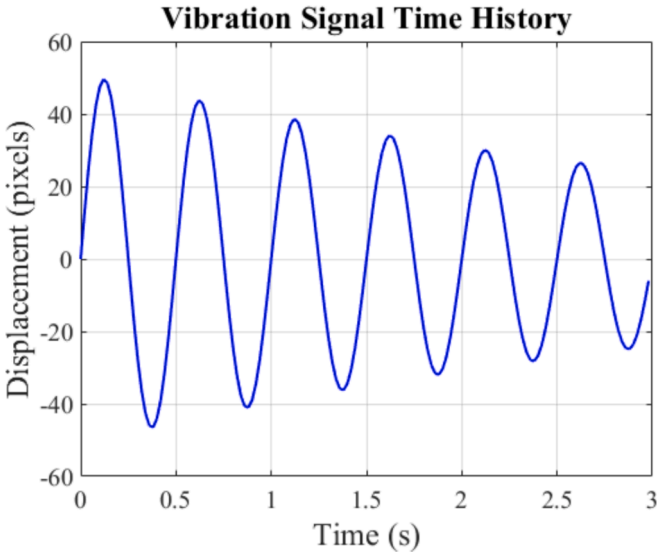


Fig. 3. The vibration curve of the circular target in numerical video when A is set to be 0.04 × 1280 pixels.

phase ϕ is obtained, which is the phase ϕ_{j,θ_i} with respect to any scale and any orientation in Eq. (1), it is set to be a constant, c, based on the phase constant assumption.

$$\phi(x,y,t) = c \tag{2}$$

With small motion assumption, the derivation with respect to time can be done on each side of Eq. (2) to obtain Eq. (3). In Eq. (3), ϕ_x represents the spatial phase gradient with respect to the horizontal direction of the frame, ϕ_y denotes the vertical spatial phase gradient, and ϕ_t is the temporal phase gradient. Additionally, v_x and v_y correspond to the velocities of the pixel in the horizontal and vertical directions, respectively.

$$\phi_x v_x + \phi_y v_y + \phi_t = 0 \tag{3}$$

As shown in Eq. (4), the spatial gradient of phase $\overrightarrow{\phi_c}$ is a vector aligning with the direction of the phase.

$$(\phi_x, \phi_y) = \overrightarrow{\phi_c} \tag{4}$$

In Eq. (5), the full velocity of pixel $\vec{v}$ is a vector in any direction.

$$(v_x, v_y) = \vec{v} \tag{5}$$

The vectors $\overrightarrow{\phi_c}$ and $\vec{v}$ form an angle β between them. By expressing the Eq. (3) in vector form, the following Eq. (6) is obtained.

$$\left\|\overrightarrow{\phi_c}\right\| \|\vec{v}\| \cos\beta + \phi_t = 0 \tag{6}$$

$\vec{v_c}$ is the projection of full velocity on the direction of phase, which is the component velocity that is computed from phase-based optical flow, as shown in Eq. (7).

$$\vec{v_c} = \overrightarrow{\phi_c^n} \|\vec{v}\| \cos\beta = \overrightarrow{\phi_c^n} \left(-\frac{\phi_t}{\left\|\overrightarrow{\phi_c}\right\|} \right) = \frac{(\phi_x, \phi_y)}{\left\|\overrightarrow{\phi_c}\right\|} \left(-\frac{\phi_t}{\left\|\overrightarrow{\phi_c}\right\|} \right)$$
$$= -\frac{\phi_t}{(\phi_x^2 + \phi_y^2)} (\phi_x, \phi_y) \tag{7}$$

Once the component velocity of the pixels is calculated, it can be multiplied by the time interval between two frames to determine the pixel displacement, as shown in Eq. (8), expressed in units of pixels per second.

$$\vec{d_c} = \vec{v_c} \Delta t \tag{8}$$

2.3. The scale level of CSP

When using CSP to decompose video frames, it is essential to establish two parameters, which are the number of directions and scale levels [42]. It is unnecessary to set the bandwidth for each steerable filter, as they are preliminarily designed as a one-octave bandwidth and are free from aliasing in the frequency domain [30,42].

In the field of vision-based structural response monitoring, the choice of measurement direction should depend on the structural response presented in the video. Once the direction of interest is identified, the scale level is the only parameter in CSP left to be carefully considered for efficient phase-based motion measurement.

To facilitate the comprehension of the scale level in CSP, it can be beneficial to draw an analogy with the discrete wavelet transform (DWT). In the decomposition process of DWT, a lowpass filter is applied to convolve with the subband of a certain scale level before down-sampling [48]. Consequently, the detail coefficients at higher scale levels correspond to lower frequency components. A similar process is

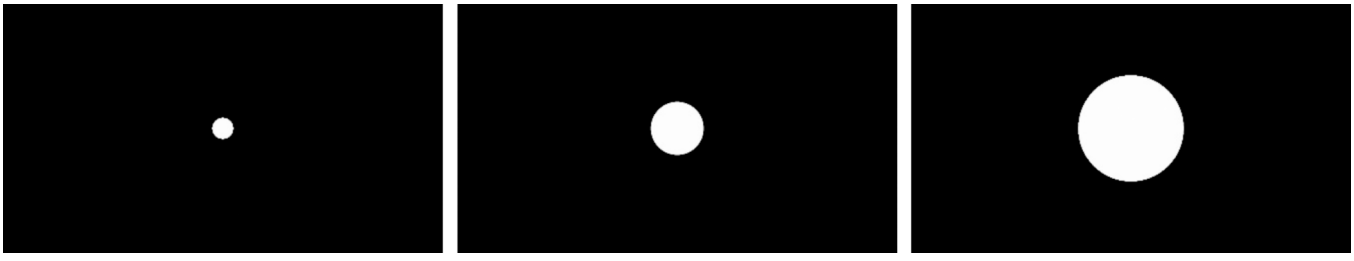
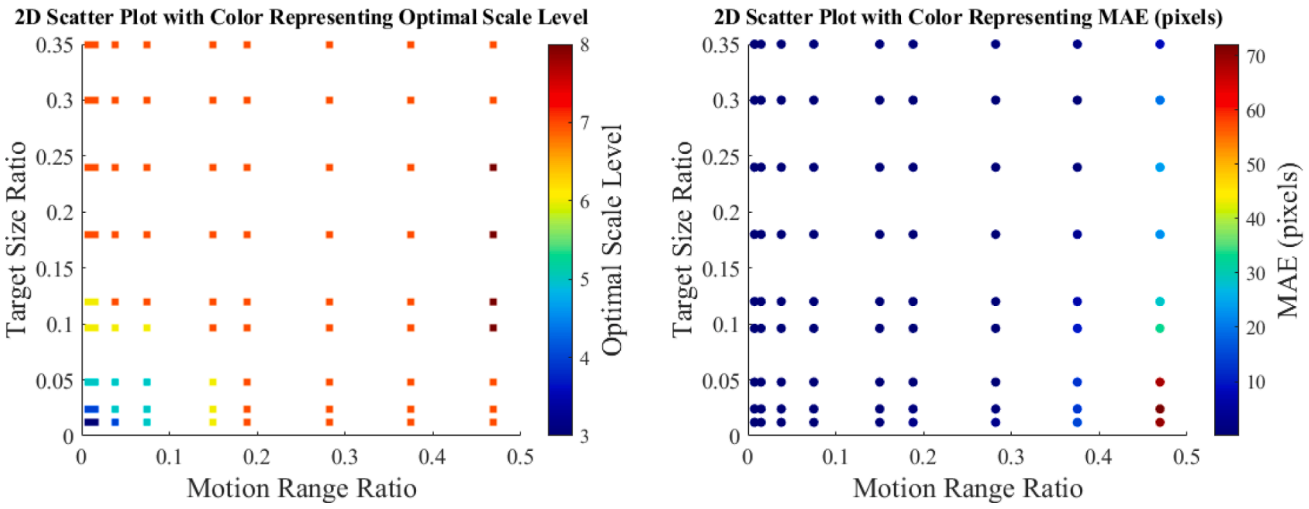


Fig. 4. The first frame of the diameter of D × 1280 pixels of the target (D = 0.048, 0.120, 0.240).



(a). Optimal scale level **(b). Accuracy (MAE)**

Fig. 5. The optimal scale levels and their accuracy under different motion ranges and target sizes.

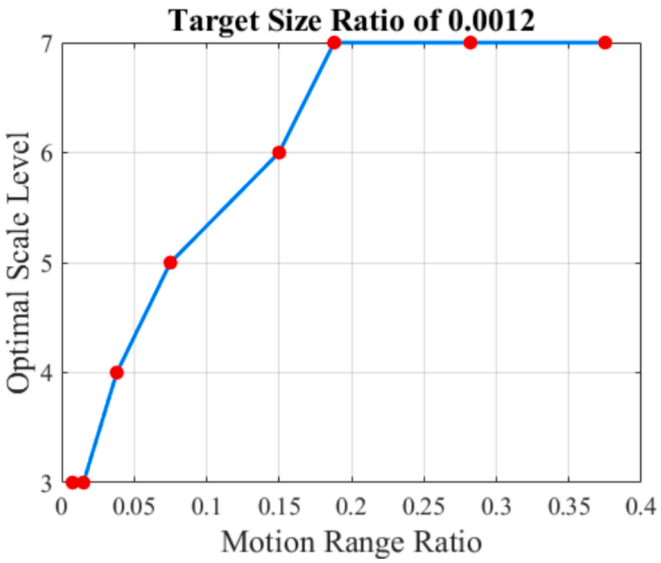


Fig. 6. The relationship between optimal scale level and motion range ratio when target size is 0.012 × 1280 pixels.

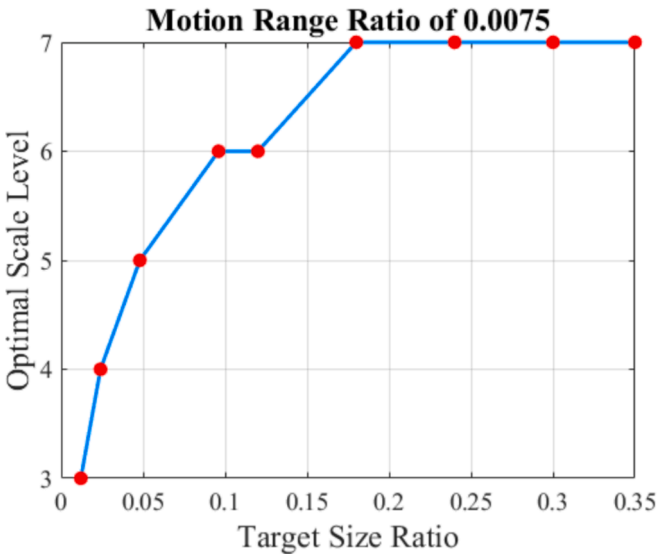


Fig. 7. The relationship between optimal scale level and target size ratio when motion range is 0.0075 × 1280 pixels.

employed in the CSP method, where a certain scale level also corresponds to specific frequency components within an image.

Due to the downsampling in the CSP decomposition process, one pixel at a higher scale level can correspond to a large pixel region in the original image, as shown in Fig. 2. Consequently, if phase information from a specific scale level is used to compute displacement, the displacement at that scale level should be multiplied by $2^{\text{scale level}}$ to convert it into the actual displacement, accurately reflecting the motion in the original image, as shown in Eq. (9).

$$\vec{d}_{c\,\text{scale level n}} \times 2^{\text{scale level n}} = \vec{d}_c \tag{9}$$

The scale level of CSP has proved to be important for motion measurement [33]. Peng et al. [46] have done systematic research on the CSP parameters' impacts on the measurement performance. However, the connections between optimal parameters of CSP and the properties of motion video are not their focus, which are important for building a practical measurement strategy.

In the field of feature detection, the frequency components of an image correspond to different objects within that image. For example, high-frequency components typically represent thin features, such as edges, whereas low-frequency components are generally associated with

Table 1
The measurement performance, MAE values of all scale levels of CSP for the video case with MRR = 0.075 and TSR = 0.024.

Video Case	Scale Level	MAE (pixels)
MRR = 0.075, TSR = 0.024	0	23.0532
	1	23.0481
	2	23.0469
	3	26.5664
	4	21.6060
	5	0.0866
	6	0.1396
	7	0.7099
	8	66.1620
	9	309.8965

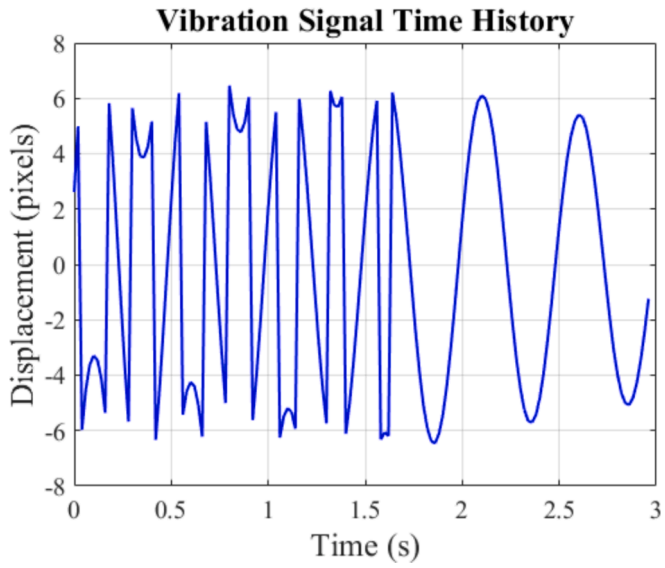


Fig. 8. The motion measurement result from the scale level 4 for the motion video case of 0.075 × 1280 pixels of motion range and 0.024 × 1280 pixels of target size.

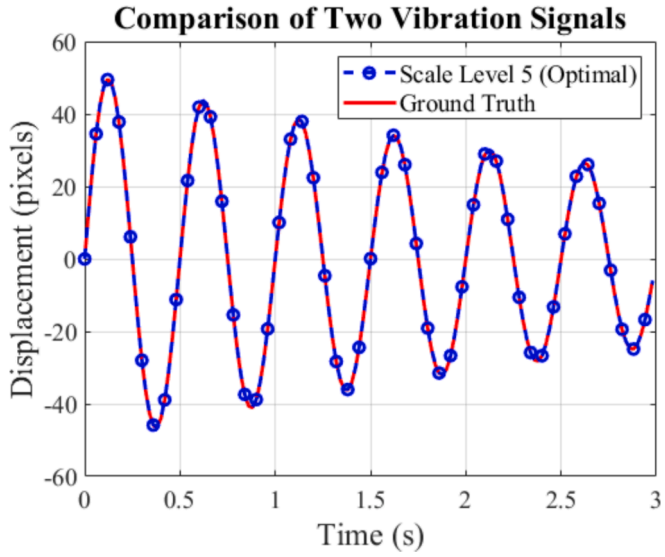


Fig. 9. The motion measurement result from the optimal scale level (level 5) of above case and its comparation with ground truth.

broader features, like surfaces [27]. The CSP method can adjust its scale level to focus on the target of different frequencies in the image.

However, for the motion estimation of a video, the motion region, which contains the motion range and target size, can also be related to a certain frequency. Therefore, it is essential to further investigate whether the scale level of CSP needs to be carefully selected to accommodate the motion range and target size. Under these circumstances, the subsequent sections evaluate the motion measurement performance of the CSP scale level using a phase-based method across various motion ranges and target sizes through both numerical and laboratory experiments.

3. The automatic selection of optimal scale level in CSP

In this section, a series of numerical videos with different motion ranges and target sizes are employed for the motion estimation tasks. By computing the phase-based optical flow on these videos using CSP, the optimal scale levels are ascertained by correlating measurement results with the established ground truths. Analysis reveals distinct trends of the optimal scale levels in relation to motion ranges and target sizes. Based on the data and relationships from the numerical experiment, a measurement strategy is formulated to determine the optimal scale level of CSP by considering the motion range and target size, without knowing the ground truth of motion.

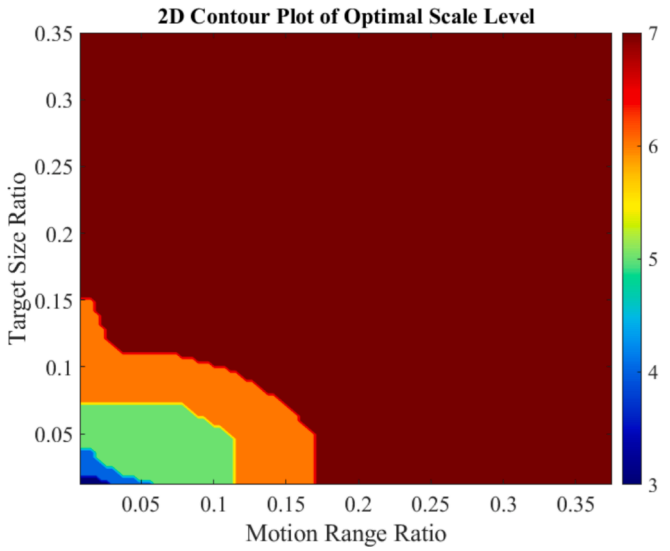


Fig. 10. The 2D contour map of optimal scale level with respect to target size ratio and motion range ratio.

3.1. Numerical experiment

3.1.1. Numerical videos

For the numerical test videos, an 8-bit depth format with a resolution of 720P (1280 × 720 pixels) is selected, considering the computational costs and versatility. The videos feature a black background (intensity value of 0) and a white circular target (intensity value of 255) initially centered on the screen. Throughout the video, the target undergoes damped simple harmonic motion horizontally, characterized by a damping coefficient of 0.25. The designed motion equation and curve are shown in Eq. (10) and Fig. 3. The frame rate is set at 50 frames per second (FPS), and the video 's total duration is 3 s. To mimic realistic noise conditions, Gaussian noise with a standard deviation of 2 pixels is added.

$$x = Ae^{-ct}\sin(\omega_n t) \tag{10}$$

For the amplitude of motion A, values of (0.004, 0.008, 0.020, 0.040, 0.080, 0.100, 0.150, 0.200, 0.250) as fractions of the frame width (1280 pixels) are selected. The motion ranges, defined by the span between the maximum and minimum displacement values, are (0.0075, 0.015, 0.038, 0.075, 0.150, 0.188, 0.282, 0.375, 0.469) as fractions of the frame width. Given the minimal motion range is very small (0.0075 × 1280 = 9.6 pixels), the videos are generated with subpixel motion. This is achieved by up-sampling the video by a factor of ten to capture the subpixel movement, followed by bicubic interpolation to restore the video to its original 720P resolution. This process ensures the video accurately reflects movements as small as 0.1 pixels.

The size of the target, defined as the length of projection of the target in the motion direction, is also deemed critical. For the circular target, the diameter is the target size. Diameters of 0.012, 0.024, 0.048, 0.096, 0.120, 0.180, 0.240, 0.300, and 0.350 as fractions of the frame width (1280 pixels) are chosen for these targets. The first frames of some cases in the series of numerical videos are shown in Fig. 4.

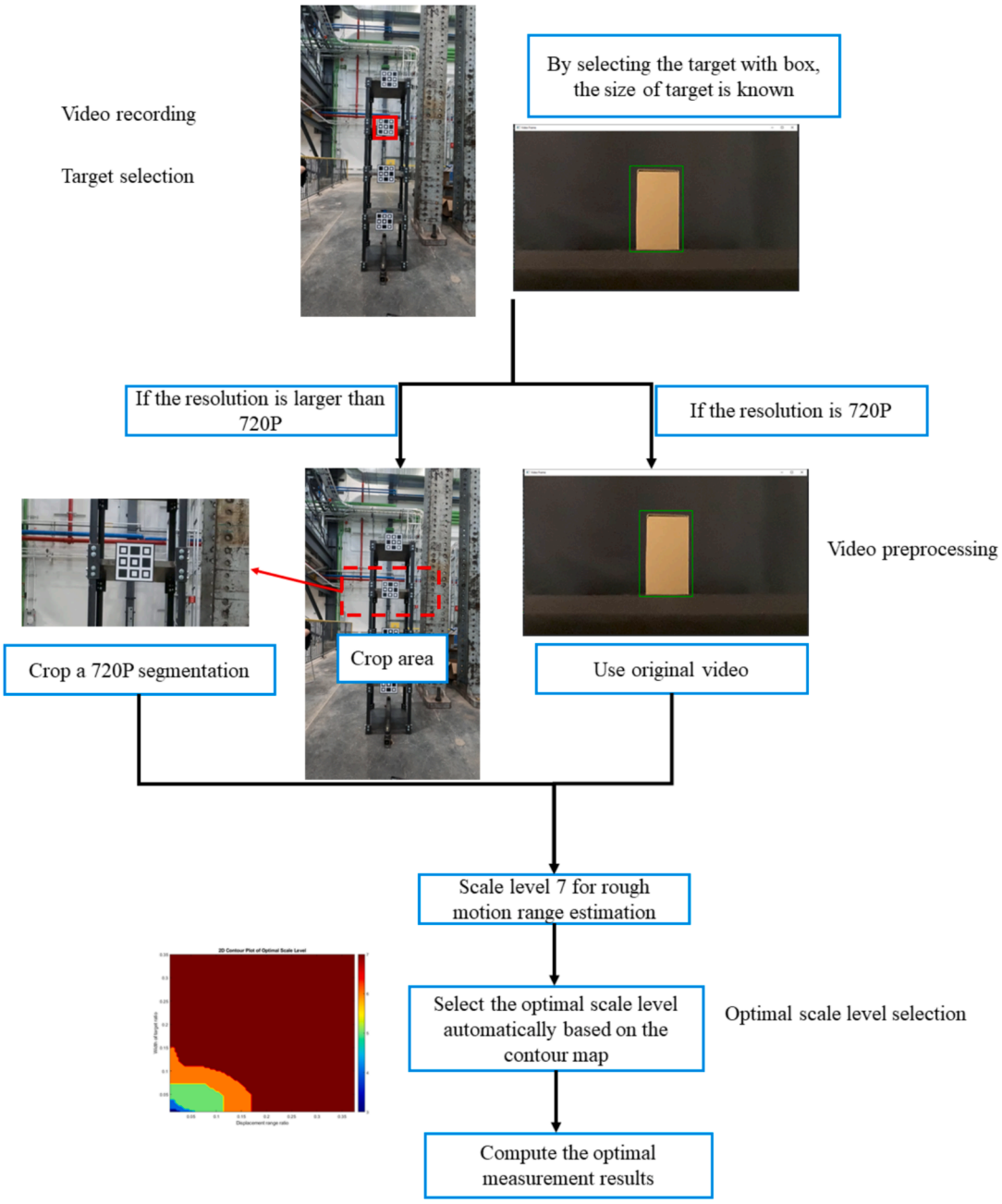


Fig. 11. The flowchart of the measurement strategy.

3.1.2. The optimal scale levels of CSP under different motion ranges and target sizes

For a resolution of 720P video, the highest scale level of CSP is 9. The frame size of scale level 0 is 1280 × 720, while the frame size of scale level 9 is 2 × 1. In order to find the optimal scale level for measurement, the phase-based motion estimations from scale level 0 to scale level 9 for all videos are conducted. The point of interest (POI) is adopted for motion estimation. For every scale level in CSP, the POI is positioned at the center of the frame.

Then, the mean average error (MAE) against the ground truth displacement is calculated to evaluate the measurement performance of each scale level and select the optimal scale level for each video. The optimal scale level of CSP is the scale level that has the least MAE value among all scale levels from 0 to 9. The optimal scale levels and the corresponding accuracy for the series of numerical videos are shown in Fig. 5 (a) and (b), respectively.

Observations from the results of the optimal scale levels in Fig. 5 indicate that higher motion ranges correspond to higher optimal scale levels. As shown in Fig. 6, for a target size ratio of 0.012, the optimal scale level is 3 when the motion range ratio is 0.0075. As the motion range ratio increases to 0.188, the optimal scale level rises to 7.

Additionally, larger targets correspond to higher optimal scale levels. In Fig. 7, with a motion range ratio of 0.0075, a target size ratio of 0.012 yields an optimal scale level of 3. When the target size ratio increases to 0.18, the optimal scale level escalates to 7.

However, it should also be noted that, although the optimal scale level has positive relationships with the motion range and target size, the highest optimal scale level is always 7. It will not reach scale levels 8 or 9 as the effective optimal scale level because the resolutions of these two scale levels are too low. This low resolution results in significant information loss from the original signal and an inability to provide reliable phase information.

(a). Experiment layout **(b). Laser sensor head**

Fig. 12. Images of the experiment layout and laser displacement sensor.

Table 2
The parameters of eight experimental cases.

Case name	Distance (mm)	Motion range (rough, mm)	Focus length (mm)	Length to pixel ratio (mm/pixel)
MR30mm, FL13mm	900	30	13	1.608
MR30mm, FL26mm	900	30	26	0.804
MR30mm, FL52mm	900	30	52	0.402
MR30mm, FL78mm	900	30	78	0.268
MR152mm, FL13mm	900	152	13	1.608
MR152mm, FL26mm	900	152	26	0.804
MR152mm, FL52mm	900	152	52	0.402
MR152mm, FL78mm	900	152	78	0.268

As mentioned in section 2.3, one pixel in a high-scale level of CSP could correspond to a vast region in the original frame. The region in the original frame that POI at a certain scale level would cover is the focus region of the certain scale level.

Although scale level 7 can handle many cases of large motion ranges and target sizes, some failed measurement situations also occur when the motion range and target size exceed the focus region of scale level 7. For instance, when the motion range ratio is 0.469, almost all cases cannot have reliable measurements, as the MAE values of these cases are observed to be extremely high in Fig. 5 (b).

From what is observed from the MAE results of all scale levels for the case of MRR = 0.075 (MRR is the motion range ratio) and TSR = 0.024 (TSR is the target size ratio) in Table 1, when the scale level is lower than the optimal scale level, the MAE results are significantly greater than the MAE value of the optimal scale level. The situation may be caused by the fact that the focus region of a certain scale level cannot cover the whole motion region, leading to poor measurement performance. This point can also be observed by the displacement from the measurement result of a certain scale level of CSP. As shown in Fig. 8, the displacement result from the scale level 4 is inaccurate at the beginning, where the higher amplitude part is incorrect and missing. However, the trend of lower amplitude motion towards the end of the video is measured accurately. In Fig. 9, the scale level 5 is the optimal scale level for this case, showing great consistency with the ground truth.

When the scale level is higher than the optimal scale level but within the scale level 7, the MAE result is slightly greater than the optimal scale level, which is still reliable but less accurate. As shown in Table 1, MAE values of the higher scale levels 6 and 7 are slightly larger than the optimal scale level 5. This is mainly caused by the extra noise (Gaussian noise) introduced by the large focus region of a higher scale level of CSP. Consequently, the scale level of CSP is the key parameter significantly affecting the performance of phase-based motion estimation.

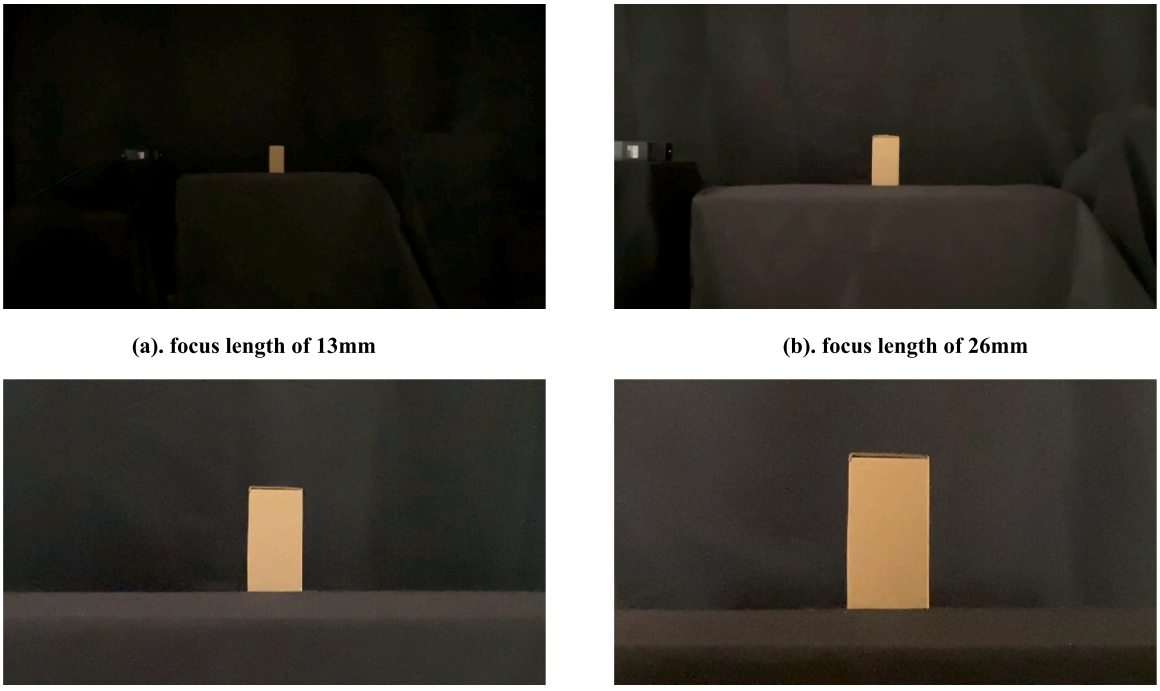
3.2. The trends of optimal scale level in relation to motion range and target size

In order to better understand the trends of optimal scale level for motion estimation with respect to motion range and target size, a 2D contour map of optimal scale level in relation to motion range and target size is proposed based on the results from numerical motion video estimation. From what is observed in Fig. 10, the optimal scale level has positive relationships with both the motion range and target size. The motion range and target size can be viewed together as the motion region, where any pixels from the target can appear. Considering this, the optimal scale level also has a positive relationship with motion region.

The discussion on the motion range ratio and target size ratio in the contour map is confined to values within 0.4. This restriction is due to the observation that, for cases exceeding this range, the majority fail in motion estimation. The cause of these failures, as detailed in Section 3.1.2, is attributed to the superior high scale level required for such cases, which has a significantly reduced resolution, thereby yielding unreliable phase information.

As observed in Fig. 10, within the region of effective measurement cases, the optimal scale level 7 occupies a substantial area, approximately 88.37 % of the region, indicating its significant adaptability for varying properties of motion videos. Conversely, optimal scale level 6 accounts for only 6.77 % of the region. Additionally, optimal scale level 5 corresponds to 4.31 %, optimal scale level 4 to 0.45 %, and optimal scale level 3 to a mere 0.10 % of the effective measurement cases region. It is evident that the adaptability of optimal scale levels decreases as the height of optimal scale level decreases.

This 2D contour map will also be helpful in predicting the optimal scale level when motion range and target size are known for a certain



(a). focus length of 13mm (b). focus length of 26mm

(c). focus length of 52mm (d). focus length of 78mm

Fig. 13. The first frames of video in the laboratory experiment with different focus lengths.

Table 3
The estimated optimal scale levels based on measurement strategy.

Cases	Rough motion range ratio from scale level 7	Target size ratio	Estimated optimal scale level
MR30mm, FL13mm	0.0149	0.025	4
MR30mm, FL26mm	0.0337	0.051	5
MR30mm, FL52mm	0.0671	0.101	6
MR30mm, FL78mm	0.0994	0.151	7
MR152mm, FL13mm	0.0549	0.025	5
MR152mm, FL26mm	0.1344	0.051	6
MR152mm, FL52mm	0.2850	0.101	7
MR152mm, FL78mm	0.3852	0.151	Non effective measurement

video motion measurement task.

3.3. The measurement strategy for the automatic selection of optimal scale level of CSP

For the better application of CSP for phase-based optical flow, a measurement strategy is proposed to help the automatic selection of the optimal scale level of CSP without knowing the ground truth of displacement.

Based on the observed MAE values from previous data, the scale levels ranging from the optimal scale level to scale level 7 are all relatively reliable for motion measurement. However, the optimal scale level exhibits the highest accuracy. Therefore, scale level 7 is always suitable for rough motion range estimation under any motion ranges and target sizes in a certain region. Given these circumstances, a measurement strategy for video with higher than or equal to the resolution of

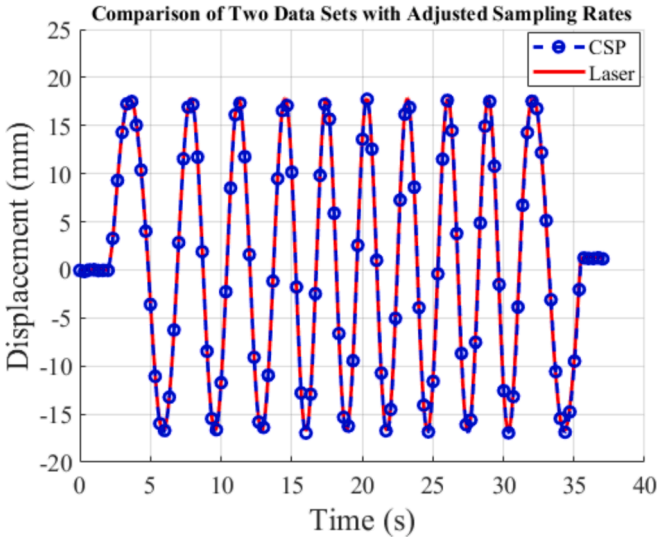


Fig. 14. The motion estimation result of the estimated optimal scale level of CSP for the case with MR of 30 mm and FL of 52 mm.

Table 4
The optimal scale levels of all experiment cases.

Case	Optimal scale level	MAE (mm)
MR30mm, FL13mm	4	0.4960
MR30mm, FL26mm	5	0.3457
MR30mm, FL52mm	6	0.2571
MR30mm, FL78mm	7	0.2134
MR152mm, FL13mm	5	1.5573
MR152mm, FL26mm	6	0.7792
MR152mm, FL52mm	7	0.6520
MR152mm, FL78mm	Can not handle	meaningless

Table 5
The measurement performance (MAE) of all scale levels of CSP for the experiment case with MR of 30 mm and FL of 26 mm.

Case	Scale level	MAE (mm)
MR30mm, FL26mm	0	10.0516
	1	9.8611
	2	10.0031
	3	9.5241
	4	10.0663
	5	0.3457
	6	0.4125
	7	0.6006
	8	4.7232
	9	7.8739

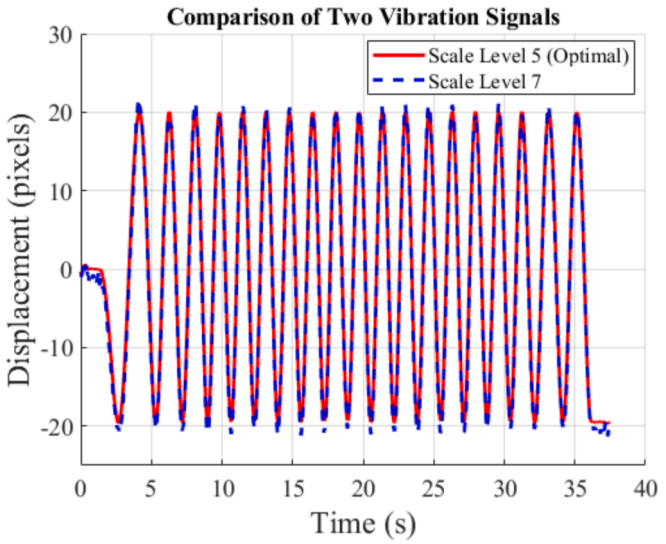


Fig. 15. Comparison of motion measurement results between scale level 5 (optimal) and scale level 7 for the case with MR of 30 mm and FL of 26 mm.

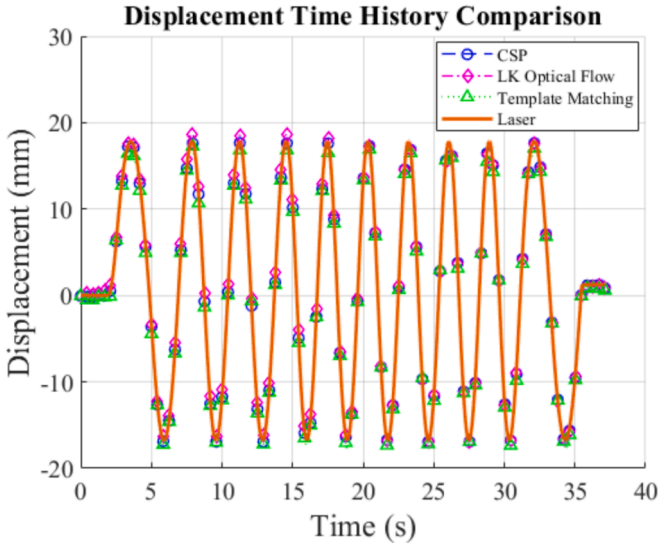


Fig. 16. Comparison of motion measurement results between different vision-based methods.

720P and one-dimensional motion is proposed to achieve the optimal motion estimation automatically using the CSP for the phase-based optical flow method, as shown in following contents and Fig. 11.

Table 6
The measurement performances of different vision based methods.

	CSP	LK	TM
MAE (mm)	0.25715	0.462734	0.51193

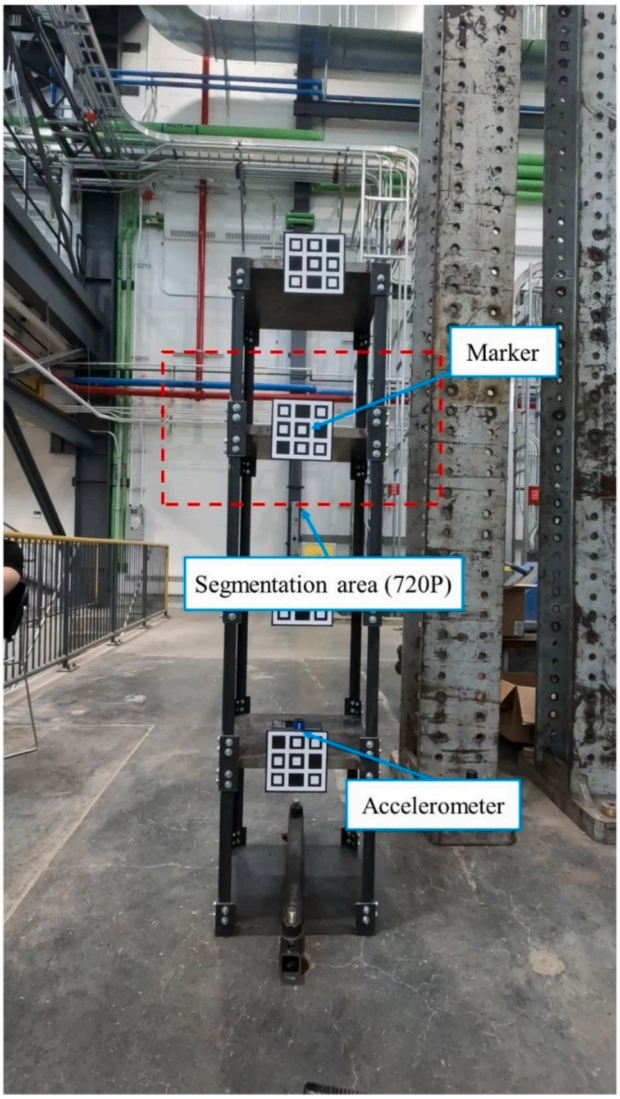


Fig. 17. The setup of experiments and the location of segmentation area.

(1) With a camera, record the one-dimensional motion of a target along the direction with the width or height direction of the frame, as long as the length of this direction is equal to or larger than 1280 pixels.
(2) Select the target of interest with a bounding box in the first frame of the video. If the video resolution is 720P, go to step (3). Else, crop a 720P resolution segmentation around this target in this video, make the width direction of the 720P video aligned with the direction of motion.
(3) The length of the target along the motion direction is automatically obtained from the size of the box. Then, the size of the target is determined.
(4) Use scale level 7 to estimate the displacement of the target first. Then, the approximate range of motion can be obtained.
(5) With the target size and approximate motion range known, use the 2D contour map of the optimal scale level shown in Fig. 10 to automatically identify the optimal scale level of CSP for the

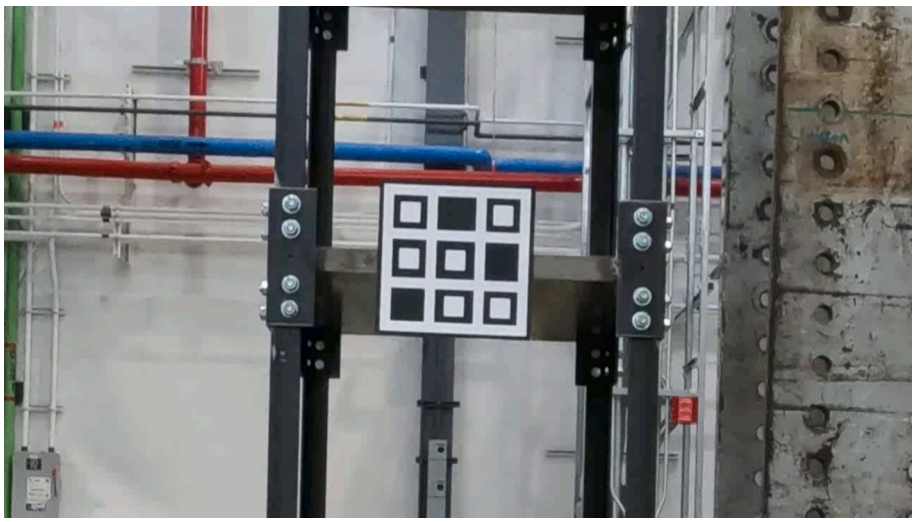
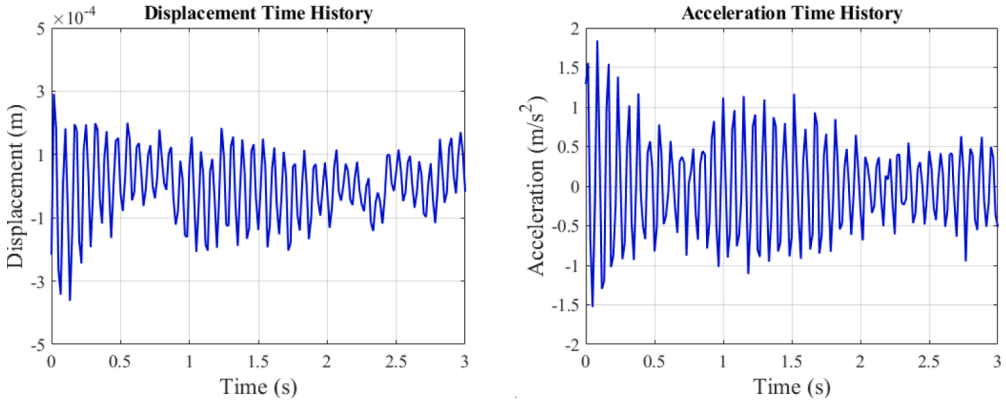
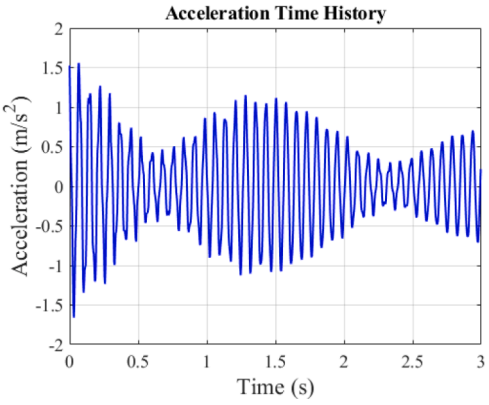


Fig. 18. The first frame of cropped video in the process of measurement strategy.



(a). Displacement from optimal scale level **(b). Acceleration from optimal scale level**



(c). Acceleration from accelerometer

Fig. 19. The displacement time series from from accelerometer and optimal scale level of CSP.

phase-based optical flow measurement, without the need to know the ground truth of displacement.

(6) Use the phase information from the optimal scale level of CSP to compute the phase-based optical flow; the optimal displacement measurement result is then obtained.

However, for videos with a resolution lower than 720P, the background extension can be used to expand the video to 720P. Considering the scarcity of cameras that record at resolutions lower than 720P, this case will not be further discussed in detail.

4. Validation experiments

After the measurement strategy is proposed, a series of laboratory experiments are conducted to verify it. The layout of the experiment is shown in Fig. 12 Fig. 12(a). The motion target of the experiment is a

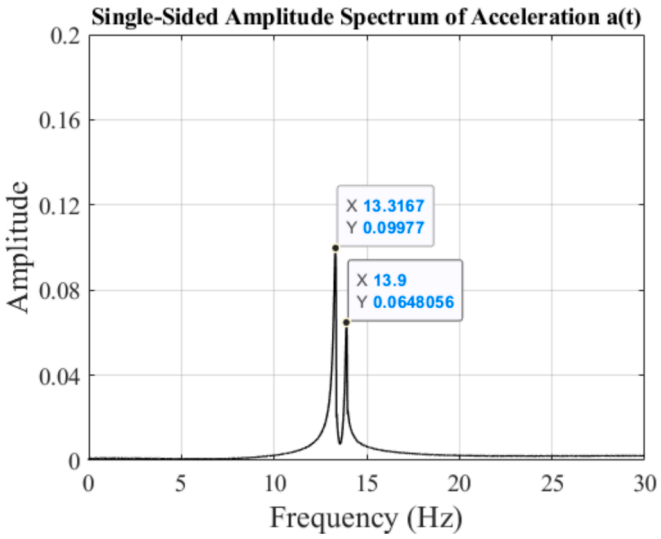


Fig. 20. The frequency spectrum of vibration signal from the accelerometer.

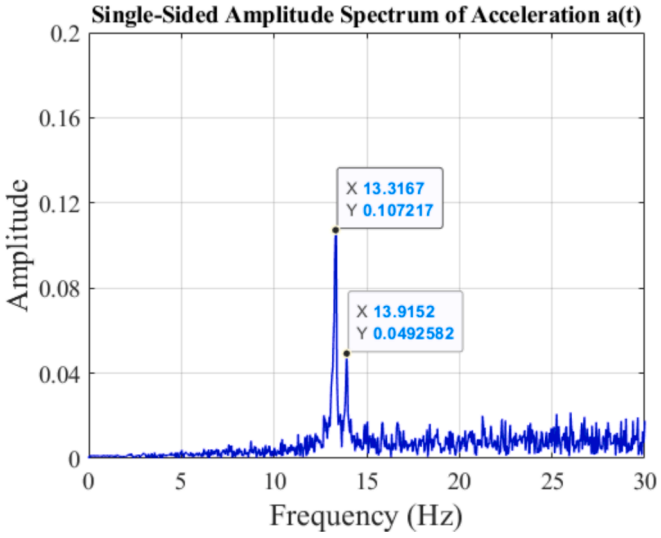


Fig. 21. The frequency spectrum of vibration signal from the automatically obtained optimal scale level using phase-based method.

column fixed on the cylinder head of a shaker. The shaker can produce reciprocating motion with adjustable speed. The range of motion is adjustable, roughly from 30 mm to 152 mm. Considering that the motion speed could be reduced to minimal levels, an iPhone 11 camera is employed to record the motion using 720P video at a frame rate of 24 FPS. The camera is positioned 900 mm away from the moving target, with the optical axis perpendicular to the direction of the target's motion. The ground truths of motion are measured by a laser displacement sensor (LDS) of Keyence LK-G3000, with an LK-G405 sensor head (effective measurement range 400 mm $\pm$ 100 mm), as shown in Fig. 12 (b).

To validate the measurement strategy across varying motion ranges and target sizes, adjustments are made to the shaker's motion range and the camera's focal length, thereby establishing diverse experimental conditions. It should be noted that when altering the focal length, the phone camera generally utilizes a digital zoom strategy. Unlike optical zoom, which changes the physical focal length, digital zoom employs a computational approach by cropping, interpolating, and applying proprietary image enhancement techniques digitally to achieve the desired zoom effect. The camera's focal length (FL) is adjusted to 13 mm, 26 mm, 52 mm, and 78 mm to achieve varying target sizes for the experimental cases. Additionally, the range of motion can also be altered by varying the focal lengths. The shaker's stroke is set to approximately 30 mm and 152 mm to produce varying motion ranges (MR) for the experimental cases.

Consequently, eight experimental cases are conducted to validate our measurement strategy, as shown in Table 2. As the focal length is varied across the series of experimental cases, the length-to-pixel ratio is changed for each experimental case.

Fig. 13 shows the initial frames for various focal length settings. It is evident that the target size varies with different focal lengths.

Assuming the ground truth of the motion is unknown, our measurement strategy begins by measuring the target size in the direction of motion, specifically the width of the rectangular target. Subsequently, the scale level 7 is employed for a preliminary estimation of the motion range. Once the target size and rough motion range ratios are determined, the estimated optimal scale levels of CSP for different experimental cases can be obtained by searching on the 2D contour map of the optimal scale level, as shown in Table 3.

Fig. 14 displays the displacement measurement results from the estimated optimal scale level of CSP compared to those obtained by the laser displacement sensor for the case with MR of 30 mm and FL of 52 mm, showing great consistency with the laser results.

To verify the effectiveness of the measurement strategy, it is necessary to compute the displacements of target at all scale levels of CSP and compare these results with the displacement measurements from the laser displacement sensor. The sample rate of the laser displacement sensor is set to 1000 Hz, with a sampling duration of 60 s. Due to the absence of a method to synchronize the camera video with the laser sensor measurements, post-processing alignment is employed to achieve temporal synchronization. MAE is used as a metric for accuracy. By comparing the MAE values across various scale levels of CSP in different experimental cases, the optimal scale levels for all eight experimental cases are determined.

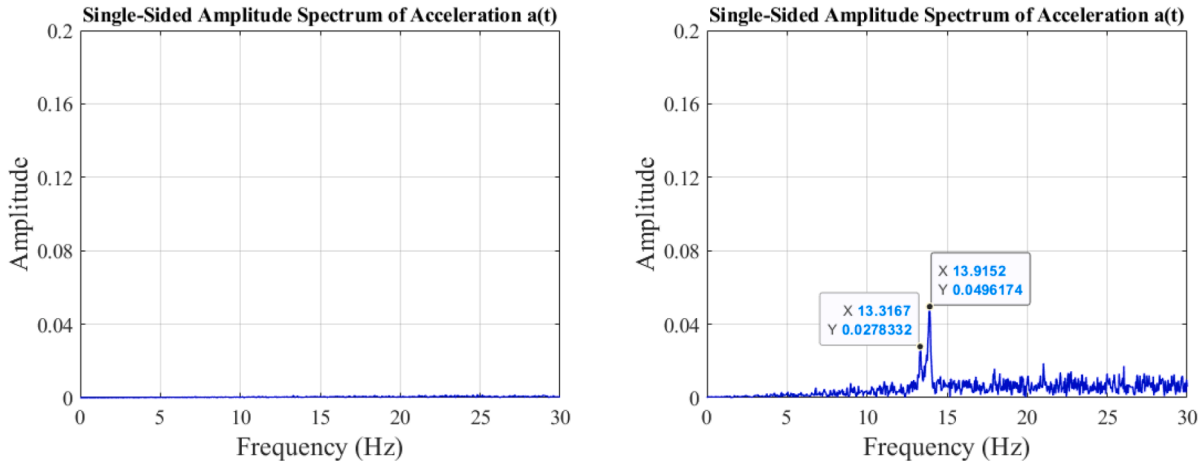
As shown in Table 4, all optimal scale levels determined by the MAE values are consistent with the estimated optimal scale levels derived from our measurement strategy. This consistency confirms the effectiveness of our measurement strategy for the practical application of the CSP phase-based optical flow method.

In Table 5, the measurement performances of all scale levels for the case with MR of 30 mm and FL of 13 mm clearly demonstrate the effectiveness of our measurement strategy. The lowest MAE value, 0.3457 mm, is achieved at scale level 5, indicating it as the optimal scale level for the case. Meanwhile, scale level 7 yielded an MAE value of 0.6006 mm, which, while slightly higher, is still sufficiently accurate for estimating rough amplitudes of motion. Under such circumstances, the approximate motion range can be derived from the measurement results at scale level 7, as illustrated in Fig. 15.

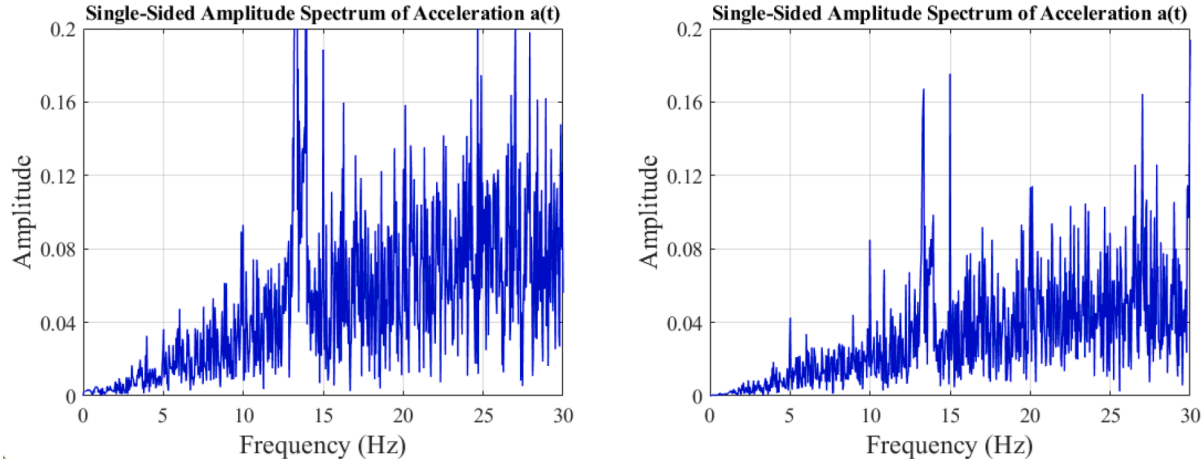
To further verify the superiority of proposed measurement strategy, the measurement video for the case with MR of 30 mm and FL of 26 mm is further processed with traditional vision-based methods. For the LK optical flow, the POI is set at the vertical edge of the column, where the feature is dense with respect to moving direction. For the template matching, the normalized cross-correlation coefficient is employed for matching, and a parabolic fitting method around the peak match location is employed for subpixel level accuracy. As observed in Fig. 16 and Table 6, the estimation results from optimal scale level show better accuracy than LK optical flow and template matching.

5. The dynamic property test on a four-floor frame structure

A vision-based dynamic property test of a four-floor frame structure is conducted to verify the proposed measurement strategy. The frame structure is composed of steel plates and steel symmetric angle sections connected by bolts and brackets. The structure has four stories, each with a height of 0.4 m, and a floor area of 0.4 m $\times$ 0.4 m. Column

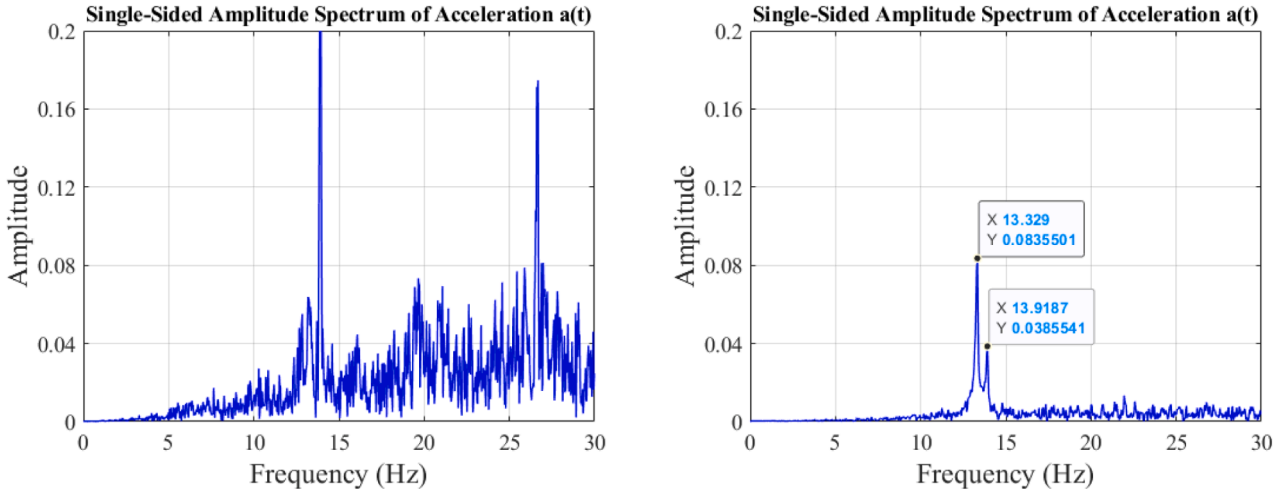


(a). Scale level 5 (b). Scale level 6



(c). Scale level 8 (c). Scale level 9

Fig. 22. The frequency spectrum of vibration signal from other scale levels of CSP using phase-based methods.



(a). Template matching (b). LK optical flow

Fig. 23. The frequency spectrum of vibration signal from the template matching and LK optical flow.

sections L 31.75 × 31.75 × 4.76 mm are used for the columns. The structure's base is rigidly clamped to the ground via an HSS beam. The frequency identification of each floor is performed using accelerometer measurements. The accelerometers are fixed at the center of each floor and are able to measure acceleration in two directions along the horizontal plane, with a sampling frequency of 256 Hz. Designed markers are fixed on each floor as targets for vision-based methods. videos recording the vibration of the frame structure are captured using a GoPro10, with a resolution of 3840 × 2160 pixels and 60 FPS. The vibration of the frame structure is induced by the impact of a hammer. The setup for this experiment is shown in Fig. 17.

Following our measurement strategy, the vibration of structure is recorded with camera video. The dynamic property of third floor is studied for verification. Then, the marker at third floor is selected as the target of interest with box window. A videosegment with 720P resolution is extracted, ensuring the target is centered in the frame, as shown in Fig. 17 and Fig. 18. The target size (231 pixels) and approximate motion range (2 pixels) are obtained within our measurement strategy. Based on those properties of motion video, the estimated optimal scale level for measurement is automatically obtained as 7.

The displacement time series measured by phase-based method are shown in Fig. 19 (a). Taking the second derivative of displacement series yields acceleration series, as shown in Fig. 19 (b). The acceleration time series from accelerometer and optimal scale level shows a similar time series pattern, as presented in Fig. 19 (b) and (c). The dynamic properties from accelerometer are shown in Fig. 20, with a dominant frequency of 13.3167 Hz. Considering the sample rate of camera, the dynamic properties within 30 Hz of third floor based on the results from optimal scale level are shown in Fig. 21. The measurements shows good estimation of the dynamic properties of the frame structure, with the same dominant frequency as the accelerometer. Besides, there has a second peak with a frequency of 13.9 Hz in the frequency spectrum of accelerometer in Fig. 20. Due to close distance between the two peaks in the frequency domain, this situation is referred to as the modal coupling problem, which indicates that the frame system has natural frequencies that are very close to each other. As for the the optimal scale level of CSP, it also detects the second peak with the right frequency of 13.9152 Hz.

In the measurement strategy for this case, the estimated optimal scale level is 7. To show its superior performance among different scale levels, the measured dynamic properties from scale level 5,6,8,9, are also presented for the comparison in Fig. 22. Although the result from scale level 6 is close to scale level 7, the dominant frequency from scale level 6 is different from the accelerometer, possessing a frequency of 13.9152 Hz. Meanwhile, the second frequency from scale 6 is 13.3167 Hz, corresponding to the dominant frequency of the accelerometer. The reversed two peaks in the frequency spectrum of scale level 6 indicate an inappropriate measurement. Furthermore, irrelevant frequency components become more prominent at scales 8 and 9, indicating a failure in measurements. At scale level 5, the magnitude of the frequency spectrum is obviously incorrect, demonstrating the measurement failure. These cases demonstrate the superior measurement performance at the optimal scale level.

As we observed during the motion range estimation, the pixel motion in the video is very small, around 2 pixels. This is because the frame structure is quite stiff, and the amplitude of vibration is small, leading to minimal pixel motion. This presents a challenge for vision-based methods. However, the high subpixel accuracy of the phase-based method is effective in this case.

Traditional vision-based methods, such as the LK optical flow and template matching, are used for comparison. As shown in Fig. 23 (b), the LK optical flow method can also provide accurate frequency estimation with the dominant frequency of 13.329 Hz and the second frequency of 13.9187 Hz, since it is designed to have good subpixel level accuracy. However, for the template matching method, due to the small pixel motion, it cannot have good estimation regarding the dynamic property of the frame structure, as shown in Fig. 23 (a).

6. Conclusion

In this study, the performance of phase-based optical flow measurement across various scale levels in CSP is systematically studied, considering different motion ranges and target sizes in motion videos. This research identifies trends in optimal scale levels in relation to motion range and target size, and proposes a measurement strategy to optimize performance of CSP phase-based optical flow for the displacement estimation. The conclusions and measurement strategy are substantiated and validated through both numerical and laboratory vision-based measurement experiments. The key conclusions are summarized as follows:

(1). The scale level of CSP affects the measurement performance of phase-based displacement estimation significantly. There exists an optimal scale level for displacement estimation of a certain motion video.

(2). The optimal scale level for displacement measurement has positive relationships with both the motion range and target size of motion video. And this is further verified in a series of laboratory experiments.

(3). A 2D contour map of the optimal scale level in relation to motion range ratio and target size ratio is proposed based on the numerical experiment data of optimal scale levels under different motion ranges and target sizes.

(4). Based on the 2D contour map of the optimal scale level, a measurement strategy is proposed to automatically select the optimal CSP scale level for phase-based motion estimation. This strategy takes into account the motion range and target size, without requiring knowledge of the motion's ground truth, and it is further verified in a series of laboratory experiments and structural dynamic property estimation experiments.

As previously investigated in this article, the case with large motion region is unable to deal with in this strategy, due to the unreliable phase information at very high scale level of CSP. Besides, this measurement strategy is only focusing one-dimensional displacement estimation. Further investigation is warranted to expand this strategy to two-dimensional situations.

CRediT authorship contribution statement

Sida Ai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chuanzhi Dong:** Writing – review & editing, Supervision, Investigation, Conceptualization. **Qipei Mei:** Writing – review & editing, Supervision, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was financially supported by the China Scholarship Council (No. 202106370030).

Data availability

Data will be made available on request.

References

[1] C.-Z. Dong, F.N. Catbas, A review of computer vision–based structural health monitoring at local and global levels, Struct. Health Monit. 20 (2021) 692–743, https://doi.org/10.1177/1475921720935585.

[2] S.C. Saxena, S.B.L. Seksena, A self-compensated smart LVDT transducer, IEEE Trans. Instrum. Meas. 38 (1989) 748–753, https://doi.org/10.1109/19.32186.

[3] F. Gomez, J.-W. Park, B.F. Spencer, Reference-free structural dynamic displacement estimation method, Struct. Control Health Monit. 25 (2018) e2209.

[4] D. Feng, M.Q. Feng, Vision-based multipoint displacement measurement for structural health monitoring: Vision-Based Displacement Measurement for SHM, Struct. Control Health Monit. 23 (2016) 876–890, https://doi.org/10.1002/stc.1819.

[5] Y. Chen, D. Joffre, P. Avitabile, Underwater Dynamic Response at Limited Points Expanded to Full-Field Strain Response, J. Vib. Acoust. 140 (2018) 051016, https://doi.org/10.1115/1.4039800.

[6] Y. Xu, J.M.W. Brownjohn, Review of machine-vision based methodologies for displacement measurement in civil structures, J. Civ. Struct. Health Monit. 8 (2018) 91–110, https://doi.org/10.1007/s13349-017-0261-4.

[7] D. Feng, M.Q. Feng, Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review, Eng. Struct. 156 (2018) 105–117, https://doi.org/10.1016/j.engstruct.2017.11.018.

[8] Y. Miao, Y. Kong, H. Nam, S. Lee, G. Park, Phase-based vibration imaging for structural dynamics applications: Marker-free full-field displacement measurements with confidence measures, Mech. Syst. Signal Process. 198 (2023) 110418, https://doi.org/10.1016/j.ymssp.2023.110418.

[9] J.M.W. Brownjohn, Y. Xu, D. Hester, Vision-Based Bridge Deformation Monitoring, Front. Built Environ. 3 (2017), https://doi.org/10.3389/fbuil.2017.00023.

[10] Y. Xu, J. Brownjohn, D. Kong, A non-contact vision-based system for multipoint displacement monitoring in a cable-stayed footbridge, Struct. Control Health Monit. 25 (2018) e2155.

[11] C.-Z. Dong, S. Bas, F.N. Catbas, A completely non-contact recognition system for bridge unit influence line using portable cameras and computer vision, Smart Struct. Syst. 24 (2019) 617–630, https://doi.org/10.12989/SSS.2019.24.5.617.

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

[13] X. Huang, X. Wan, D. Peng, Robust Feature Matching with Spatial Smoothness Constraints, Remote Sens. 12 (2020) 3158, https://doi.org/10.3390/rs12193158.

[14] T. Khuc, F.N. Catbas, Computer vision-based displacement and vibration monitoring without using physical target on structures, Struct. Infrastruct. Eng. 13 (2017) 505–516, https://doi.org/10.1080/15732479.2016.1164729.

[15] Q. Hu, S. He, S. Wang, Y. Liu, Z. Zhang, L. He, F. Wang, Q. Cai, R. Shi, Y. Yang, A High-Speed Target-Free Vision-Based Sensor for Bus Rapid Transit Viaduct Vibration Measurements Using CMT and ORB Algorithms, Sensors 17 (2017) 1305, https://doi.org/10.3390/s17061305.

[16] J. Baqersad, P. Poozesh, C. Niezrecki, P. Avitabile, Photogrammetry and optical methods in structural dynamics – A review, Mech. Syst. Signal Process. 86 (2017) 17–34, https://doi.org/10.1016/j.ymssp.2016.02.011.

[17] M.N. Helfrick, C. Niezrecki, P. Avitabile, T. Schmidt, 3D digital image correlation methods for full-field vibration measurement, Mech. Syst. Signal Process. 25 (2011) 917–927, https://doi.org/10.1016/j.ymssp.2010.08.013.

[18] M. Hagara, F. Trebuňa, R. Huňady, M. Kalina, M. Schrötter, Experimental Identification of Modal Parameters of Thin Metal Sheets by using of DIC, Procedia Eng. 48 (2012) 180–188, https://doi.org/10.1016/j.proeng.2012.09.503.

[19] B.K.P. Horn, B.G. Schunck, Determining optical flow, Artif. Intell. 17 (1981) 185–203, https://doi.org/10.1016/0004-3702(81)90024-2.

[20] B.D. Lucas, T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision, (n.d.).

[21] Jianbo Shi, Tomasi, Good features to track, in: Proc. IEEE Conf. Comput. Vis. Pattern Recognit. CVPR-94, IEEE Comput. Soc. Press, Seattle, WA, USA, 1994: pp. 593–600. Doi: 10.1109/CVPR.1994.323794.

[22] S. Bhowmick, S. Nagarajaiah, Spatiotemporal compressive sensing of full-field Lagrangian continuous displacement response from optical flow of edge: Identification of full-field dynamic modes, Mech. Syst. Signal Process. 164 (2022) 108232, https://doi.org/10.1016/j.ymssp.2021.108232.

[23] J. Lv, M. Lv, J. Xiao, L. Wen, Q. Lou, A point tracking method of TDDM for vibration measurement and large-scale rotational motion tracking, Measurement 193 (2022) 110827, https://doi.org/10.1016/j.measurement.2022.110827.

[24] V. Hoskere, J.-W. Park, H. Yoon, B.F. Spencer, Vision-Based Modal Survey of Civil Infrastructure Using Unmanned Aerial Vehicles, J. Struct. Eng. 145 (2019) 04019062, https://doi.org/10.1061/(ASCE)ST.1943-541X.0002321.

[25] M.C. Morrone, R.A. Owens, Feature detection from local energy, Pattern Recognit. Lett. 6 (1987) 303–313, https://doi.org/10.1016/0167-8655(87)90013-4.

[26] D.J. Fleet, A.D. Jepson, Computation of component image velocity from local phase information, Int. J. Comput. vis. 5 (1990) 77–104, https://doi.org/10.1007/BF00056772.

[27] M.C. Morrone, J. Ross, D.C. Burr, R. Owens, Mach bands are phase dependent, Nature 324 (1986) 250–253, https://doi.org/10.1038/324250a0.

[28] J.L. Barron, D.J. Fleet, S.S. Beauchemin, Performance of optical flow techniques, Int. J. Comput. vis. 12 (1994) 43–77, https://doi.org/10.1007/BF0142098.

[29] T. Gautama, M.A. Van Hulle, A phase-based approach to the estimation of the optical flow field using spatial filtering, IEEE Trans. Neural Netw. 13 (2002) 1127–1136, https://doi.org/10.1109/TNN.2002.1031944.

[30] N. Wadhwa, M. Rubinstein, F. Durand, W.T. Freeman, Phase-based video motion processing, ACM Trans. Graph. 32 (2013) 1–10, https://doi.org/10.1145/2461912.2461966.

[31] J.G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W.T. Freeman, O. Buyukozturk, Modal identification of simple structures with high-speed video using motion magnification, J. Sound Vib. 345 (2015) 58–71, https://doi.org/10.1016/j.jsv.2015.01.024.

[32] J.G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W.T. Freeman, O. Buyukozturk, Structural Modal Identification Through High Speed Camera Video: Motion Magnification, in: J. De Clerck (Ed.), Top. Modal Anal. Vol. 7, Springer International Publishing, Cham, 2014: pp. 191–197. Doi: 10.1007/978-3-319-04753-9_19.

[33] D.H. Diamond, P.S. Heyns, A.J. Oberholster, Accuracy evaluation of sub-pixel structural vibration measurements through optical flow analysis of a video sequence, Measurement 95 (2017) 166–172, https://doi.org/10.1016/j.measurement.2016.10.021.

[34] S. Collier, T. Dare, Accuracy of phase-based optical flow for vibration extraction, J. Sound Vib. 535 (2022) 117112, https://doi.org/10.1016/j.jsv.2022.117112.

[35] Y. Yang, C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, D. Mascareñas, Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification, Mech. Syst. Signal Process. 85 (2017) 567–590, https://doi.org/10.1016/j.ymssp.2016.08.041.

[36] A. Sarrafi, Z. Mao, C. Niezrecki, P. Poozesh, Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification, J. Sound Vib. 421 (2018) 300–318, https://doi.org/10.1016/j.jsv.2018.01.050.

[37] Y. Miao, J.Y. Jeon, Y. Kong, G. Park, Phase-based displacement measurement on a straight edge using an optimal complex Gabor filter, Mech. Syst. Signal Process. 164 (2022) 108224, https://doi.org/10.1016/j.ymssp.2021.108224.

[38] Y. Wang, W. Hu, J. Teng, Y. Xia, Phase-based motion estimation in complex environments using the illumination-invariant log-Gabor filter, Mech. Syst. Signal Process. 186 (2023) 109847, https://doi.org/10.1016/j.ymssp.2022.109847.

[39] G. Liu, M.Z. Li, Z. Mao, Q.S. Yang, Structural motion estimation via Hilbert transform enhanced phase-based video processing, Mech. Syst. Signal Process. 166 (2022) 108418, https://doi.org/10.1016/j.ymssp.2021.108418.

[40] M.Z. Li, G. Liu, Z. Mao, Q.S. Yang, J.W. Gu, Two-dimensional motion estimation using phase-based image processing with Riesz transform, Mech. Syst. Signal Process. 188 (2023) 110044, https://doi.org/10.1016/j.ymssp.2022.110044.

[41] E.P. Simoncelli, W.T. Freeman, E.H. Adelson, D.J. Heeger, Shiftable multiscale transforms, IEEE Trans. Inf. Theory 38 (1992) 587–607, https://doi.org/10.1109/18.119725.

[42] E.P. Simoncelli, W.T. Freeman, The steerable pyramid: a flexible architecture for multi-scale derivative computation, in: Proc. Int. Conf. Image Process., IEEE Comput. Soc. Press, Washington, DC, USA, 1995: pp. 444–447. Doi: 10.1109/ICIP.1995.537667.

[43] Z. Shang, Z. Shen, Multi-point vibration measurement and mode magnification of civil structures using video-based motion processing, Autom. Constr. 93 (2018) 231–240, https://doi.org/10.1016/j.autcon.2018.05.025.

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

[45] Z. Ma, J. Choi, P. Liu, H. Sohn, Structural displacement estimation by fusing vision camera and accelerometer using hybrid computer vision algorithm and adaptive multi-rate Kalman filter, Autom. Constr. 140 (2022) 104338, https://doi.org/10.1016/j.autcon.2022.104338.

[46] Z. Peng, X. Wang, Z. Wang, W. Liu, M. Liu, Evaluation of the impact of filter types and parameters upon the accuracy of phase-based optical flow method with a complex steerable pyramid, PLoS One 19 (2024) e0308943, https://doi.org/10.1371/journal.pone.0308943.

[47] J. Portilla, E.P. Simoncelli, A Parametric Texture Model Based on Joint Statistics of Complex Wavelet Coefficients, (n.d.).

[48] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (1989) 674–693, https://doi.org/10.1109/34.192463.