



Holistically Nested Edge Detection and particle filtering for subtle vibration extraction

Nicholas A. Valente^a, Zhu Mao^{a,b,*}, Christopher Niezrecki^a

^a Department of Mechanical Engineering, University of Massachusetts Lowell, Lowell, MA 01854, USA

^b Department of Mechanical and Materials Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

ARTICLE INFO

Communicated by Javad Baqersad

Keywords:

Holistically Nested Edge Detection
Motion Magnification
Deep Learning
Particle Filtering
Computer Vision-Based Modal Analysis
Image Processing

ABSTRACT

Extracting and analyzing subtle motion via optical sensing techniques is becoming a more important research area within the vibration community. Computer vision approaches and machine learning architecture have been adopted to contest the complexity concerning hand instrumentation of sensors. Pre-treatment using a stochastic pattern for optical analysis or traditional wired sensing is not always permitted for large structures. In this work, Holistically Nested Edge Detection (HED) is utilized to highlight high-spatial-frequency content features in images for structural dynamic analysis. Following the accentuation of edge features, phase-based motion magnification (PMM) is used to amplify motion that is not visible to the naked eye. This permits the band-passing of higher order resonant frequencies for structural dynamic evaluation. Implementation of two-dimensional particle filtering (PF) is utilized to track the accentuated edge features in the presence of ghosting artifacts that appear post magnification. Within this work, compared to traditional edge-thresholding, the combination of HED and PF (HED-PF) is more proficient to handle noise and ultimately improves magnified image quality and modal parameter extraction. Uncertainties related to using PF for vibration measurement are analytically presented to provide guidance on the behavior of the algorithm. To demonstrate the effectiveness of HED-PF, experimental modal testing takes place on both a tier-structure and cantilever beam to demonstrate the approaches' capability in handling varying sorts of excitation. Finally, introduction of background occlusions and non-ideal lighting scenarios are used to test the limitations of the approach. Ultimately, this will highlight HED-PF's robust tracking capabilities compared to the current state of the art.

1. Introduction

The use of non-contact sensing is a widely investigated research topic in the field of structural dynamics. This is largely due to its non-invasive approach of gathering quantitative data. The mounting of physical sensors, although accurate, requires cabled connections in addition to large data acquisition systems. Recent improvements in the field of optical sensing have permitted extraction of full-field dynamics using camera data [1,2]. Digital Image Correlation (DIC) is a technique that has been adopted to acquire sub-pixel resolution of displacement using speckled patterns applied to a structure, image correlation, and either a single or one or more camera pairs using triangulation. The utilization of two cameras provides a three-dimensional extension of the approach, which can aid in

* Corresponding author at: Department of Mechanical and Materials Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA.
E-mail address: zmao2@wpi.edu (Z. Mao).

extracting out-of-plane displacements and operating or mode shapes [3,4]. This computer vision methodology has been applied to structures such as commercial aircraft [5], industrial components [6], and civil infrastructure [7]. In typical vibration applications, displacement is often unperceivable to the naked eye. Motion magnification, a computer vision technique, was introduced by Liu et al. to amplify subtle motion seen in video [8]. Wu et al. further developed the method by linearly amplifying motion using pixel intensities, which revealed the approach's sensitivity to noise [9]. Taking that information into account, Wadhwa et al. proposed a more noise-robust manner to magnify motion by using an image's phase data [10,11]. Later modifications to the original algorithm took place, such that, choice of filter bank was determined based on the gradient of motion seen in video across pixels [12]. Following amplification of subtle motion, higher-order dynamics can be extracted non-invasively for structural analysis [13–15] and operational modal analysis [16,17]. The Phase-Based Motion Magnification (PMM) algorithm has had a wide array of applications including the extraction of structural dynamics wind-turbine blades [18,19] and structural health monitoring [20–24]. The capturing of subtle motion is not limited solely to the engineering domain. This technique has proven to be impactful in the medical setting, where slight motions are observed during surgical and medical diagnostics [25–27]. Large amounts of amplification ultimately creates spatial distortion artifacts that contaminate magnified data [28]. Adoption of both spatial filtering techniques [29–31] and machine learning models [32–35] has proved to be effective in limiting the presence of these ghosting artifacts. There typically exists a tradeoff in being able to amplify subtle structural motion and distorting the image, thereby preventing the quantification of the true motion [28]. The focus of this paper is to develop a new method of motion quantification that can address this issue.

The tracking of a particular region of interest in a video has been commonly achieved using correlation techniques followed by some form of interpolation [36–39]. This requires computation of dynamic motion for each pixel within an image throughout a series of frames. To address this challenge, various forms of edge-detection and target tracking have been utilized to extract dynamic behavior from structures [40]; however, hand-selection of pixels that are of interest is still required. Adaptive thresholding and other forms of image binarization have shown to be effective in highlighting edges, but due to the nature of background occlusions some pixels of interest may be discarded. Holistically Nested Edge Detection (HED) is a computer vision approach that was generated; such that, prominent background and foreground pixels are not discarded [41]. With most edges defined throughout a sequence of frames, extraction of amplified subtle motion is attainable. Particle filtering, a stochastic Monte Carlo process, is a customizable algorithm that aids in tracking objects seen in video using pixel intensities in color channels [42–44]. This methodology has been extended to structural dynamic applications; however, pre-treatment of a structure was necessary for modal parameter extraction [45,46]. Within this work, rather than relying on optical targets or a stochastic pattern, a two-dimensional particle filter (PF) is utilized to track edge features that are accentuated using HED.

The novelty of the proposed work is utilizing a hybrid methodology, Holistically Nested Edge Detection Particle Filtering (HED-PF), to extract phase-based magnified time data that can be used for improved non-contact experimental modal analysis. This approach does not rely on the use of a mounted optical target; rather, the method relies on the prominent high-spatial-frequency edge features that result following the use of HED. Derivation of the statistical parameters related to the proposed methodology is also provided to evaluate the effectiveness and accuracy of HED-PF in the presence of noise. The proposed approach is compared to current state-of-the-art methodologies for conducting experimental modal analysis after magnifying motion in particular frequency bands. This includes the use of two-dimensional point tracking, phase motion estimation (PME), and traditional edge-thresholding for magnified motion extraction. Experimental testing takes place on a tier-structure for extraction of magnified displacement. The approach's capabilities are then extended to extract operational deflection shapes (ODS) from a cantilever beam in the presence of distracting background scenery. The experimental modal data is compared to analytical mode shapes obtained using a finite element model. The remainder of the work is organized as follows: theoretical background, uncertainty quantification of two-dimensional particle filtering, experimental validation, followed by conclusions.

2. Theoretical background

2.1. Holistically Nested Edge Detection (HED)

Edge detection is a common practice in computer vision for identifying high-frequency features that are present in an image. Holistically Nested Edge Detection (HED) is a deep learning based approach, which utilizes multi-layered convolutional neural networks to produce edge-defining images [41]. Several factors such as lighting conditions and cluttered background scenery can affect the binary image that ensues post convolution. The process begins by taking a raw image input, X , and subsequently computing a binary edge map, Y . To differentiate from traditional edge detection methodologies, HED is formulated, such that, background and foreground features are differentiated from one another. Convolutional neural networks (CNN) are typically employed to learn low dimensional patterns seen in image data. There are a multitude of image layers that are computed using a convolution neural network (CNN) to accentuate prominent and subtle edge features. Here, Equations (1)–(4) summarize the work conducted by Xie et al. Equation (1) expresses the objective function for the collection of all network layer parameters, W , with corresponding weights, w , for each side-output layer's classifier.

$$\mathcal{L}_{\text{side}}(W, w) = \sum_{q=1}^Q \beta_q \mathcal{L}_{\text{side}}^{(q)}(W, w^{(q)}) \quad (1)$$

Xie et al. defines the objective function $\mathcal{L}_{\text{side}}$ for Q number of side-output layers. For simplistic indexing, q represents one individual side-output layer that totals to Q number of layers. Each layer also contains an associative learning rate β_q and an image level loss function $\mathcal{L}_{\text{side}}^{(q)}$ respectively. Following the computation of individual $\mathcal{L}_{\text{side}}$ layers, a weighted-fusion layer $\mathcal{L}_{\text{fuse}}$ is computed to gain

insight into the weights associated with each $\mathcal{L}_{\text{side}}$ layer. The weights computed give the network insight into what features are deemed important as the network trains.

$$\mathcal{L}_{\text{fuse}}(\mathbf{W}, \mathbf{w}, \mathbf{h}) = \text{Dist}\left(Y, \hat{Y}_{\text{fuse}} = \text{sigmoid}\left(\sum_{q=1}^Q h_q A_{\text{side}}^{(q)}\right)\right) \quad (2)$$

In Equation (2), the $\hat{Y}_{\text{fuse}}$ term is comprised of edge map predictions made by the CNN model. The activation functions of each side output layer $A_{\text{side}}^{(q)}$ and the associate fusion weight h_q are computed over the sigmoid function where classification of an edge feature is bounded using a probability bounded between 0 and 1. The $\text{Dist}(Y, \hat{Y}_{\text{fuse}})$ between the network's predictions and the binary edge map are then compared using cross-entropy loss to determine the accuracy of edge classification. Utilizing both Equation (1) and (2), Xie et al. minimizes the objective function using stochastic gradient descent.

$$(\mathbf{W}, \mathbf{w}, \mathbf{h})^* = \underset{[\mathcal{L}_{\text{side}}(\mathbf{W}, \mathbf{w}), \mathcal{L}_{\text{fuse}}(\mathbf{W}, \mathbf{w}, \mathbf{h})]}{\text{argmin}} \left(\mathcal{L}_{\text{side}}(\mathbf{W}, \mathbf{w}) + \mathcal{L}_{\text{fuse}}(\mathbf{W}, \mathbf{w}, \mathbf{h}) \right) \quad (3)$$

The updated weights $(\mathbf{W}, \mathbf{w}, \mathbf{h})^*$ are then used for testing where the output from the CNN results in edge maps that are computed for each fusion $\hat{Y}_{\text{fuse}}$ and side $\hat{Y}_{\text{side}}$ layer. An averaging of each side and fusion layer results in the final edge-defined output $\hat{Y}_{\text{HED}}$.

$$\hat{Y}_{\text{HED}} = \sum_{q=1}^Q \left(\hat{Y}_{\text{side}}, \hat{Y}_{\text{fuse}}^{(1)}, \dots, \hat{Y}_{\text{fuse}}^{(q)} \right) \quad (4)$$

Fig. 1 displays the machine learning approach that is proposed by Xie et al. [41].

The goal of HED is to locate edge defining pixels and distribute an associative weight with respect to the sum of all pixels found in an image. The resulting image in Fig. 1(c) contains highlighted pixels that are characterized as edge features. Unlike traditional edge detection techniques, as shown in Fig. 1(b), HED can preserve edge features regardless of insufficient lighting scenarios. The rubber duck was chosen for this example to display how HED can address a variety of textures that typically cause issues with other computer vision techniques. Specifically, the high frequency spatial features that are present on the duck would typically be difficult to capture if using traditional thresholding. It should be noted that for this example a monochromatic background was chosen to highlight the ability of the machine learning approach. HED ultimately simplifies the edge-feature tracking problem due to its ability to limit pixel discontinuities. The HED processed image can accentuate the edge features, while limiting the number of discarded pixels following appropriate thresholding. This fact aids in using pixels for displacement measurement without having to rely on pixel interpolation.

2.2. Phase-Motion estimation (PME)

For dynamic motion, pixel intensities found in the images will vary over a period of data collection. The phase-valued complex map can be used to track the motion between sequential frames. Consider a magnified image $\tilde{I}(x, y, t_n)$ at a specific location within the image (x, y) . In the next frame, sequential motion is expressed as some change in location $(\Delta x, \Delta y)$ applied to the previous image where $\tilde{I}(x + \Delta x, y + \Delta y, t_{n+1})$. For simplification purposes, brightness consistency is assumed such that the illumination of an object is constant from t_n to t_{n+1} . A Gabor wavelet is used to decompose images into their respective amplitude and phase components in the complex domain [47]. As described by [48], the kernel for the Gabor wavelet $g(x, y; \lambda, \theta, \psi, \sigma_g, \gamma)$ is defined as

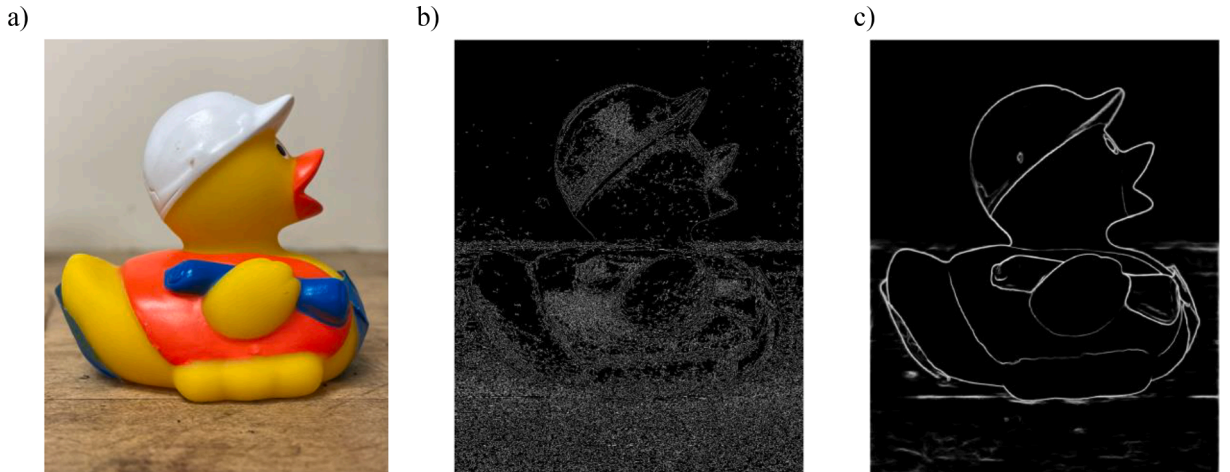


Fig. 1. (a) Original captured image of Ernie the Engineer; Computation and highlight of high-frequency spatial features using (b) Canny edge detector and (c) Holistically Nested Edge Detection (HED) output as a result of Equation (4).

$$g(x, y; \lambda, \theta, \psi, \sigma_g, \gamma) = e^{\left(-\frac{x^2 + y^2}{2\sigma_g^2} \right)} e^{i \left(2\pi \frac{x}{\lambda} + \psi \right)}. \quad (5)$$

The kernel is defined by a sequence of parameters where: γ is the spatial aspect ratio of the kernel, λ is the wavelength of the kernel, σ_g is the standard deviation of the Gaussian envelope, ψ is the phase offset, and θ is the orientation of the wavelet. The orientation of the kernel can be expressed concisely as $(x', y') = (-x \bullet \cos\theta + y \bullet \sin\theta, x \bullet \sin\theta + y \bullet \cos\theta)$. Motion capture using phase values is determined based on the Gabor kernel's orientation. For example, a Gabor wavelet containing $\theta = 0^\circ$ is well suited to capture horizontal motion between sequential images as shown in Fig. 2.

The first stage of PME is to decompose each of the magnified frames $\tilde{I}$ captured in video into the complex domain using the following convolution

$$C(u, v, t_n) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{I}(x, y, t_n) e^{\left(-\frac{(x-u)^2 + y^2}{2\sigma_g^2} \right)} e^{i \left(2\pi \frac{x-u}{\lambda} + \psi \right)} dx dy. \quad (6)$$

Similarly to Equation (6), the complex coefficient can be expressed forward in time as

$$C(u, v, t_{n+1}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \tilde{I}(x + \Delta x, y + \Delta y, t_{n+1}) e^{\left(-\frac{(x-u)^2 + y^2}{2\sigma_g^2} \right)} e^{i \left(2\pi \frac{x-u}{\lambda} + \psi \right)} dx dy. \quad (7)$$

After mathematical manipulation, as derived in its entirety by [18], the complex coefficients $C(u, v, t_n)$ and $C(u, v, t_{n+1})$ are expressed in the rotational plane as

$$C(u, v, t_n) = A_{t_n}(u, v, t_n) e^{i\tilde{\phi}(u, v, t_n)}, C(u, v, t_{n+1}) = A_{t_{n+1}}(u, v, t_{n+1}) e^{i\tilde{\phi}(u, v, t_{n+1})}. \quad (8)$$

From Equation (8), true motion between two consecutive images can be approximated using the difference in phase [18]:

$$\begin{aligned} \Delta x \propto \tilde{\phi}(u, v, t_{n+1}) - \tilde{\phi}(u, v, t_n) & \quad \text{if } \theta = 0 \\ \Delta y \propto \tilde{\phi}(u, v, t_{n+1}) - \tilde{\phi}(u, v, t_n) & \quad \text{if } \theta = \frac{\pi}{2} \end{aligned} \quad (9)$$

The extracted phase is cyclic in nature and therefore its values are bounded from $(-\pi, \pi)$. This boundedness limits how much motion can be captured between sequential frames.

2.3. Phase-Based Motion Magnification (PMM)

The amplification of subtle motion in video permits a non-invasive way of showing higher order dynamic behavior of structures [49–51]. The phase-based approach, proposed by Wadhwa et al. modified the original motion magnification technique to acquire more

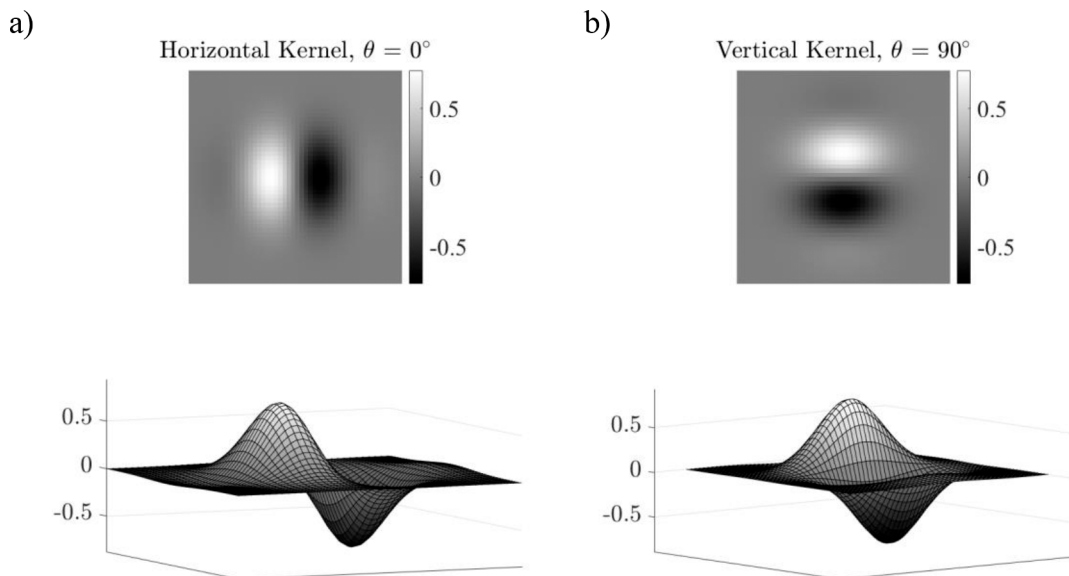


Fig. 2. Gabor wavelet designed for capturing motion in the (a) horizontal, $\theta = 0^\circ$ and (b) vertical, $\theta = 90^\circ$ direction respectively.

robust noise handling capabilities [10]. The PMM approach is now concisely described, and the detailed mathematical formulation for the approach can be found in [8–10]. Fleet and Jepson showed that phase variation in images is proportional to true motion [52]. Wadhwa et al. temporally filtered and amplified phase-variations found in images $\phi(t)$ by applying a linear band-pass filter, B of a single frequency ω . In the time domain, where the time index t ranges from 1 to n , the filtered (u, v) phase plane can be expressed as

$$B_\omega(\phi(u, v, t_{1:n})) = \omega \bullet \phi(u, v, t_{1:n}) \quad (10)$$

Once filtered, a reconstruction of the altered images takes place where the phase is re-compiled with associative amplitude, $A(u, v, t_{1:n})$ which results in a complex valued image. The phase information is amplified by a magnification factor, α , that exaggerates subtle motion within a particular frequency band.

$$\tilde{I}(x, y, t_{1:n}) = A(u, v, t_{1:n}) e^{i(B_\omega(\phi(u, v, t_{1:n})) + (1+\alpha)B_\omega(\phi(u, v, t_{1:n})))} \quad (11)$$

Equation (11) expresses the series of reconstructed frames $\tilde{I}$, which reveals subtle motion within a specified frequency band. For this application, the frequency band is determined by utilizing the a priori information that comes because of an experimental dynamic test. It should be noted that an increase in α is known to not only exaggerate motion, but also produce ringing effects that could make quantification of amplified motion more complex [28].

2.4. Two-dimensional particle filter

Particle filtering is commonly used in estimation theory with several applications in multiple object tracking [43,53,54]. Common algorithms such as Kalman filters have shown success in extracting dynamic motion; however, they are limited to linear dynamics and Gaussian noise [54]. The benefit of using the particle filtering algorithm, is the ability to estimate high-dimensional state systems in the presence of noise. Particle filters are used to estimate system dynamics or states, which can be used to predict further states as they change in time [55]. There are three major components of the particle filtering algorithm, which include: the predictive state, observation state, and assignment of particle weights. The sequential Monte Carlo method begins by building a discrete dynamic model for each time step t , which is used to estimate a state vector $\mathbf{x}_t$, where

$$\mathbf{x}_t = F(\mathbf{x}_{t-1}, \mathbf{w}). \quad (12)$$

The state transition model, F , is typically chosen as a design parameter to accurately observe position and velocity quantities. This matrix is defined as a constant velocity model for this application; however, investigation into a better suited model to aid convergence is beyond the scope of this work. In Equation (12), $\mathbf{w}$ represents the associative process noise that is seen during the transitional procedure of predicting particles from state to state. Following the development of a state transition model, a discrete observation model is formulated based on prior measurements of particles. The observed vector of states, $\mathbf{z}_t$, is expressed in terms of an observed transition model H and measurement noise $\mathbf{v}$.

$$\mathbf{z}_t = H(\mathbf{x}_t, \mathbf{v}) \quad (13)$$

The collection of available observed measurements can be expressed as $\mathbf{z}_{1:t-1}$. Ultimately, the goal is to estimate the current state of the dynamic system, $\mathbf{x}_t$ given a collection of observed measurements in terms of a probability density $p(\mathbf{x}_t | \mathbf{z}_{1:t})$. A Bayesian estimation permits computation of $p(\mathbf{x}_t | \mathbf{z}_{1:t})$ in terms of the posterior density at a previous time step where

$$p(\mathbf{x}_t | \mathbf{z}_{1:t}) = \frac{p(\mathbf{z}_t | \mathbf{x}_t) p(\mathbf{x}_t | \mathbf{z}_{1:t-1})}{p(\mathbf{z}_t | \mathbf{z}_{1:t-1})}. \quad (14)$$

To do so, there is a need to compute both a prediction and update step for the sequential stochastic process. The normalizing of the collected observed measurements with respect to a previous time-step is defined as

$$p(\mathbf{z}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{z}_{t-1}) p(\mathbf{z}_t | \mathbf{x}_t) d\mathbf{x}_t. \quad (15)$$

Computationally, it is difficult to calculate the distribution analytically due to high-dimensional integration. The iterative Monte Carlo process generates random samples (i.e., particles) taking into account the prior state, $t-1$ for each particle m where

$$\mathbf{x}_{t-1}^{[m]}, \{m = 1 : M\}. \quad (16)$$

An investigation into how the total number of particles M ultimately plays a role in convergence is analyzed in the uncertainty quantification portion of this work. For this context, $t = 1$ represents the reference frame and $t = n$ is the last frame in the sequence of magnified HED images. After an observation is collected at a specific time step, the histogram similarity is evaluated according to the dynamic state of each particle. The observation model shows the log-likelihood that a measurement of a single particle $\mathbf{z}_t^{[m]}$ appears given a change in state $\mathbf{x}_t$ with respect to a red, green, blue (RGB) intensity target value or $RGB_T = [R_T, G_T, B_T]$ and its associative variance σ_{RGB} . These intensity targets range from 0 to 255 where the minimum and maximum are black and white respectively. The pixels observed in the image are represented by $RGB_p = [R_p, G_p, B_p]$.

$$p(\mathbf{z}_t | \mathbf{x}_t^{[m]}) = \frac{1}{\sqrt{2\pi}\sigma_{RGB}} e^{-\frac{(\zeta_p - \mu)^2}{2\sigma_{RGB}^2}}, \zeta = \|RGB_T - RGB_p\|_2 \quad (17)$$

Equation (17) expresses the log-likelihood function where ζ represents the Euclidean distance between RGB_T and RGB_P . The newly generated samples are a realization of the predicted probability distribution $p(x_t | z_{1:t-1})$. Given the measurements z_t , the update process provides each sampled particle at a time stamp $x_t^{[m]}$ with an associative weight according to Equation (17). The weight for each sampled particle $\xi_t^{[m]}$, is expressed as

$$\xi_t^{[m]} = \frac{p(z_t | x_t^{[m]})}{\sum_{m=1}^M p(z_t | x_t^{[m]})} \quad (18)$$

The pixel-value based particle filter is designed such that the most dominant geometric features in an image will be tracked in a particular region of interest. Fig. 3 displays the sequential steps associated with the particle filtering process.

Due to the contrast rich images and accentuation of edge features outputted by HED, sub-pixel displacement can be captured at each time step i by estimating the expected value of the generated two-dimensional point cloud of particles. In short, the generated samples are sequentially updated to minimize the distance between the criteria and observed measurements. Once the particle is plotted in the two-dimensional cartesian plane, its displacement coordinates are stored for each iteration of the tracking process. The expected values of the displacement coordinates are used to gather subpixel resolution of motion. This prevents having to rely on lengthy correlation and subsequent bilinear interpolation used in two-dimensional target tracking.

3. Maximum likelihood evaluation of two-dimensional particle filter

3.1. Formulation of expected value and associate variance

The particle filtering algorithm has been utilized in multiple object tracking; however, an understanding of the number of particles and variance associated with vibration measurements has yet to be discussed. From the previous section, it was noted that following the assignment of particle weight a two-dimensional point cloud is formed around a target pixel value. The likelihood function, expressed in Equation (17), can be used to estimate the center or expected value of the dense point cloud with M number of particles where

$$f(\zeta_{1:M}) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^M e^{-\sum_{j=1}^M \frac{(\zeta_j - \mu)^2}{2\sigma^2}} \quad (19)$$

Equation (19) expresses the likelihood measurement for each sample generated by the particle filter. Taking the natural logarithm of both sides of Equation (19) results in

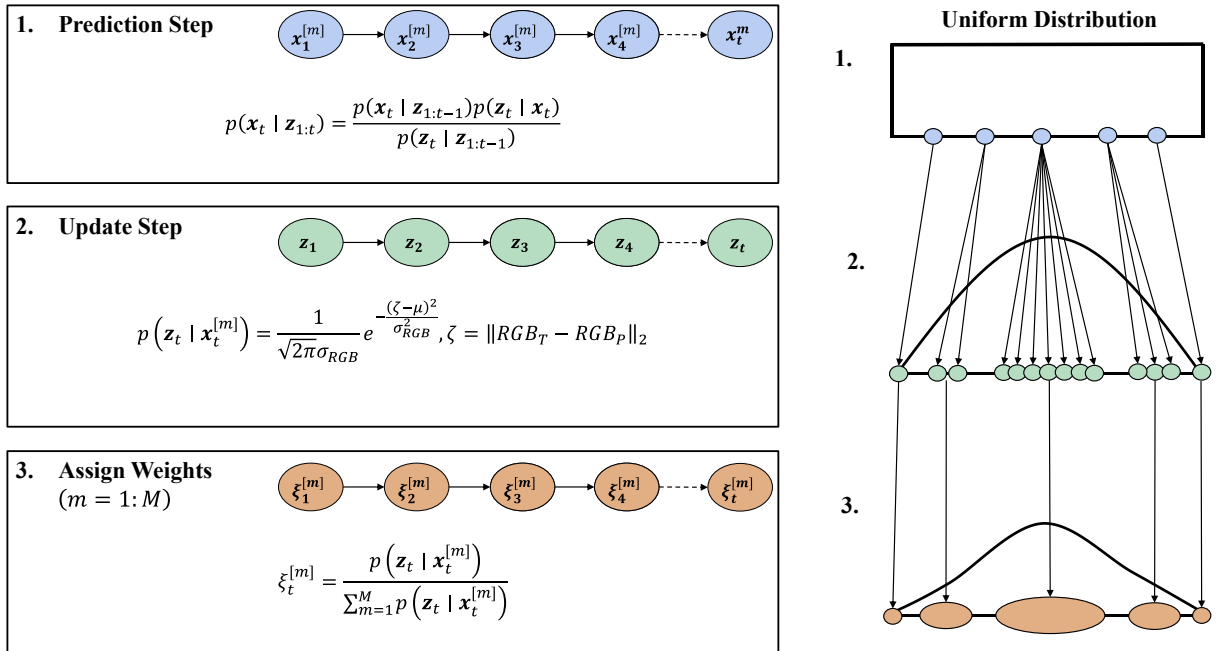


Fig. 3. Visual representation of using the particle filter for the three individual steps: likelihood, updating, and weight assignment.

$$\ln f(\zeta_{1:M}) = -\frac{M}{2} \ln(2\pi) - M \ln(\sigma) - \sum_{j=1}^M \frac{(\zeta_j - \mu)^2}{2\sigma^2}. \quad (20)$$

From this point, an estimation of the mean $\hat{\mu}_c$ and variance $\hat{\sigma}_c$ associated with the two-dimensional point cloud cluster can be computed. This minimization of both the mean and standard deviation will ultimately provide the maximum likelihood that the predictions match the RGB_r value. This is done by taking the partial derivative of the likelihood function by its two variables, μ and σ_{RGB} where

$$\frac{\partial f(\zeta_{1:M})}{\partial \mu} = \sum_{j=1}^M \frac{(\zeta_j - \mu)}{\sigma^2} = 0, \quad (21)$$

and

$$\frac{\partial f(\zeta_{1:M})}{\partial \sigma} = -\frac{M}{\sigma} + \sum_{j=1}^M \frac{(\zeta_j - \mu)^2}{\sigma^3} = 0 \quad (22)$$

Setting both Equations (21) and (22) to zero results in the derivation of the expected value $\hat{\mu}_c$ and variance $\hat{\sigma}_c$ of the two-dimensional point cloud.

$$\hat{\mu}_c = \frac{\sum_{j=1}^M \zeta_j}{M} \quad (23)$$

$$\hat{\sigma}_c^2 = \frac{\sum_{j=1}^M (\zeta_j - \mu)^2}{M} \quad (24)$$

Now that the two-dimensional point cloud parameters have been established, $\hat{\mu}_c$ will be used to estimate the sub-pixel displacement of the tracked region from frame to frame. When glancing at Equation (24), it can be deduced that the variance seen in displacement measurement is a function of the number of particles. The variance ultimately plays a key role in determining how effective magnified motion can be captured in video.

3.2. Traditional tracking approaches on perceivable motion

Displacement amplitudes in most vibration applications are very small, which motivates the need for motion amplification. In some cases, displacement between sequential frames may be perceivable which makes computer vision tracking more simplistic. A perceivable motion test, as depicted in Fig. 4(a), was conducted to portray the HED-PF's accuracy compared to current state-of-the-art methodologies in the presence of large motion. Here, the optical target mounted to the rubber duck will be tracked using the proposed approach, PME, and traditional edge thresholding. The magnitude of horizontal motion was measured to be 22 (cm).

It can be seen in Fig. 4(b) that all three computer vision methodologies are successful in tracking large displacement between sequential frames. The images of interest were uniformly cropped, such that solely the optical target was present in the frame. This ultimately ensures that each technique is presented with the same pixels of interest to track over the sequence of frames. This accuracy

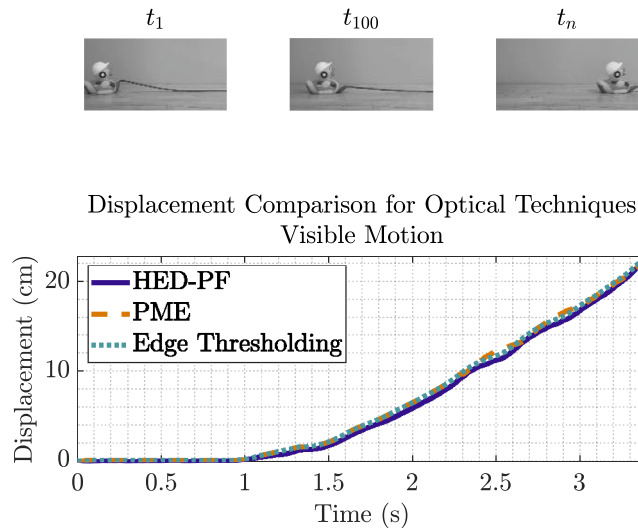


Fig. 4. (a) Perceivable motion of Ernie the Engineer rubber duck subjected to displacement in the horizontal direction; (b) Displacement comparison of the optical target between the proposed HED-PF and current state-of-the-art methodologies (i.e., PME and Edge Thresholding).

can be attributed to the high signal-to-noise ratio (SNR) that exists for the optical target placed on the rubber duck. The unmodified optical target makes tracking via computer vision techniques simplistic due to the minimal amount of noise contaminating the edge features. In the case where motion amplification is necessary, the degradation of high frequency spatial features (i.e. edges) will ultimately affect the accuracy of displacement estimation [28]. As presented in the following section, tracking amplified regions of interest poses a more complex tracking scenario when addressing large motion amplification artifacts.

3.3. Synthetic data measurement – amplified motion

To highlight the usefulness of the HED-PF approach, a synthetic data set is generated where a region of interest is created and oscillates diagonally in the (x,y) plane. Fig. 5 shows the rigid region of interest that displaces in both directions following a $f(x) = 2\sin(10\pi t)$ trajectory.

Once magnification of subtle motion takes place, ghosting artifacts begin to appear at the extrema of motion seen in an image as shown in Fig. 5(b). The grayscale variation that appears as a result of the ringing effects ultimately compromises the effectiveness of traditional edge-detection. Pixel dropouts and interpolation techniques are necessary following the use of traditional edge detection due to the pixel thresholding that takes place. Following thresholding, the calculation of edge locations becomes more cumbersome due to pixel discontinuity which can produce incorrect displacement estimation. Fig. 6 displays the differences between traditional edge-detection compared to the HED-PF combined approach.

Comparing Fig. 6(c) and Fig. 6(d), the amplitude associated with the magnified displacement calculation between sequential frames is shown to contain less noise when utilizing HED-PF. This can be attributed to the averaging of multiple particle locations that takes place in order to compute the sub-pixel coordinate at a given time step. Following the use of HED, edge features are connected at each boundary point, which results in less pixel discontinuity and the need for pixel interpolation. In regard to this particular synthetic data set, $M = 30,000$ (particles). The number of particles is ultimately a design parameter that is determined by the user as an input. Referencing Equation (24), the variance $\hat{\sigma}_c^2$ is a function of M ; therefore, an increase of number of particles produces more accurate results. It should be noted that an increase in particles, although limits variation in the observed data, results in larger computation time [56]. In future sections, the number of particles will be seen as an important parameter when approximating magnified time series. Determination of an optimal number of particles should be carefully studied as a function of convergence time, however, it is beyond the scope of this work.

4. Dynamic testing and analysis using HED-PF

Now that the theoretical background has been developed and introduced, the HED-PF methodology is used to extract dynamic content from both a tier structure in addition to a cantilever beam under non-ideal testing scenarios. Fig. 7 outlines the algorithm methodology that is used to extract subtle dynamics seen in captured video with the use of four discrete steps.

The robustness of this methodology is demonstrated by including background occlusions within both sets of tests. Section 4.1 draws comparison between the proposed methodology and the current state of the art methods such as: 2-D point tracking, Phase-Motion Estimation (PME), and edge thresholding. Section 4.2 extends the proposed methodology for computation of operational deflection shapes that come as a result of magnified time-series extraction.

4.1. Tier-structure

To draw comparisons between HED-PF to the current state of the art, hand-speckle interferometry is applied to a tier-like structure. In addition, optical targets are mounted to the structure; such that, correlation techniques in addition to interpolation can be used to extract magnified displacement. As shown in [18,19], 2-D point tracking is an effective way of extracting exaggerated motion within a particular frequency band. This will serve as the benchmark measurement that will be used to compare the effectiveness of the alternative computer vision methodologies. Fig. 8 depicts the tier structure in addition to the frequency response measurement from the gathered acceleration data following random excitation. A Modal Shop shaker outputs the random excitation to the bottom floor of

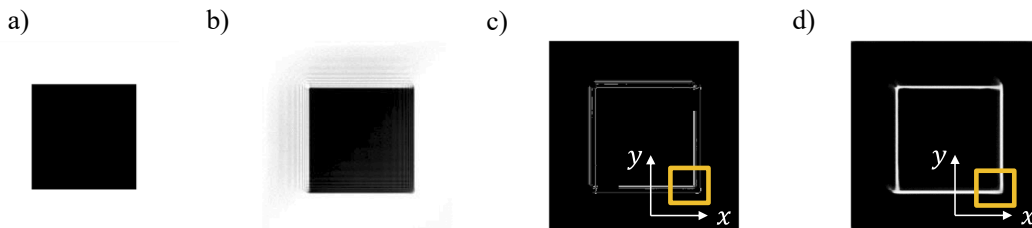


Fig. 5. Generated simulation of a 220×220 (pixel) square in a 400×400 (pixel) image with 2 (pixel) peak to peak motion; (a) is the raw data; (b) magnified image where $\alpha = 20$; (c) traditional edge-detection of a magnified image using the Canny edge-detection model; (d) HED of a magnified image. In both (c) and (d), the yellow box is to highlight the rightmost edge of the oscillating geometry. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

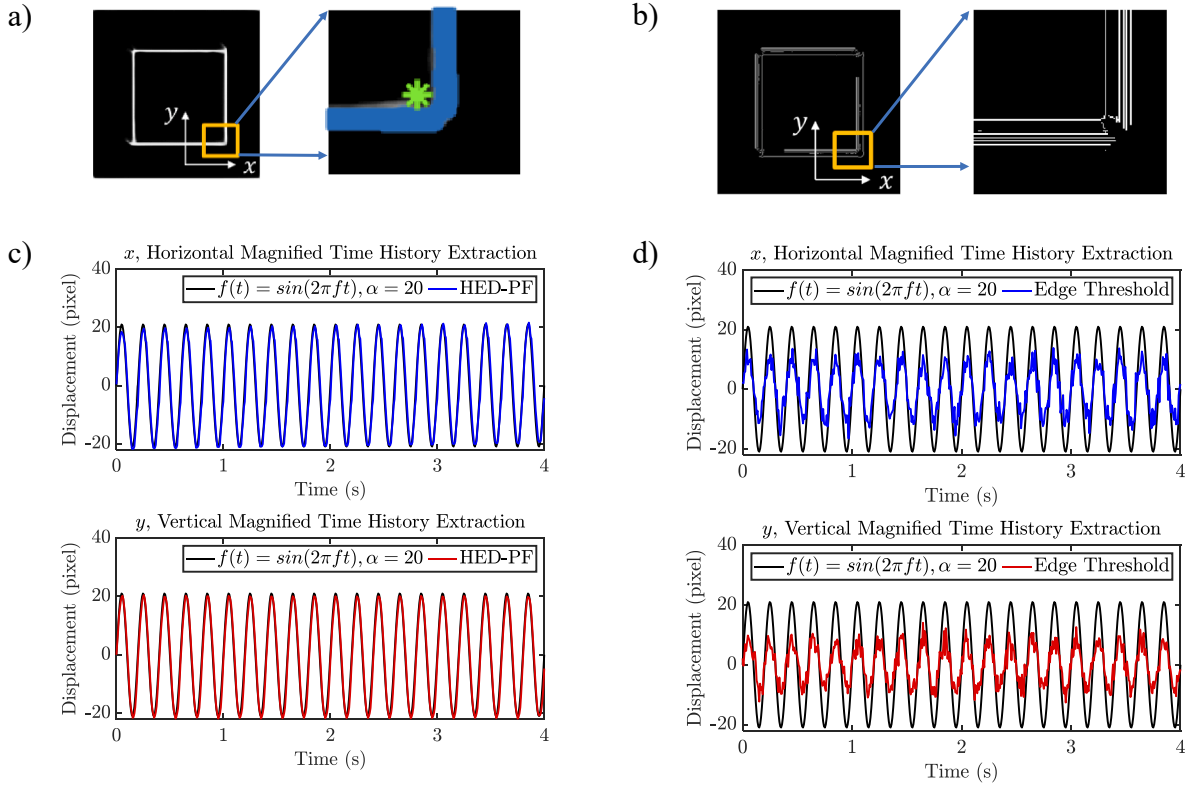


Fig. 6. (a) HED-PF for tracking a magnified edge feature using: $RGB_T = [255, 255, 255]$, $M = 30,000$ (particles). The two-dimensional point cloud is represented by the blue particles, while the green marker is the centroid, μ_c , of the generated point cloud; (b) Traditional edge-detection with thresholding to highlight the pixel discontinuities that appear at a magnified boundary; (c) Magnified time series extraction of edge-feature using HED-PF; (d) Magnified time series extraction of edge-feature using traditional edge-detection with thresholding. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Data Collection and Pre-Processing of Captured Images

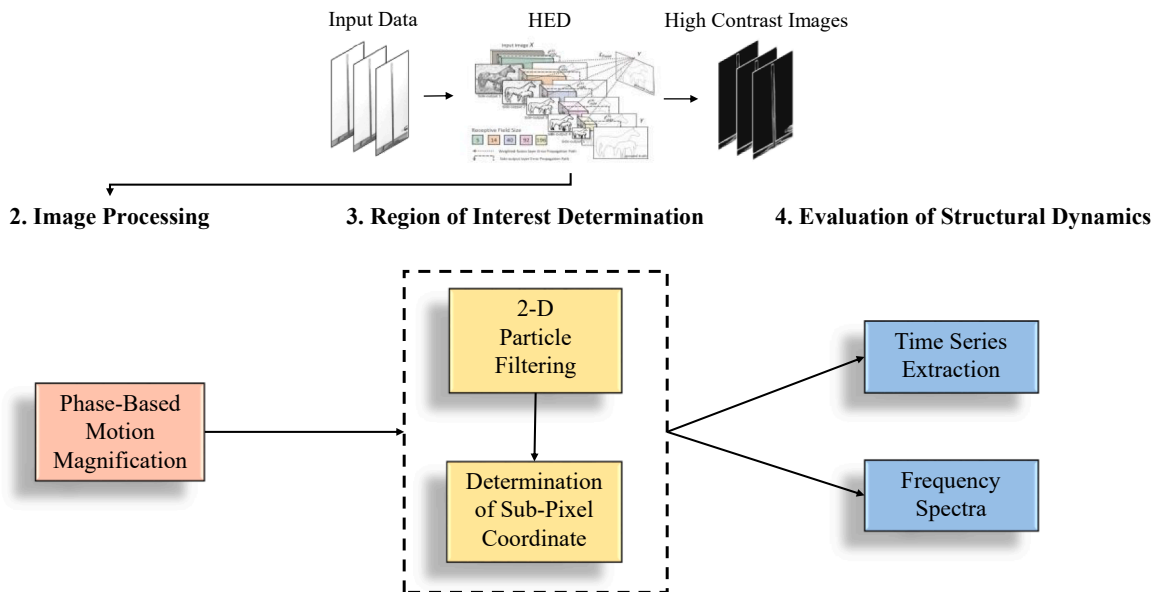


Fig. 7. Pre-trained machine learning architecture [41] and algorithm methodology for extraction of structural dynamics of varying structures.

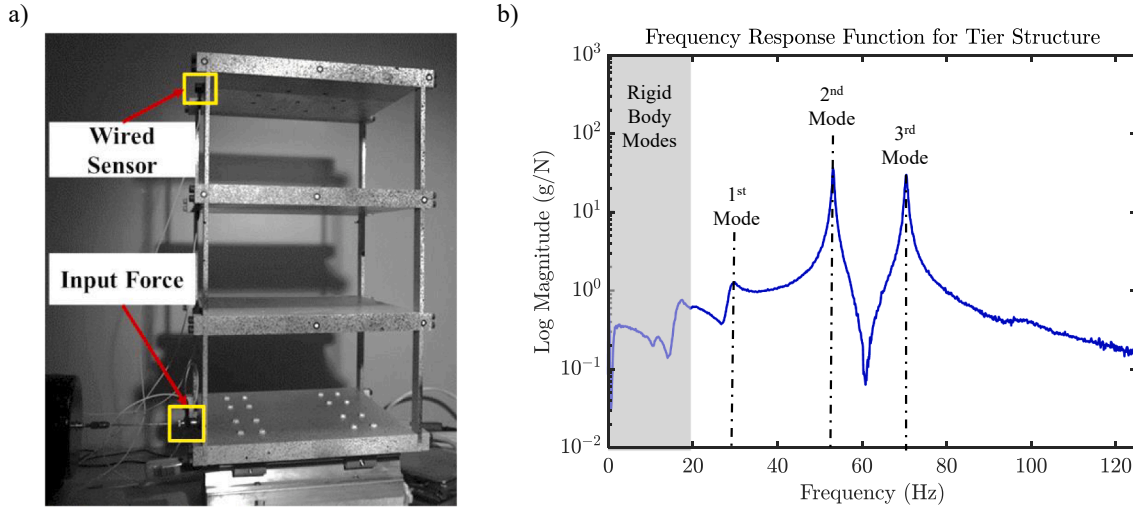


Fig. 8. (a) Tier structure with hand speckled pattern and optical targets mounted; (b) Frequency response function (FRF) of acceleration data compared to Gaussian white noise excitation.

the structure with an associative variance of 0.1. Due to the boundary conditions of this structure, there are two rigid body modes associated with this structure's dynamics. In this case, these rigid body modes include the side-to-side and rocking motion of the structure. Also, it is important to note that the tier structure is situated on rubber stoppers that sit atop a non-rigid wooden laboratory bench.

As outlined in Fig. 7, magnification of the high contrast images is a pre-processing step prior to extraction of exaggerated motion. The subtle motion is captured using a PHOTRON high-speed camera that captures data at 250 (fps). Table 1 outlines the PMM parameters that are used to isolate and exaggerate motions in particular frequency bands.

A region of interest determination is made once the high contrast images are magnified in a particular frequency band.

Fig. 9 displays the manner in which a region of interest is evaluated using the proposed HED-PF. Unlike the current state of the art, HED-PF does not require the lengthy computation of flow vectors and dynamics associated with each pixel. It should be noted, however, that techniques such as optical flow need not to be applied across the full scale of the image if motion is confined to a small number of pixels. The particle filter reduces the amount of calculation necessary by limiting its computation to a smaller number of pixels within a region of interest. Even on a smaller facet scale, the particle filter will solely be focused on the pixels that fulfill the criteria outlined by the likelihood function. The scalability of the particle filter could permit capturing of full-scale data; however, that is beyond the scope of this work. The system states being estimated are defined as the instantaneous location of the moving interest points, and at each iteration, the state is sampled based upon prior knowledge or a previous image. To further verify the uncertainty quantification of HED-PF, the number of particles used for exaggerated motion extraction is varied for the tier structure as shown in Fig. 10. It can be seen in Fig. 11, the Time Response Assurance Criterion (TRAC) values degrade if a sparse number of particles is used to approximate motion found in video. The metric for TRAC is defined as

$$TRAC = \frac{\left(\tilde{z}_m^T z_m \right)^2}{\left(\tilde{z}_m^T \tilde{z}_m \right) \left(z_m^T z_m \right)}. \quad (25)$$

In Equation (25), z_m is the computer vision approach's time series while $\tilde{z}_m$ is the time series extracted by the Digital Image Correlation Engine (DICE). The analysis and optimization mode were set to tracking and gradient-based respectively for computation of a discrete time series using DICE. TRAC values range from 0 to 1, where 0 and 1 represent no correlation and perfect correlation respectively. This finding correlates well with Equation (24), where the variation of the particles, $\hat{\sigma}_c^2$, is dictated by the number of samples generated. A sparse sampling produces a point-cloud that may contain outliers, which can greatly affect point-cloud centroid calculation and subsequent motion estimation. As the mode of vibration increases, more magnification is necessary to make

Table 1
Phase-Based Motion Magnification (PMM) parameters for the tier structure.

Mode	Frequency Band $\omega_l - \omega_h$ (Hz)	Magnification Factor α
1st Bending Mode	29.00 – 34.00	50
2nd Bending Mode	50.50 – 55.50	75
3rd Bending Mode	68.50 – 73.50	100

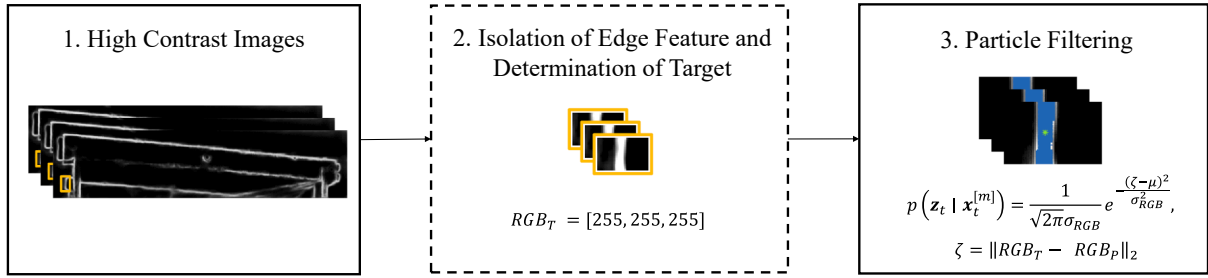


Fig. 9. Isolation of region of interest and use of particle filtering for sub-pixel accuracy of magnified displacement.

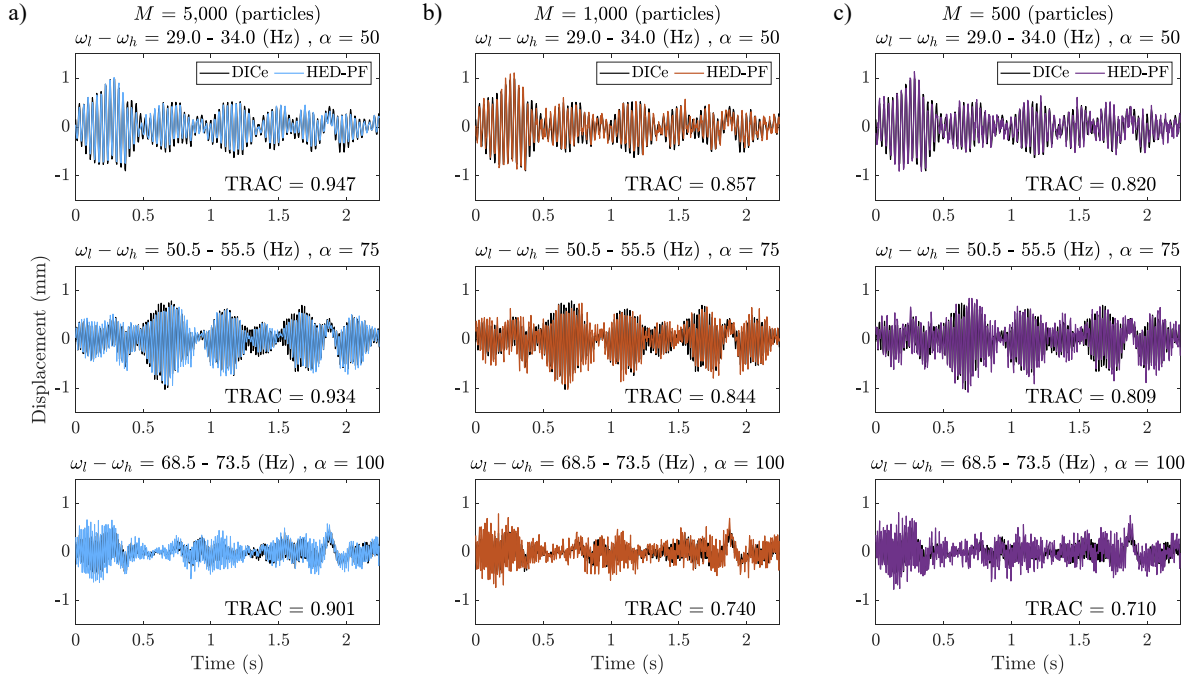


Fig. 10. Accuracy in magnified time series extraction of the specified region of interest shown in Fig. 9, for the first three bending modes using parameters outlined in Table 1; (a) $M = 5,000$ (particles); (b) $M = 1,000$ (particles); (c) $M = 500$ (particles).

imperceptible motion for higher order modes more visible. Although this makes subtle motion more apparent, quantification may be altered due to the degree of magnification [28]. It has been seen that the particle filter has various challenges that contribute to its main sources of error; however, a fine tuning of the number of particles works to combat inaccuracy in the motion estimation process [56].

Fig. 12 compares 2-D point tracking with the current state-of-the-art computer vision techniques such as PME and Edge-Thresholding. In the PME case, the direction of the Gabor kernel is oriented such that motion in the horizontal direction can be captured throughout the sequence of frames. Additionally, adaptive thresholding is used to create a binary image which highlights prominent edge features which can subsequently be tracked. DICE is used to gather sub-pixel resolution using the mounted optical targets, which is considered to be a baseline metric of comparison [57].

By inspecting Fig. 12, HED-PF is compared to current state-of-the-art methodologies such as PME and edge thresholding. Using a sample size of $M = 5,000$, HED-PF performs well compared to the other non-invasive computer vision methodologies even in the presence of large magnification factor. In Fig. 13(a), there is a clear drop in accuracy when looking at the motion estimation results for both PME and traditional edge thresholding. As aforementioned, the deterioration of natural geometric features in an image post magnification affects the accuracy of tracking exaggerated motion. In contrast, the recurrent sampling and smoothed edges following the use of HED aid in gathering an accurate estimate of sequential motion between images irrespective of additive noise. In the following section, the limitation of HED-PF will be tested by altering the manner of excitation as well as extracting operating deflection shapes (ODS) using magnified time series data.

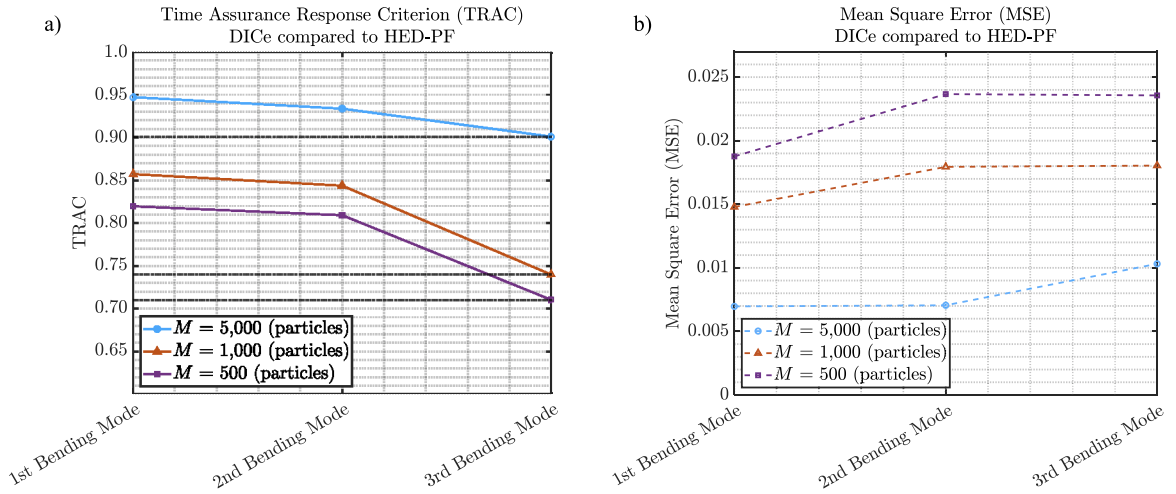


Fig. 11. (a) TRAC and (b) Mean Square Error (MSE) values comparing DICE and each M for the first three magnified bending modes.

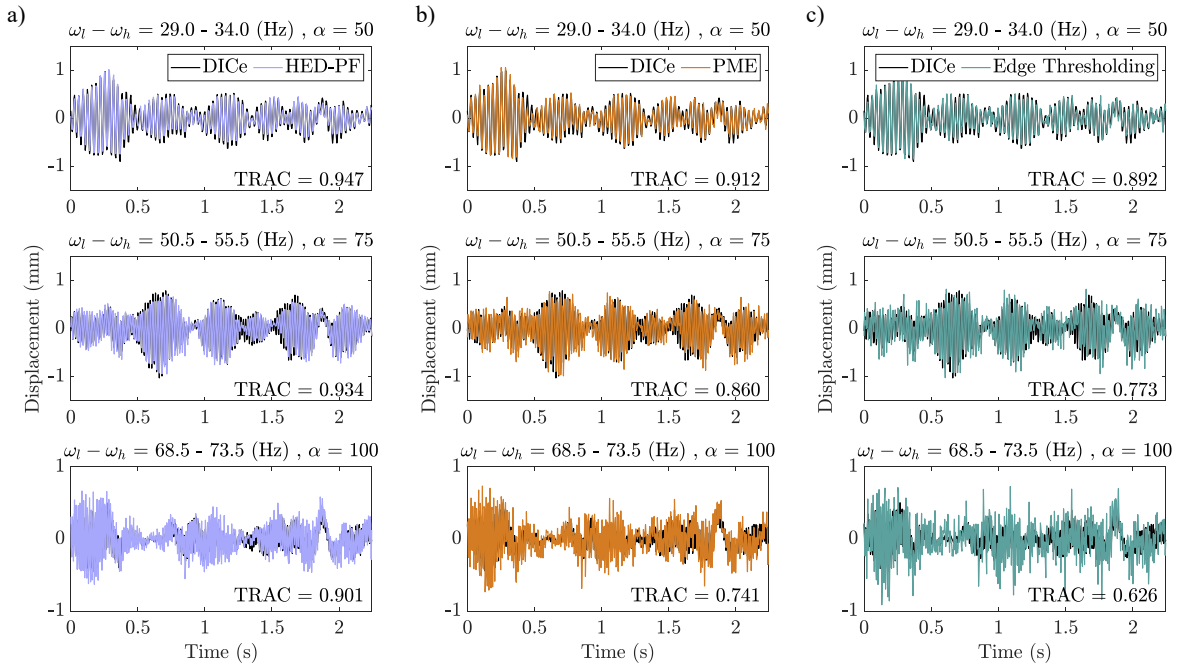


Fig. 12. Magnified time series extraction comparing 2-D point tracking via DICE and (a) HED-PF; (b) PME; (c) Edge Thresholding.

4.2. Cantilever beam with varying excitation and distracting background scenery

Now that HED-PF has been used to extract amplified subtle motion seen in video, further experimentation took place to extract modal characteristics. An experimental modal test was conducted on the cantilever structure to gather information concerning resonant frequencies. Image data was captured using a PHOTRON camera with a sampling rate of 1500 (fps) and a resolution of 2048×448 (pixel). The camera was positioned perpendicular to the test fixture with a working distance of 120 (cm) as shown in Fig. 14.

It should be noted that this test was designed; such that, several non-ideal scenarios were in place to test the robustness of the proposed approach versus the current state-of-the-art. Insufficient lighting and occlusions are common computer vision hurdles that limit the accuracy of object tracking and feature matching [58]. Fig. 15(a) depicts an example of the image data collected by the PHOTRON camera with the distracting background scenery, while Fig. 15(b) shows the frequency response function for the beam fixture.

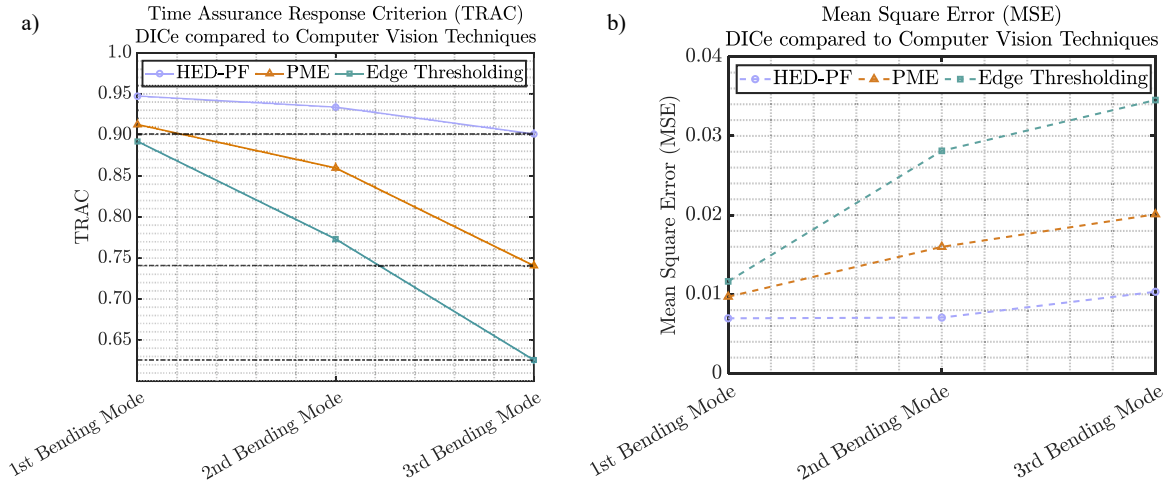


Fig. 13. (a) TRAC and (b) Mean Square Error (MSE) values comparing DICe and each computer vision technique for the first three magnified bending modes.

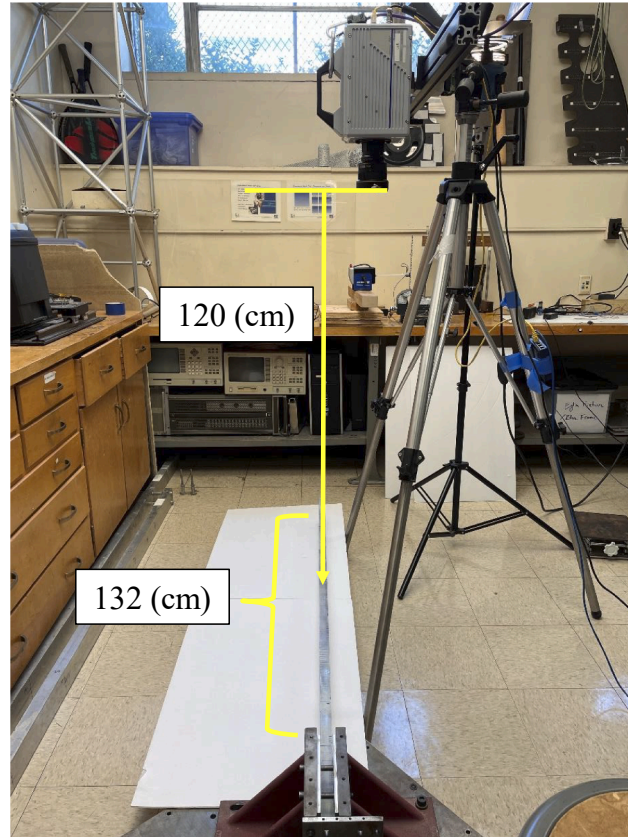


Fig. 14. Experimental setup of the cantilever beam where the orientation of the camera was positioned perpendicular to the test fixture. The working distance and length of the beam were measured at 120 and 132 (cm) respectively.

A finite element model is created for the aluminum beam; such that, analytical mode shapes can be compared to the experimental data that is gathered by HED-PF, PME, and Edge-Thresholding. The parameters and details related to the finite element model are outlined in Table 2. Mode shapes are computed using the extracted eigenvectors and scaled; such that; the minimum and maximum deflection are bounded from $(-1, 1)$. The vectors for the first four bending modes are shown in Fig. 16.

Two varying forms of excitation are used for this dynamic test, due to the magnification necessary to extract the high-frequency

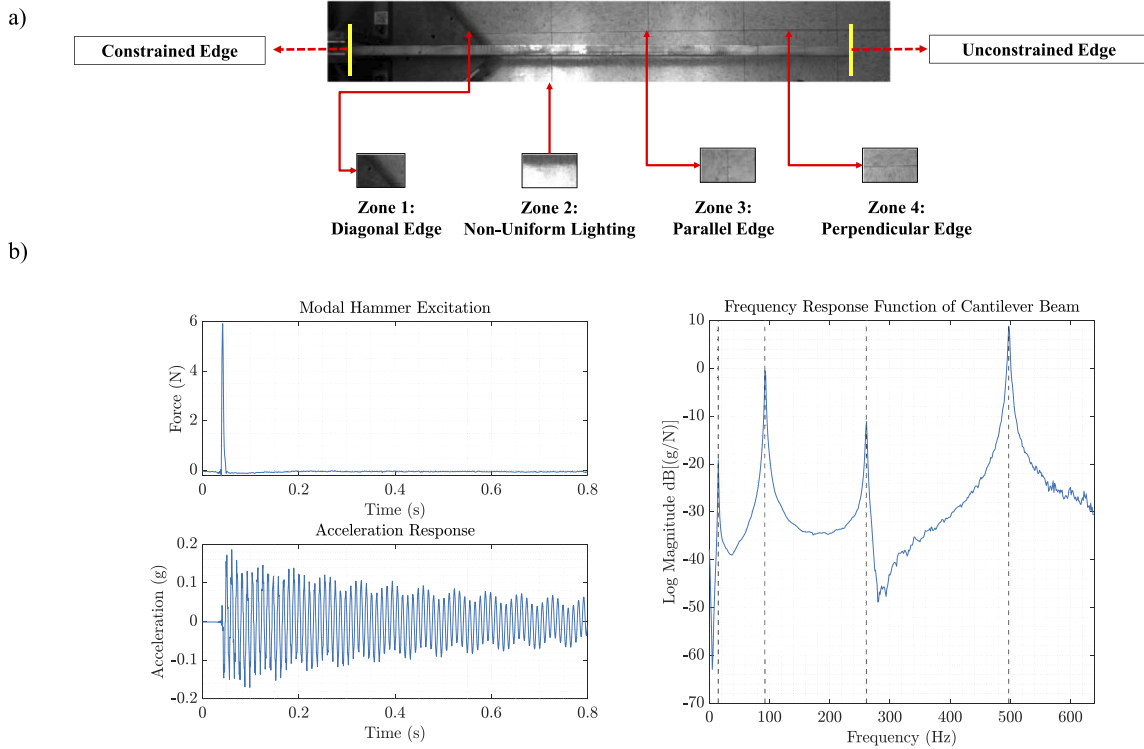


Fig. 15. (a) Cantilever beam with testing occlusions such as insufficient lighting and a non-homogenous background; (b) Modal test conducted on the beam structure where the first four resonant frequencies are identified using a sampling rate of 1280 Hz.

Table 2

Finite element model parameters used for mode shape extraction of cantilevered beam.

Property	Value
Density	2.7 (g/cm ³)
Young's Modulus	69 (GPa)
Poisson's Ratio	0.33
L, Length Dimension of Cantilever Beam	132 (cm)
h, Height Dimension of Cantilever Beam	5.0 (cm)
b, Base Dimension of Cantilever Beam	2.5 (cm)
t _b , Thickness Dimension of Cantilever Beam	0.3 (cm)
Number of Elements	20
Size of Elements (L × W × H)	(6.6 × 1.4 × 5) cm
Type of Elements	2-D beam

content of higher dynamics. Using a modal impact hammer permits modal participation spread over multiple modes, which can make computer vision tracking more difficult. For this test, ten impacts were imparted and averaged to conduct modal analysis on the beam. These impacts took place at a 2 (cm) distance unconstrained edge pictured in Fig. 15. To limit the interference with further analysis, image data was captured after the excitation took place using the high-speed cameras at 1500 (fps). An increase in magnification factor can aid in making imperceivable motion more visible, but the integrity of key features in the image diminish quickly [28]. As an alternative form of excitation, a PCB MiniShaker is driven at each of the first four resonant frequencies that were identified in Fig. 15 (b). Using HED-PF, extraction of operating data will take place on both modal hammer and shaker excitation. This will permit conclusions to be drawn on how HED-PF's ability to extract frequency spectra with unimodal and multimodal excitation. Table 3, outlines the frequencies gathered using the experimental finite element, and computer vision data. The experimental frequencies will be used to excite the first four bending modes using the PCB MiniShaker.

The imaginary component of the Fast Fourier Transform (FFT) is used to extract operating data using the magnified time series. The Modal Assurance Criterion (MAC) is used to quantitatively measure the correlation between mode shapes. The MAC value is similar to TRAC where the value of unity is representative of high correlation between two mode shape vectors. Equation (26) express the MAC where comparisons are drawn between experimental operational deflection shapes $\tilde{\vartheta}_m$ and finite element model ϑ_m mode shapes, where H represents the Hermitian operator.

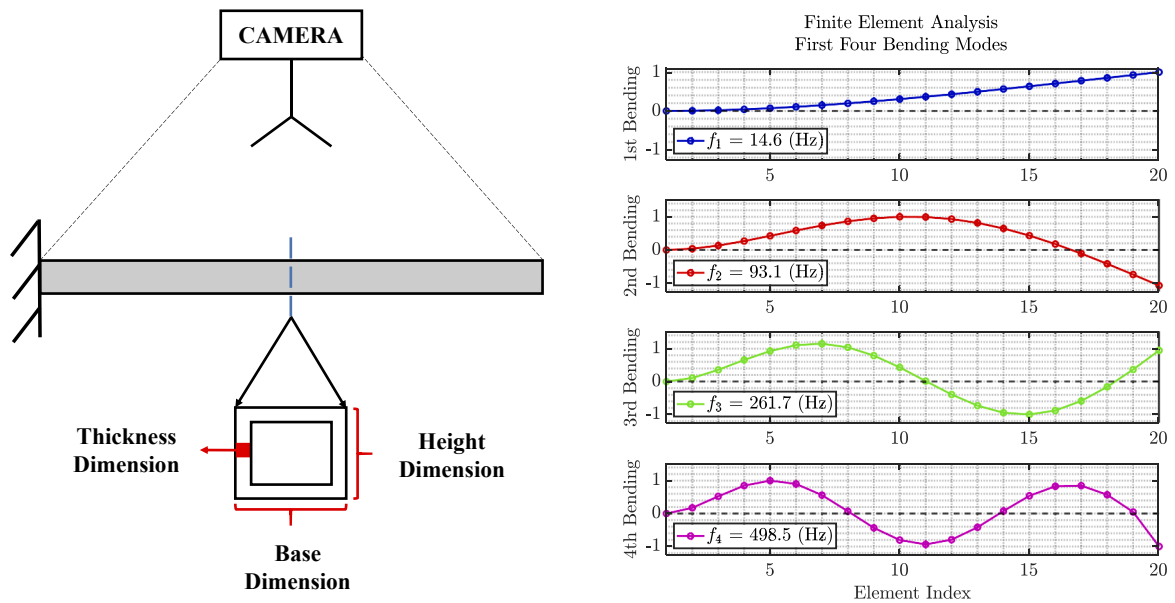


Fig. 16. Display of the experimental test fixture and its appropriate cross section with extracted analytical mode shapes of the cantilever beam structure.

Table 3

Comparison between experimental, finite element model, and computer vision frequencies.

Bending Mode	Accelerometer Frequencies (Hz)	Finite Element Frequencies (Hz)	HED Frequencies (Hz)	PME Frequencies (Hz)	Edge-Thresholding Frequencies (Hz)
1st Bending	15.00	14.95	14.97	14.32	13.90
2nd Bending	92.50	93.10	92.10	92.75	91.54
3rd Bending	261.30	261.70	261.60	259.92	258.33
4th Bending	497.50	498.50	497.23	496.14	493.40

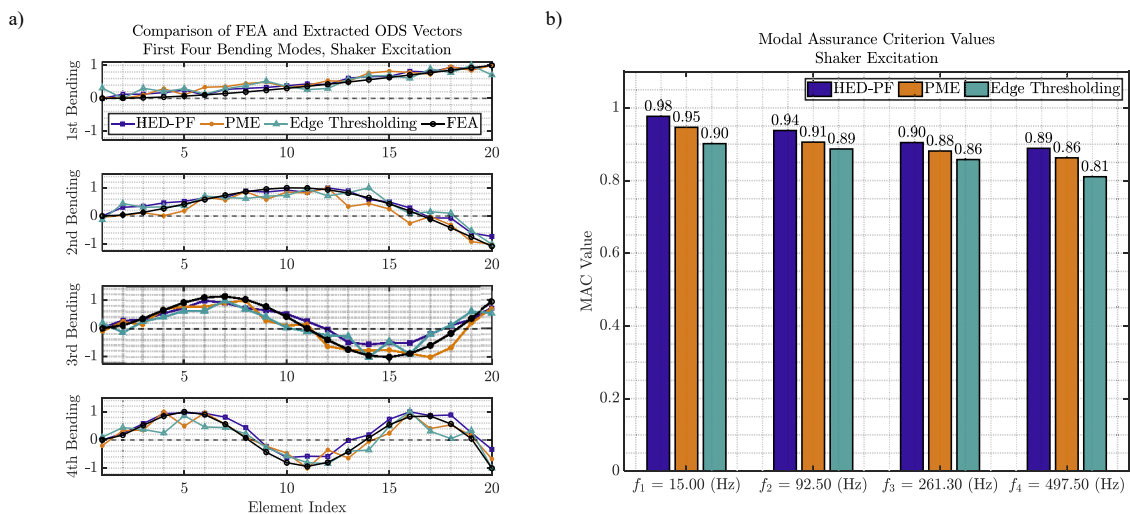


Fig. 17. (a) HED-PF, PME and Edge-Thresholding extraction of ODS using a shaker for excitation referencing the testing parameters outlined in Table 3; (b) Modal Assurance Criterion (MAC) values comparing ODS extraction for each computer vision approach compared to the Finite Element Model for unimodal excitation.

$$MAC = \frac{(\tilde{\vartheta}_m^H \vartheta_m)^2}{(\tilde{\vartheta}_m^H \tilde{\vartheta}_m)(\vartheta_m^H \vartheta_m)}. \quad (26)$$

Fig. 17 and Fig. 18 display the extracted operational deflection shapes using the computer vision techniques with their corresponding MAC values for the first four resonant frequencies. When comparing the results of the shaker excitation test, it can be seen the HED-PF is successfully able to extract operating shapes of the cantilever beam with a high degree of accuracy when compared to the finite element model. The results show that both PME and edge thresholding struggle in the presence of ghosting artifacts from PMM, background occlusions, and insufficient lighting scenarios. Fig. 18 depicts the quantitative results from the modal hammer excitation test. These results display the robust capability of HED-PF to identify subtle changes in motion that spans multiple frequencies. With regards to the shaker test, an excitation at a resonant frequency will be more simplistic to identify using computer vision approaches because the magnification is specified within the bandpass filter's width.

The modal impact hammer, depending on the hammer tip used, can span a large range of frequencies; thus, amplification of resonant frequencies may not be as prominent post magnification. Glancing at the correlation results, HED-PF is successfully able to extract operating data semi-autonomously without having to rely on hand-speckling or lengthy image correlation techniques. It appears that the correlation begins to decline as the mode of vibration increases, which can be attributed to a multitude of reasons. One in particular, is the reliance on PMM to amplify the subtle excitation captured in video. To minimize this effect and achieve higher correlation, a larger form of excitation would be necessary to limit the amount of amplification needed to make motion more apparent. Additional factors that are worth mentioning are the introduction of the various trouble zones that are depicted in Fig. 15. Specifically, the oversaturation that appears in Zone 2 is an example of non-uniform lighting. The proposed approach can handle such occlusions based on the learning of high-frequency spatial features. The alternative computer vision that has been seen in literature is extremely sensitive to changes in pixel intensity. Therefore, MAC values would likely increase for the current-state-of-the-art approaches if this test fixture was subjected to ideal testing conditions.

5. Conclusions

In this work, HED-PF is presented as an alternative way to extract subtle displacement from video to improve identification of operating shapes. To this point, there has been seldom investigations into using two-dimensional particle filtering as a manner for extracting subtle motion. Due to the ghosting artifacts that come as a result of motion amplification, Holistically Nested Edge Detection (HED) is utilized to highlight high-frequency edge features in images without eliminating pixels using adaptive thresholding. The combination of particle filtering and HED eliminates the need for exhaustive correlation techniques and subsequent interpolation between displacement measurements. Correlation techniques, although highly accurate, require computation of vectors at each pixel within a facet.

The implementation of HED-PF permits a more unsupervised approach for extracting dynamic motion. HED-PF is shown to outperform common computer vision techniques such as Phase Motion Estimation (PME) and Edge-Thresholding when extracting amplified motion. HED-PF was able to extract operational deflection shapes which showed large agreement to an analytical finite

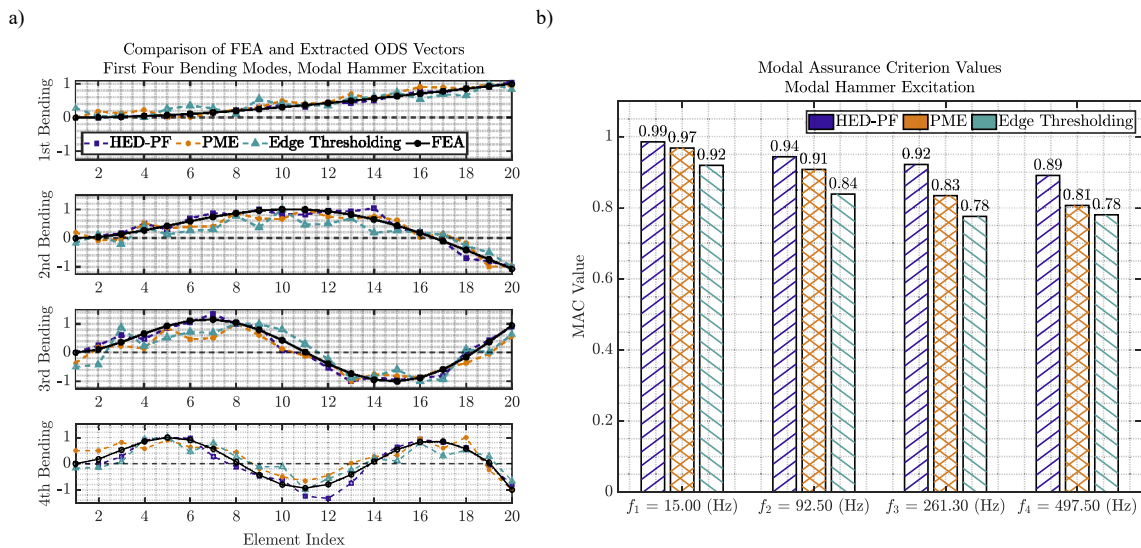


Fig. 18. (a) HED-PF, PME and Edge-Thresholding extraction of ODS using a modal impact hammer for excitation; (b) Modal Assurance Criterion (MAC) values comparing ODS extraction for each computer vision approach compared to the Finite Element Model.

element model, thus displaying its potential for handling more complex engineering structures. Also, investigation into an optimal number of particles to ensure accuracy was displayed. As aforementioned, the incorporation of additional particles increases the accuracy in experimental results, however, this comes at a computational cost. Alternative statistical approaches that include notions concerning the observed physical phenomena have yet to be explored. Conducting said analysis could prove to decrease the number of samples necessary to approximate a change in displacement over a sequence of frames. Future studies will investigate added uncertainties that are observed when using the pre-trained HED machine learning model.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Zhu Mao reports financial support was provided by National Science Foundation.

Data availability

The data that has been used is confidential.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1762809. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- [1] J. Javh, J. Slavić, M. Boltežar, The subpixel resolution of optical-flow-based modal analysis, *Mech. Syst. Sig. Process.* 88 (2017) 89–99, <https://doi.org/10.1016/j.ymssp.2016.11.009>.
- [2] B. Pan, H. Xie, J. Gao, A. Asundi, Improved speckle projection profilometry for out-of-plane shape measurement, *Appl. Opt.* 47 (2008) 5527–5533, <https://doi.org/10.1364/AO.47.005527>.
- [3] M. Angelosanti, S. Debetwar, E. Currá, A. Sabato, 3D-DIC analysis for BIM-oriented SHM of a lab-scale aluminium frame structure, *J. Phys. Conf. Ser.* 2041 (1) (2021/10/01 2021.), 012009, <https://doi.org/10.1088/1742-6596/2041/1/012009>.
- [4] A. Sabato, N.A. Valente, C. Niezrecki, Development of a camera localization system for three-dimensional digital image correlation camera triangulation, *IEEE Sens. J.* 20 (19) (2020) 11518–11526, <https://doi.org/10.1109/JSEN.2020.2997774>.
- [5] Á.J. Molina-Viedma, E. López-Alba, L. Felipe-Sesé, F.A. Díaz, Operational deflection shape extraction from broadband events of an aircraft component using 3D-DIC in magnified images, *Shock Vib.* 2019 (2019) 1–9, <https://doi.org/10.1155/2019/4039862>.
- [6] L. Felipe-Sesé, Á.J. Molina-Viedma, E. López-Alba, F.A. Díaz, FP+DIC for low-cost 3D full-field experimental modal analysis in industrial components, *Mech. Syst. Sig. Process.* 128 (2019) 329–339, <https://doi.org/10.1016/j.ymssp.2019.04.004>.
- [7] D. Reagan, A. Sabato, C. Niezrecki, “Unmanned aerial vehicle acquisition of three-dimensional digital image correlation measurements for structural health monitoring of bridges,” vol. 10169, ed. 2017.
- [8] C. Liu, A. Torralba, W.T. Freeman, F. Durand, E. Adelson, Motion magnification, *ACM Trans. Graph.* 24 (2005) 519–526, <https://doi.org/10.1145/1186822.1073223>.
- [9] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, W. Freeman, Eulerian video magnification for revealing subtle changes in the world, *ACM Trans. Graph.* 31 (4) (2012) pp. <https://doi.org/10.1145/2185520.2185561>.
- [10] N. Wadhwa, M. Rubinstein, F. Durand, W.T. Freeman, Phase-based video motion processing, *ACM Transactions on Graphics (TOG)* 32 (4) (2013) 1–10.
- [11] N. Wadhwa, H.-Y. Wu, A. Davis, M. Rubinstein, E. Shih, G.J. Mysore, J.G. Chen, O. Buyukozturk, J.V. Guttag, W.T. Freeman, F. Durand, Eulerian video magnification and analysis, *Commun. ACM* 60 (1) (2016) 87–95.
- [12] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, “Riesz Pyramids for Fast Phase-Based Video Magnification,” *Computational Photography (ICCP)*, 2014 IEEE International Conference on, 2014, doi: 10.1109/ICCPHOT.2014.6831820.
- [13] Q. Qiu, D. Lau, Defect detection in FRP-bonded structural system via phase-based motion magnification technique, *Struct. Control Health Monit.* 25 (2018) e2259–e, <https://doi.org/10.1002/stc.2259>.
- [14] A.J. Molina-Viedma, L. Felipe-Sesé, E. López-Alba, F.A. Díaz, High frequency mode shapes characterisation using Digital Image Correlation and phase-based motion magnification, *Mech. Syst. Sig. Process.* 102 (2018) 245–261, <https://doi.org/10.1016/j.ymssp.2017.09.019>.
- [15] A.J. Molina-Viedma, L. Felipe-Sesé, E. López-Alba, F.A. Díaz, 3D mode shapes characterisation using phase-based motion magnification in large structures using stereoscopic DIC, *Mech. Syst. Sig. Process.* 108 (2018) 140–155, <https://doi.org/10.1016/j.ymssp.2018.02.006>.
- [16] 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.
- [17] D.M. Siringoringo, S. Wangchuk, Y. Fujino, Noncontact operational modal analysis of light poles by vision-based motion-magnification method, *Eng. Struct.* 244 (2021) 112728, <https://doi.org/10.1016/j.engstruct.2021.112728>.
- [18] 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>.
- [19] P. Poozesh, A. Sarrafi, Z. Mao, P. Avitabile, C. Niezrecki, Feasibility of extracting operating shapes using phase-based motion magnification technique and stereo-photogrammetry, *J. Sound Vib.* 407 (2017) 350–366.
- [20] C.T. do Cabo, N. A. Valente, Z. Mao, “Motion magnification for optical-based structural health monitoring,” in *Health Monitoring of Structural and Biological Systems XIV*, P. Fromme and Z. Su, Eds., 2020, vol. 11381: SPIE, pp. 221 – 227-221 – 227, doi: 10.1117/12.2559266.
- [21] 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>.
- [22] 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. Sig. Process.* 85 (2017) 567–590.
- [23] M. Ghandil, Ö. Dabanli, H.T. Riahi, An enhanced indirect video-based measurement procedure for dynamic structural system identification applications, *Measurement* 182 (2021) 109759, <https://doi.org/10.1016/j.measurement.2021.109759>.
- [24] Q. Li, G. Wang, A. Sarrafi, X. Niu, W. Lu, and Z. Mao, “Dynamic Characteristics Identification of an Arch Dam Model via the Phase-Based Video Processing,” *KSCSE J. Civil Eng.*, vol. 25, no. 1, pp. 140–152-140–152, 2021, doi: 10.1007/s12205-020-0400-z.

- [25] J. Brievea, E. Moya-Albor, "Phase-based motion magnification video for monitoring of vital signals using the Hermite transform," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 2017/november, vol. 10572, pp. 105720M-105720M, doi: 10.1117/12.2285959.
- [26] J. Won, P.-C. Huang, S.A. Boppart, Phase-based Eulerian motion magnification reveals eardrum mobility from pneumatic otoscopy without sealing the ear canal, *J. Phys.: Photonics* 2 (3) (2020) 034004, <https://doi.org/10.1088/2515-7647/ab8a59>.
- [27] E. Moya-Albor, J. Brievea, H. Ponce, O. Rivas-Scott, C. Gómez-Peña, "Heart Rate Estimation using Hermite Transform Video Magnification and Deep Learning," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2595-2598, 2018, doi: 10.1109/EMBC.2018.8512879.
- [28] N.A. Valente, C.T. do Cabo, Z. Mao, C. Niezrecki, Quantification of phase-based magnified motion using image enhancement and optical flow techniques, *Measurement* 189 (2022) 110508.
- [29] M. Verma, S. Raman, "Edge-Aware Spatial Filtering-Based Motion Magnification," in Proceedings of 2nd International Conference on Computer Vision & Image Processing, B. B. Chaudhuri, M. S. Kankanalli, and B. Raman, Eds., 2018: Springer Singapore, pp. 117–128-117–128, doi: 10.1007/978-981-10-7898-9_10.
- [30] X. Wu, X. Yang, J. Jin, Z. Yang, Amplitude-based filtering for video magnification in presence of large motion, *Sensors* 18 (2018) 2312, <https://doi.org/10.3390/s18072312>.
- [31] N.A. Valente, Z. Mao, M. Southwick, C. Niezrecki, "Implementation of Total Variation Applied to Motion Magnification for Structural Dynamic Identification," in Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6, D. Di Maio and J. Baqersad, Eds., 2020: Springer International Publishing, pp. 139–144-139–144, doi: 10.1007/978-3-030-47721-9_17.
- [32] M. Verma, R. Ghosh, S. Raman, "Saliency Driven Video Motion Magnification," in Computer Vision, Pattern Recognition, Image Processing, and Graphics, R. Rameshan, C. Arora, and S. Dutta Roy, Eds., 2018: Springer Singapore, pp. 89–100-89–100, doi: 10.1007/978-981-13-0020-2_9.
- [33] M. Sushma, A. Gupta, J. Sivaswamy, "Semi-automated Magnification of Small Motions in Videos," in Pattern Recognition and Machine Intelligence, P. Maji, A. Ghosh, M. N. Murty, K. Ghosh, and S. K. Pal, Eds., 2013: Springer Berlin Heidelberg, pp. 417–422-417–422, doi: 10.1007/978-3-642-45062-4_57.
- [34] C. Shi, G. Luo, A streaming motion magnification core for smart image sensors, *IEEE Trans. Circuits Syst. Express Briefs* PP (2017) 1, <https://doi.org/10.1109/TCSIL.2017.2775583>.
- [35] T.-H. Oh et al., "Learning-Based Video Motion Magnification," in Computer Vision – ECCV 2018, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., 2018: Springer International Publishing, pp. 663–679-663–679, doi: 10.1007/978-3-030-01225-0_39.
- [36] J. N. Sarvaiya, S. Patnaik, S. Bombaywala, "Image Registration by Template Matching Using Normalized Cross-Correlation," in 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies, 2009, pp. 819-822, doi: 10.1109/ACT.2009.207.
- [37] D. Gorjup, J. Slavič, A. Babnik, M. Boltežar, Still-camera multiview Spectral Optical Flow Imaging for 3D operating-deflection-shape identification, *Mech. Syst. Sig. Process.* 152 (2021), 107456, <https://doi.org/10.1016/j.ymssp.2020.107456>.
- [38] 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. Sig. Process.* 164 (2022) 108232, <https://doi.org/10.1016/j.ymssp.2021.108232>.
- [39] N.A. Valente, C.T. do Cabo, Z. Mao, C. Niezrecki, "Template Matching and Particle Filtering for Structural Identification of High- and Low-Frequency Vibration," in Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6, Cham, D. Di Maio and J. Baqersad, Eds., 2023// 2023: Springer International Publishing, pp. 43-50, doi: 10.1007/978-3-031-04098-6_5.
- [40] S. Bhowmick, S. Nagarajaiah, Identification of full-field dynamic modes using continuous displacement response estimated from vibrating edge video, *J. Sound Vib.* 489 (2020) 115657, <https://doi.org/10.1016/j.jsv.2020.115657>.
- [41] S. Xie, Z. Tu, "Holistically-Nested Edge Detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2015/december.
- [42] J. Czyz, B. Ristic, B. Macq, A particle filter for joint detection and tracking of color objects, *Image Vis. Comput.* 25 (8) (2007) 1271–1281, <https://doi.org/10.1016/j.imavis.2006.07.027>.
- [43] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, L. Van Gool, Robust tracking-by-detection using a detector confidence particle filter, in: 2009 IEEE 12th International Conference on Computer Vision, 2009, pp. 1515–1522, <https://doi.org/10.1109/ICCV.2009.5459278>.
- [44] K. Nummiaro, E. Koller-Meier, L. Van Gool, An adaptive color-based particle filter, *Image Vis. Comput.* 21 (1) (2003) 99–110, [https://doi.org/10.1016/S0262-8856\(02\)00129-4](https://doi.org/10.1016/S0262-8856(02)00129-4).
- [45] A. Sarrafi, Z. Mao, "Mapping Motion-Magnified Videos to Operating Deflection Shape Vectors Using Particle Filters," in Rotating Machinery, Optical Methods & Scanning LDV Methods, Volume 6, C. Niezrecki, J. Baqersad, and D. Di Maio, Eds., 2019: Springer International Publishing, pp. 75–83-75–83, doi: 10.1007/978-3-030-12935-4_8.
- [46] N.A. Valente, A. Sarrafi, Z. Mao, C. Niezrecki, Streamlined particle filtering of phase-based magnified videos for quantified operational deflection shapes, *Mech. Syst. Sig. Process.* 177 (2022), 109233, <https://doi.org/10.1016/j.ymssp.2022.109233>.
- [47] I. Fogel, D. Sagi, Gabor filters as texture discriminator, *Biol. Cybern.* 61 (2) (1989) 103–113.
- [48] W.T. Freeman, E.H. Adelson, The design and use of steerable filters, *IEEE Trans. Pattern Anal. Mach. Intell.* 13 (9) (1991) 891–906.
- [49] A.C. Le Ngo, R.C.W. Phan, "Seeing the invisible: Survey of video motion magnification and small motion analysis," *ACM Computing Surveys (CSUR)*, vol. 52, no. 6, pp. 1–20-1–20, 2019, doi: 10.1145/3355389.
- [50] W. Fan et al., "A new Eulerian temporal filtering method for boosting endoscopic video motion magnification," in Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling, C. A. Linte and J. H. Siewerdsen, Eds., 2021, vol. 11598: SPIE, pp. 460 – 467-460 – 467, doi: <https://doi.org/10.1117/12.2581934>. [Online]. Available: <https://doi.org/10.1117/12.2581934>.
- [51] O. Shabi, S. Natan, A. Kolel, A. Mukherjee, O. Tchaicheeyan, H. Wolfenson, N. Kiryati, A. Lesman, J. Bernardino de la Serna, Motion magnification analysis of microscopy videos of biological cells, *PLoS One* 15 (11) (2020) e0240127.
- [52] D.J. Fleet, A.D. Jepson, Stability of phase information, *IEEE Trans. Pattern Anal. Mach. Intell.* 15 (12) (1993) 1253–1268, <https://doi.org/10.1109/34.250844>.
- [53] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little, and D. G. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking," in Computer Vision - ECCV 2004, T. Pajdla and J. Matas, Eds., 2004: Springer Berlin Heidelberg, pp. 28–39-28–39, doi: 10.1007/978-3-540-24670-1_3.
- [54] M. Marron et al., "Comparing a Kalman Filter and a Particle Filter in a Multiple Objects Tracking Application," in 2007 IEEE International Symposium on Intelligent Signal Processing, 3-5 Oct. 2007 2007, pp. 1-6, doi: 10.1109/WISP.2007.4447520.
- [55] F. Cheli, P. Mazzoleni, M. Pezzola, E. Ruspini, E. Zappa, Vision-based measuring system for rider's pose estimation during motorcycle riding, *Mech. Syst. Sig. Process.* 38 (2) (2013) 399–410, <https://doi.org/10.1016/j.ymssp.2013.01.009>.
- [56] J. Elfring, E. Torta, R. van de Molengraft, "Particle Filters: A Hands-On Tutorial," *Sensors*, vol. 21, no. 2, 2021, doi: 10.3390/s21020438.
- [57] D. Turner, P. Crozier, P. Reu, "Digital Image Correlation Engine, Version 00," ed, 2015.
- [58] S. Ricco, C. Tomasi, "Dense Lagrangian motion estimation with occlusions," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 16-21 June 2012 2012, pp. 1800-1807, doi: 10.1109/CVPR.2012.6247877.