



Full-length article

Methods to estimate subpixel level small motion from video of vibrating cutting tools



Anshid Nuhman P., Aditya Singh, Rohit Lambora, Mohit Law*

Machine Tool Dynamics Laboratory, Department of Mechanical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India

ARTICLE INFO

Available online xxxx

Keywords:

Subpixel motion
High speed camera
Extension tube
Optical flow
Vision-based modal analysis

ABSTRACT

Cutting tools can vibrate with frequencies of up to a few kHz with amplitudes ranging from a few μm to hundreds of μm . To characterize these vibrations using vision-based modal analysis methods, video of vibrating tools must be acquired at high speed and resolution. High speed acquisition will ensure modes are temporally resolved without aliasing, whereas high resolution acquisition ensures that the pixel size is small enough to resolve small motion. However, since most digital cameras trade speed for resolution, estimating small motion of high frequency modes becomes difficult. To address this issue, this paper presents new methods to estimate small motion from video of vibrating cutting tools. Two methods are illustrated on a video of a slender end mill. One is hardware-based, and the other is software-based. In the hardware-based method, we use combinations of extension tubes and a reverse ring fitted with a standard lens to achieve a pixel size of as less as $1.3 \mu\text{m}$. In the software-based method, we leverage the capability of intensity- and phase-based motion registration methods to detect small subpixel level motion, even when the pixel size is $83 \mu\text{m}$. In both methods we use the object's own features to detect and register motion, thus overcoming the limitation of digital image correlation schemes that need markers for target tracking. Modal parameters evaluated from motion estimated with both methods are found to agree with those extracted from twice integrated accelerations. Since methods proposed herein proffer new ways to register small oscillatory motion, we believe that our findings will further help leverage the adoption of vision-based modal analysis methods in machine tool systems as an alternative to the more traditional experimental modal analysis procedures.

© 2022 CIRP.

Introduction

Vision-based modal analysis methods offer several advantages [1–3]. Since the method is non-contact, it avoids pitfalls associated with potential mass-loading due to measurement transducers used in more traditional experimental modal analysis (EMA) procedures. Moreover, since every pixel is a virtual sensor, full-field mode shape analysis is possible with fewer video recordings than having to rove the sensor or the actuator over a measurement grid as is required in traditional EMA procedures. Furthermore, measurements need just a digital camera, and analysis requires a computer running image processing algorithms. Sophisticated and expensive multi-channel data acquisition systems together with modal measurement and analysis software packages are hence not required. These advantages make vision-based modal analysis a promising alternative to the

more traditional EMA procedures for modal analysis of machine tool systems [4–7]. However, their true potential can only be realized when motion to be estimated from video is properly resolved in space and in time.

Characterizing cutting tool vibrations using vision-based methods requires videos to be acquired at high speeds and resolutions. This is necessary since cutting tools can vibrate with frequencies of up to a few kHz with amplitudes ranging from a few μm to hundreds of μm . High speed acquisition will ensure modes are temporally resolved without aliasing, whereas high resolution acquisition will ensure that the pixel size is small enough to resolve small motion. However, since most digital cameras trade speed for resolution, high frequency small amplitude tool response may either be temporally aliased, spatially aliased, or both. This makes it difficult to estimate small motion of high frequency modes.

Proper temporal resolution requires respecting the Nyquist rate. This is possible with use of high-speed cameras [4,5,7]. And, for measurements using inexpensive digital cameras that may or not be able to acquire images at rates to avoid aliasing, true modal

* Corresponding author.
E-mail address: mLaw@iitk.ac.in (M. Law).

parameters may still be recovered from potentially aliased signals using measurements that are fractionally uncorrelated [8,9]. However, whether temporally aliased or not, if modes are dynamically stiff, motion might be small and of the subpixel level, making motion spatially ill-resolved and hence not observable.

It has been suggested that like there is a Nyquist rate at which data must be sampled to avoid aliasing, likewise there is also a Nyquist criteria to avoid spatial aliasing which requires that motion should span at least two pixels within a frame for it to be resolved [10]. However, even with resolutions of 1080×1080 pixels, the per-pixel resolution can be of the order of $\sim 37.5 \mu\text{m}$, as was shown in prior work [4–6,8,9]. As such, respecting the Nyquist spatial criterion would require motion of the order of $75 + \mu\text{m}$ for it to be properly resolved. And, though some cutting tool modes may indeed have motion of the order of hundreds of μm , dynamically stiffer modes usually vibrate with amplitudes of only a few μm , and motion for such modes might remain undetected.

To address the issue of and need for estimating subpixel level small motion from video recordings of vibrating cutting tools, this paper proposes two solutions. One is hardware-based, and the other is software-based. Registering motion using these two methods is the main new technical contribution of this paper.

In the hardware-based method, we show how combinations of extension tubes and a reverse ring fitted with a standard lens can be used to achieve a pixel size of as less as $1.3 \mu\text{m}$. The size of the pixel without these tubes and/or rings is $83 \mu\text{m}$. The extension tubes hence change the imaging-system resolution from being camera/detector limited to lens limited. This multi-fold increase in the magnification beyond the resolution of the original lens is like an 'empty magnification'. Since these tubes contain no optics there is no distortion of the image and/or added detail or contrast in the image. The tube simply acts as a spacer between the lens and the camera body. By moving the lens away from the body, the extension tube adjusts the focal point and allows focusing on subjects closer to the camera. More than one extension tube can be stacked together as necessary. A macro lens serves the same purpose as that of extension tube, albeit at higher costs. Due to their magnification potential, extension tubes and/or macro lenses have been used in many shape, photogrammetry, displacement, and strain measurement applications that needed to resolve small features [11–13] and displacements [10,14,15].

Though extension tubes and/or macro lenses have been used in other applications, there are no reports of using them to measure small motion of vibrating objects. As such, our experimental demonstrations through this paper are the first such reports showing how this method can be leveraged to detect small oscillatory motion. Moreover, prior work that used tubes to estimate small displacements did so using digital image correlation subpixel motion registration schemes that need markers and/or speckle patterns for target tracking and motion estimation [10,14,15]. Placing markers and/or making speckles on tools is not feasible since it may damage the coating and render the tool unusable. As such, we use the tool's own features to detect and register motion using edge detection and tracking schemes [5]. Even though the edge detection and tracking scheme is not strictly a subpixel level motion registration method, since the method averages the spatial location of the edge over several pixels within the region of interest, it sometimes can register motion smaller than the pixel size [5].

In addition to being the first to report on the use of extension tubes to estimate small oscillatory motion, our use of a reverse ring as an alternate method to magnify the image to detect small oscillatory motion is also the first such report. In this method the lens is simply connected backwards with the aid of a lens-specific mount. Since the lens and its optics are reversed in this setup, the image gets magnified. Magnification can be further enhanced by placing one or more extension tubes between the

camera body and the reverse mounted lens, as will be shown in this paper.

Though there is a trade-off in the loss of depth of field with the use of extension tubes and reverse rings, since we are in any case only interested in the motion of the cutting edge(s), the shallow depth of field is not of major concern in our application of interest. Moreover, since use of extension tube(s) and/or reverse rings changes the effective focal length and the aperture, and since less light falls on the image sensor, adjusting the intensity of the light becomes necessary for every measurement. And, since the focal length, aperture, and illumination are all correlated, optimally tuning these parameters can sometimes become difficult. As such, we also discuss software-based methods to detect subpixel level small oscillatory motion.

In the software-based methods, we leverage the capability of optical flow motion registration methods [16–18] to detect subpixel level small motion for when the video is recorded only with a standard lens mounted on the camera. In this case too we use the tool's own features to register motion. Within the family of optical flow methods, we prefer to use the intensity-based method as well as the phase-based method. Intensity-based optical flow approaches are based on the optical flow equation that can provide velocity in the direction of the intensity gradient. Displacements are estimated by monitoring pixel intensities between pairs of images by assuming that brightness is conserved [5,16,17]. Since intensity-based optical flow methods are susceptible to noise and a change in illumination, we also explore the use of phase-based methods that determine the displacements by tracking the constant phase contours of the images. Phase contours are relatively insensitive to photometric deformations that may result from changes in lighting conditions and as such are more robust than intensities [18,19]. Our use of the phase-based motion registration scheme to detect small cutting tool motion is also the first such report in the context of machine tool systems.

We demonstrate methods to estimate subpixel level small motion from video for a representative end mill mounted in a 3-axis CNC machine. The setup to do so along with the setups with the extension tubes and the reverse ring are described in the next section. Following which we describe methods to estimate motion using software-based methods. Estimates of motion using the hardware and software-based methods are contrasted thereafter and benchmarked with twice integrated accelerations. That section includes a discussion on the modal parameters estimated from registered motion. We conclude the paper by discussing how our findings can further facilitate adoption of vision-based modal analysis methods in machine tool systems.

Detecting small motion using extension tubes and a reverse ring

The experimental setup to measure vibrations is shown in Fig. 1(a). We measured the response of a 16 mm end mill with four flutes mounted in a power chuck type tool holder with a HSK63 interface that was held in the spindle of a 3-axis vertical CNC milling machine. The overhang of the tool from the spindle nose was kept at 36 mm. We excited the tool with a modal hammer (make: Dytran; model number: 5800B4) and measured the impulse-like response using a Chronos 2.1 monochrome high-speed camera that records video at 2142 frames per second. The tool has a dominant mode at $\sim 619 \text{ Hz}$ as confirmed by measurements with an accelerometer (make: Dytran; model number: 3035B1G) mounted at the tool tip. As such, the frame rate for recording video respects the Nyquist criterion and ensures that the response is not temporally aliased. The resolution at this frame rate was 720×1280 pixels. With a standard 18–55 mm focal length lens mounted on the camera body – shown in Fig. 1(a), the per-pixel resolution was $83 \mu\text{m}$. To achieve smaller per-pixel resolutions by magnifying the image, we experimented

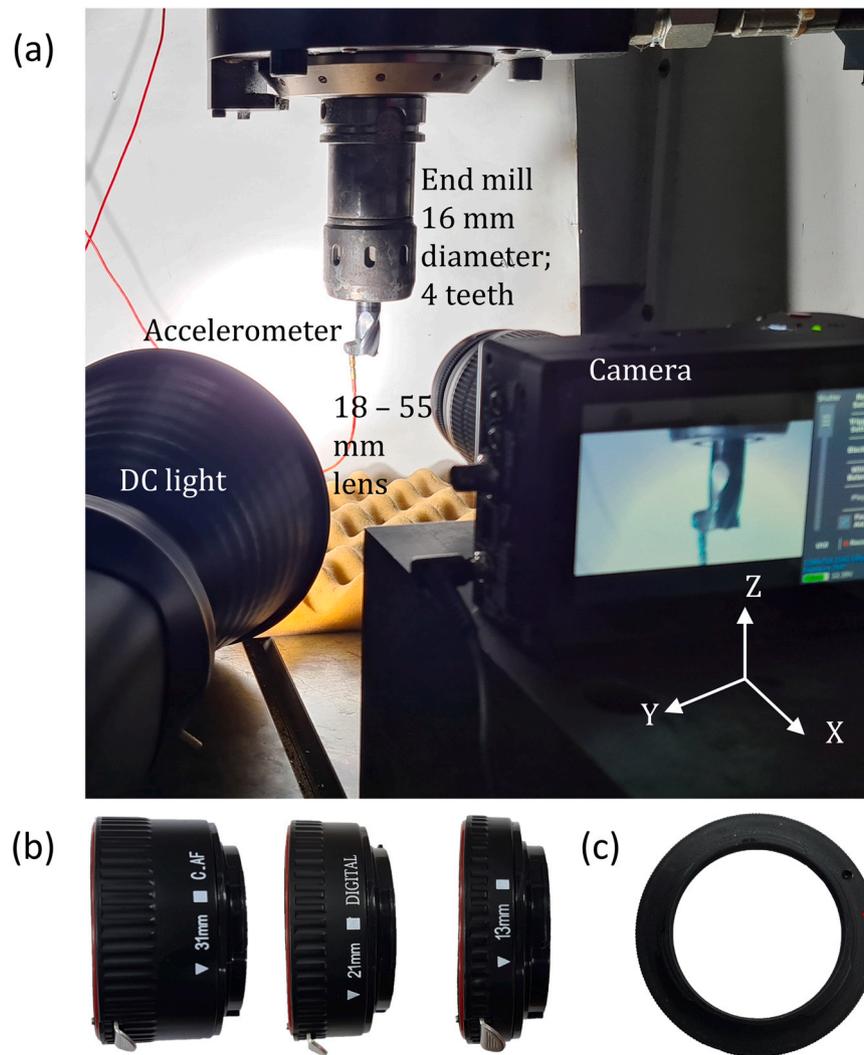


Fig. 1. (a) Experimental setup showing the camera with a standard 18–55 mm lens, the tool, and the DC light, (b) three extension tubes and (c) the reverse ring.

with seven additional and different combinations of extension tubes and a reverse ring. We experimented with three different extension tubes of length 13 mm, 21 mm, and 31 mm. These tubes are shown in Fig. 1(b). We also experimented with the lens mounted backwards on the camera with a reverse ring that is shown in Fig. 1(c). With the reverse ring, we conducted three additional experiments with the three different extension tubes.

For all experiments, the aperture was kept wide open, the shutter speed was adjusted to be 460 μ s, and the ISO setting was 2000. The acquisition rate was kept fixed at 2142 frames per second, and the resolution was 720×1280 pixels. Since less light falls on the detector when using tubes and/or the reverse ring, a 100 W DC light (make: Godox; model number: SL100D) with adjustable intensity settings was used to illuminate the tool. All measurements were conducted with a white background to minimize the influence of background noise and blur. Each experiment was conducted at least twice, and measurements were observed to be repeatable.

The single camera setup shown in Fig. 1(a) can only record in-plane motion in the direction perpendicular to the edge of the tool. As such, measurements are illustrated herein with the excitation and response being in the Y-direction. Out of plane motion for in plane excitation is expected to be small and is ignored.

A summary of the eight experiments is shown in Fig. 2. The figure shows a schematic of each experimental setup. The lens and the size of the extension tube is listed in Fig. 2. Since the focal point changes

with every combination of lens and extension tube, the distance between the lens and the tool is adjusted accordingly, and that distance too is listed in Fig. 2. We keep the distance between the DC light and the tool to be fixed and adjust only the light's intensity to obtain better contrast in the image. That intensity setting in terms of the percentage of the full intensity of 100 W is also listed in the figure. Also shown in Fig. 2 is one representative frame from every recorded video to show the field of view with the region of interest highlighted. The regions of interest are the edges that are tracked in every frame to extract displacements using the edge detection and tracking algorithm [4,5]. Since the field of view is known for every image, as are the tool's dimensions, the per-pixel resolution is easily estimated and those too are listed in Fig. 2 for each of the different setups.

As is evident from Fig. 2, the use of extension tubes and/or the reverse ring with and without the extension tubes allowed us to move the camera system closer to the tool than with just using a lens. This in turn increased the magnification. The field of views also make it evident that with the introduction of additional tubes between the lens and the camera body, there is a greater magnification of the image. Since the camera system with tubes and/or the reverse ring is closer to the tool, less light falls on the detector, and as such, the intensity of the light is increased to obtain a sharper image with better contrast for these cases. Yet, even with the increased light intensity, the images with tubes appear darker than those without.

Configuration	Schematic of the setup	Tool	Distance between lens and tool in mm	Intensity of DC light in %	Field of view	Size of the pixel in μm
Camera with 18 - 55 mm lens			115	19	720 x 1280 px 	83
Camera with 18 - 55 mm lens and a 13 mm extension tube			55	25	720 x 1280 px 	16.4
Camera with 18 - 55 mm lens and a 13 mm and a 21 mm extension tube			25	28	720 x 1280 px 	8.8
Camera with 18 - 55 mm lens and 13 mm, 21 mm, and 31 mm extension tubes			13	34	720 x 1280 px 	5.3
Camera with 18 - 55 mm lens fitted backwards with a reverse ring			35	49	720 x 1280 px 	2.7
Camera with 18 - 55 mm lens fitted backwards with a reverse ring and a 13 mm extension tube			35	60	720 x 1280 px 	2
Camera with 18 - 55 mm lens fitted backwards with a reverse ring and a 13 mm and a 21 mm extension tube			35	60	720 x 1280 px 	1.6
Camera with 18 - 55 mm lens fitted backwards with a reverse ring and 3 mm, 21 mm, and 31 mm extension tubes			35	60	720 x 1280 px 	1.3

Fig. 2. Summary of the eight different experimental configurations with the lens connected to the camera body with and without extension tubes and a reverse ring.

Also evident is that the images with tubes are less defined than those without. This is thought to be due to the low-pass spatial filtering effect with tubes that tends to filter out sharper features in the image [10].

Introducing one 13 mm extension tube between the 18–55 mm lens and the camera body improves the per-pixel resolution from 83 μm to 16.4 μm . Introducing another 21 mm tube to this combination further improves the per-pixel resolution to 8.8 μm . And, with

the 13 mm, the 21 mm, and the 31 mm tubes all stacked together and in between the 18–55 mm lens and the camera body, the per-pixel resolution was found to improve to 5.3 μm . This improvement is not insignificant. However, experiments with the 18–55 mm lens mounted backwards with a reverse ring further improve the per-pixel resolution to 2.7 μm , suggesting that reversing the lens might be better than adding three tubes to a lens mounted in the usual forward fashion. Introducing the 13 mm tube with the reverse ring

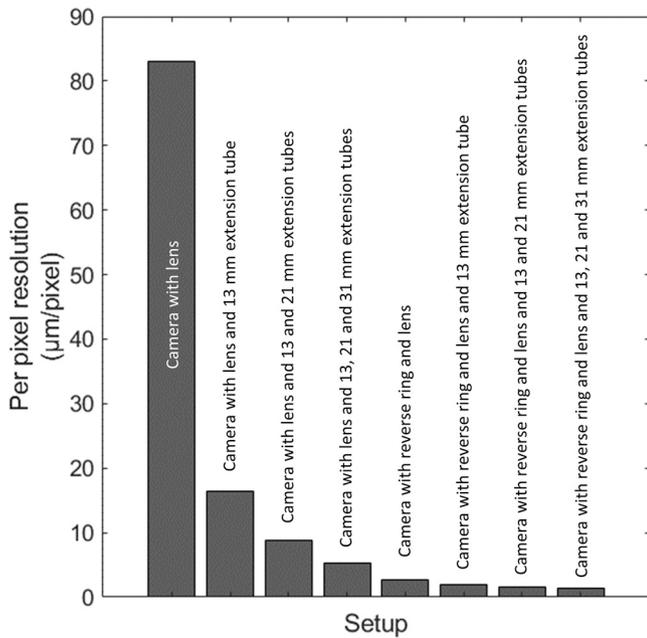


Fig. 3. Per-pixel resolutions changing with different combinations of lens mountings and extension tubes.

setup makes the per-pixel resolution 2 µm. Adding the 21 mm tube to this configuration makes the per-pixel resolution 1.6 µm. And all three tubes stacked and sandwiched between the camera body and the lens mounted backwards results in a per-pixel resolution to 1.3 µm. Changing size of pixels for different combinations of lens mountings and tubes is summarized in Fig. 3.

This reduction in the per-pixel resolution from 83 µm for the 18–55 mm lens to 1.3 µm using a combination of extension tubes and the lens mounted backwards is significant. It is expected that small tool motion will be easily resolved with this and other such small pixel sizes – as is discussed in the section after next that compares motion extracted using vision-based methods with those obtained by twice integrating accelerations.

Intensity and phase-based optical flow methods to detect subpixel level motion

This section outlines procedures to estimate subpixel level small motion from recorded video. For this we use only the video acquired using the 18–55 mm lens connected to the camera body and with which the per-pixel resolution was found to be 83 µm. To estimate motion that is potentially smaller than this pixel size, at first, we outline the intensity-based method, implementation of which follows reports in [5,16,17]. Those discussions are followed by outlining the phase-based method, implementation of which follows report in [18,19]. For extracting displacements in both methods, we decompose the video into its individual grayscale frames and then apply the scheme of interest. And, though we are only interested in motion along the direction of excitation, i.e., the Y-direction, the formulations presented below are kept generalized for the projection coordinates being x and y .

Intensity-based optical flow

An overview of the procedure to extract tool displacements from the intensity-based optical flow-based motion tracking scheme is summarized in Fig. 4. This implementation is based on our prior work [5], which had detailed the main steps. As such, we provide only a brief overview herein, and direct the reader to [5] for details.

In the first step, we crop all images to retain regions at or near the cutting edge(s) of the end mill. We then convolve every frame by a Gaussian kernel that has a standard deviation of one to filter the influence of any potential noise in the image. The second main step involves estimating displacements by monitoring pixel intensities between pairs of images using the gradient-based optical flow-based scheme.

The intensity-based optical flow method assumes brightness constancy, such that the light intensity (I) of a point on the tool in the image (x_i, y_j) that moved by a small amount to another location ($x_i + \Delta x, y_j + \Delta y$) in a short sample time Δt is constant:

$$I(x_i, y_j, t) = I(x_i + \Delta x, y_j + \Delta y, t + \Delta t). \tag{1}$$

Assuming small motion, the image intensity function can be approximated by a first order Taylor expansion as:

$$I(x_i + \Delta x, y_j + \Delta y, t + \Delta t) = I(x_i, y_j, t) + \frac{\partial I}{\partial x} \Delta x \Big|_{x=x_i} + \frac{\partial I}{\partial y} \Delta y \Big|_{y=y_j} + \frac{\partial I}{\partial t} \Delta t \Big|_t. \tag{2}$$

For small motion and for short time intervals, the optical flow equation from Eq. (1) and Eq. (2) becomes:

$$\frac{\partial I}{\partial x} \Delta x \Big|_{x=x_i} + \frac{\partial I}{\partial y} \Delta y \Big|_{y=y_j} + \frac{\partial I}{\partial t} \Delta t \Big|_t = 0. \tag{3}$$

Since $\frac{\partial I}{\partial t} \Delta t$ at some time t in Eq. (3) is the change in intensity at a point between two frames, i.e., since $\frac{\partial I}{\partial t} \Delta t \Big|_t = I(x_i, y_j, t + \Delta t) - I(x_i, y_j, t)$, Eq. (3) can be rewritten as:

$$\frac{\partial I}{\partial x} \Delta x \Big|_{x=x_i} + \frac{\partial I}{\partial y} \Delta y \Big|_{y=y_j} = I(x_i, y_j, t) - I(x_i, y_j, t + \Delta t). \tag{4}$$

To compute gradients in Eq. (4), instead of using the Horn-Schunck kernel [20] as recommended in the standard implementation, we prefer to use the Sobel kernel that was shown to work better [5]. Since intensities ($I(t)$ and $I(t + \Delta t)$) are known, and the gradients $\partial I/\partial x$ and $\partial I/\partial y$ can be determined, Eq. (4) has two unknowns in Δx and in Δy , and to solve for these, usually an additional smoothness constraint is introduced [20]. However, since displacements are of interest only in the direction of the intensity gradients, Eq. (4) can be rewritten as:

$$|\nabla I| \Delta M(x_i, y_j, t) = I(x_i, y_j, t) - I(x_i, y_j, t + \Delta t), \tag{5}$$

wherein $|\nabla I| = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 \Big|_{x=x_i} + \left(\frac{\partial I}{\partial y}\right)^2 \Big|_{y=y_j}}$ is the intensity gradient, and ΔM is the displacement in the direction of the intensity gradient. Since tool displacements are assumed to be small, the displacements are evaluated by monitoring pixel intensities between the reference image $I_0(x_i, y_j)$ and subsequent images $I(x_i, y_j, t)$:

$$\Delta M(x_i, y_j, t) = \frac{I_0(x_i, y_j) - I(x_i, y_j, t)}{|\nabla I_0|} \tag{6}$$

wherein $|\nabla I_0|$ is the intensity gradient matrix computed with respect to the reference image.

The image derivatives required to evaluate intensity gradient in Eq. (6) is given by Eq. (7). The gradients are normalized with respect to the kernel size such that the magnitude of the mean pixel intensity remains the same in the resultant image gradient matrix:

$$I_x|_t = \frac{1}{(p-1) \cdot \text{sum}(|\mathbf{G}_x|)} [(I(t) * \mathbf{G}_x + I_0 * \mathbf{G}_x), I_y|_t] \\ = \frac{1}{(p-1) \cdot \text{sum}(|\mathbf{G}_y|)} [(I(t) * \mathbf{G}_y + I_0 * \mathbf{G}_y). \tag{7}$$

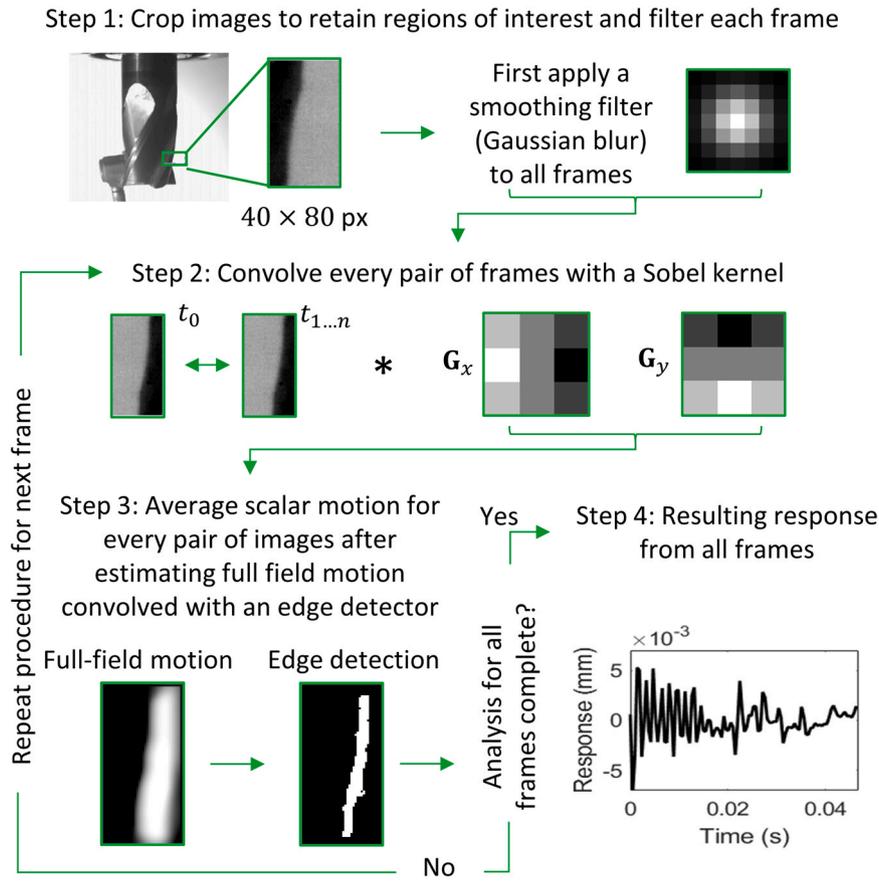


Fig. 4. Overview of the intensity-based optical flow scheme to register subpixel level motion.

wherein G_x and G_y are Sobel gradient kernels in the X and Y directions, and $\text{sum}(|G_x|)$ and $\text{sum}(|G_y|)$ represents the sum of the absolute entries, and 'p' represents the size of the kernel matrix.

Since the resulting motion matrix, M contains motion estimated for each pixel in the frame which may be influenced by background noise, we convolve the motion matrix with a Sobel edge detector, to result in the tool's averaged displacement over the detected tool's edge as:

$$y(t) = M(t) * \left[\frac{1}{\text{nnz}(\text{edge}(I))} \text{edge}(I) \right]_t \quad (8)$$

wherein the $\text{edge}(\cdot)$ operator represents the Sobel operator and $\text{nnz}(\cdot)$ represents number of non-zero entries in an input matrix. The displacements thus obtained from every pair of images are averaged over the region of interest in the third step, and the procedure is repeated for every successive pairs of images to finally extract tool motion in the final step.

Phase-based optical flow

We estimate the phase-based optical flow field of an image sequence in five steps. These steps are outlined in Fig. 5. The first step remains the same as that in the intensity-based scheme in which we crop all images to retain regions at or near the cutting edge(s) of the end mill. And, even though the phase-based scheme is said to be more robust to noise and potential changes in illumination than the intensity-based scheme, we still filter every frame with a Gaussian blur with a standard deviation of one. This is done to minimize the influence of noise. In the second main step, we spatially filter every frame with a set of quadrature pairs of the Gabor filter and compute their local phase responses. To do so, we convolve the intensity of

every spatial location with a Gabor filter kernel such that we obtain the local amplitude and local phase information as:

$$A_\theta(x, y, t) e^{i\phi_\theta(x, y, t)} = (G_2^\theta + iH_2^\theta) * I(x, y, t) \quad (9)$$

wherein $A_\theta(x, y, t)$ is the local amplitude and $\phi_\theta(x, y, t)$ is the local phase of the complex output of the filter measured for location (x, y) at frame corresponding to time t . The Gabor filter has a real (G_2^θ) and an imaginary component (H_2^θ) representing orthogonal directions and that forms a quadrature pair which differs in phase by 90° .

For the Gabor filter to work effectively, it must be tuned for the image sequence of interest. The filter can be thought of as a sinusoidal signal of a particular frequency and orientation, modulated by a Gaussian wave [21]. The parameters that control its shape, size, and performance are its orientation (θ), its wavelength and its bandwidth. Since the tool's motion of interest for us is horizontal, the orientation is kept zero, i.e., $\theta = 0$. For the video that resulted in $83 \mu\text{m}$ pixels, we tuned the Gabor filter to have a wavelength of 10 pixels per cycle and a bandwidth of three octaves. This resulted in a kernel size of 69×69 pixels.

In the phase-based approach, the contour of the local phase is assumed to be constant and its motion through time corresponds to the displacement signal [18]. Displacements are extracted from the motion of constant contour of local phase in time, which is expressed as [19]:

$$\phi_\theta(x, y, t) = c \quad (10)$$

wherein c denotes some constant. Differentiating Eq. (10) with respect to time gives:

$$\nabla \phi_\theta(x, y, t) \cdot (u, v, 1) = 0 \quad (11)$$

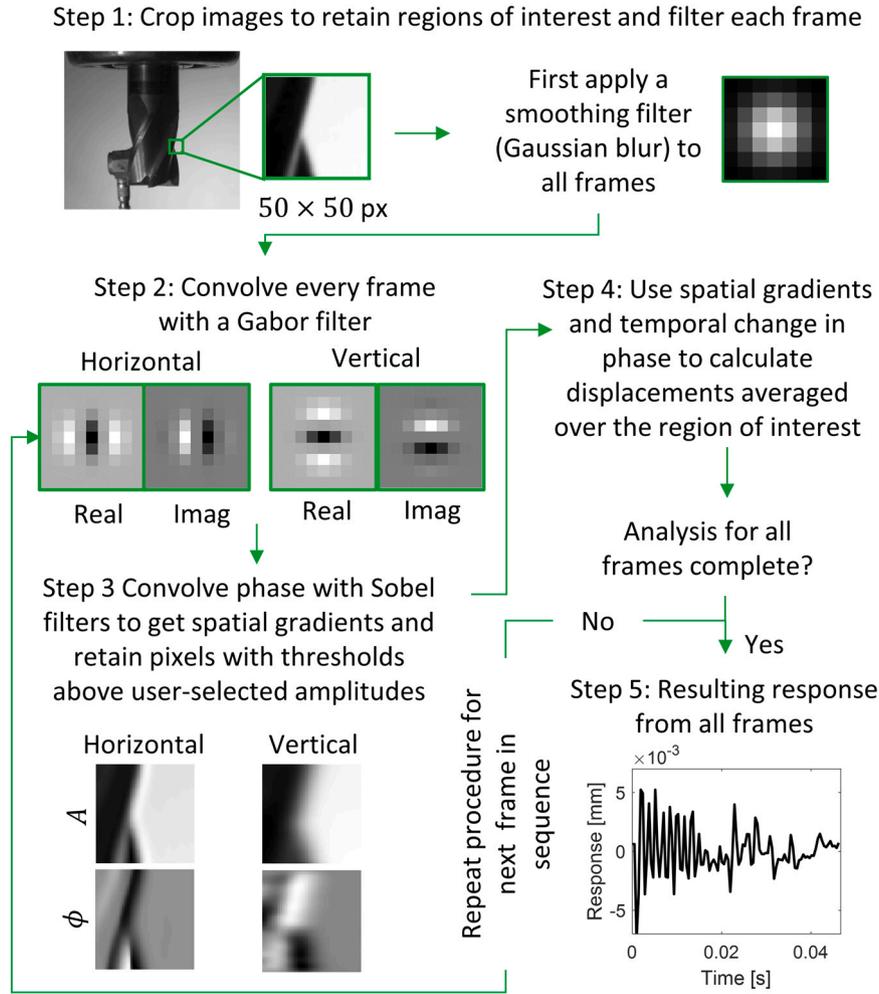


Fig. 5. Overview of the phase-based optical flow scheme to register subpixel level motion.

wherein $\nabla\phi_e(x, y, t) = (\phi_{e_x}, \phi_{e_y}, \phi_{e_t})$, and u, v denotes the horizontal and vertical velocity of the local phase contours moving between the first frame and subsequent frames. Making use of the fact that $\partial\phi_0/\partial y \approx 0$ and $\partial\phi_{90}/\partial x \approx 0$, we can write:

$$u \frac{\partial\phi_0}{\partial x} + \frac{\partial\phi_0}{\partial t} = 0 \Rightarrow u = -\left(\frac{\partial\phi_0}{\partial x}\right)^{-1} \frac{\partial\phi_0}{\partial t}, v \frac{\partial\phi_{90}}{\partial x} + \frac{\partial\phi_{90}}{\partial t} = 0 \Rightarrow v = -\left(\frac{\partial\phi_{90}}{\partial x}\right)^{-1} \frac{\partial\phi_{90}}{\partial t}. \quad (12)$$

Integrating Eq. (12) with respect to time, we obtain the displacements as:

$$d_x(t) = \int_t u(\tau) d\tau = -\left(\frac{\partial\phi_0}{\partial x}\right)^{-1} [\phi(x, y, t) - \phi(x, y, 0)],$$

$$d_y(t) = \int_t v(\tau) d\tau = -\left(\frac{\partial\phi_{90}}{\partial y}\right)^{-1} [\phi(x, y, t) - \phi(x, y, 0)]. \quad (13)$$

Spatial gradients in Eq. (12) are obtained by convolving the phase with Sobel kernels. This convolution operation forms the third main step in obtaining displacements. Furthermore, since only displacements in regions with sufficient contrast are treated as reliable, we use an amplitude threshold to retain only those pixels with intensities beyond the assigned threshold. For the frame sequence of interest herein, the amplitude threshold is set as the median of amplitude of 60 pixels with the largest amplitude in the reference frame. Thresholding also forms part of step three. Though Eq. (13)

results in a point displacement, the displacement in every frame is averaged over the region of interest to result in a scalar value. This forms part of step four. The final step involves repeating steps two to four for all subsequent frames to result in an estimate of the full displacement vector within the region of interest.

The intensity-based and the phase-based methods can both register motion at the subpixel level. It has in fact been suggested that the per-pixel resolution of these methods is governed by the intensity depth in bits [16]. For the Chronos 2.1 camera in use, since the camera is of the 12-bit intensity type, the intensity values in pixels may take any value between 0 and 4095 ($2^{12} = 4096$), giving a $1/4095$ or $\approx 2.4 \times 10^{-4}$ pixel-displacement resolution. And, since the pixel size with which these subpixel motion registration schemes are to be used was evaluated to be $83 \mu\text{m}$, the pixel resolution becomes sub-micrometers. Small motion is thus expected to be estimated correctly using these methods.

Comparative analysis of motion and modal parameters

This section compares response estimated using the different hardware setups as well as the software-based methods. Modal parameters evaluated from the registered motion are also compared. Response for the hardware-based experiments that each resulted in a different per-pixel resolution was estimated using a standard edge detection and tracking scheme. The parameters thus evaluated are compared with those evaluated from responses estimated using the intensity-based and phase-based methods applied to video recorded

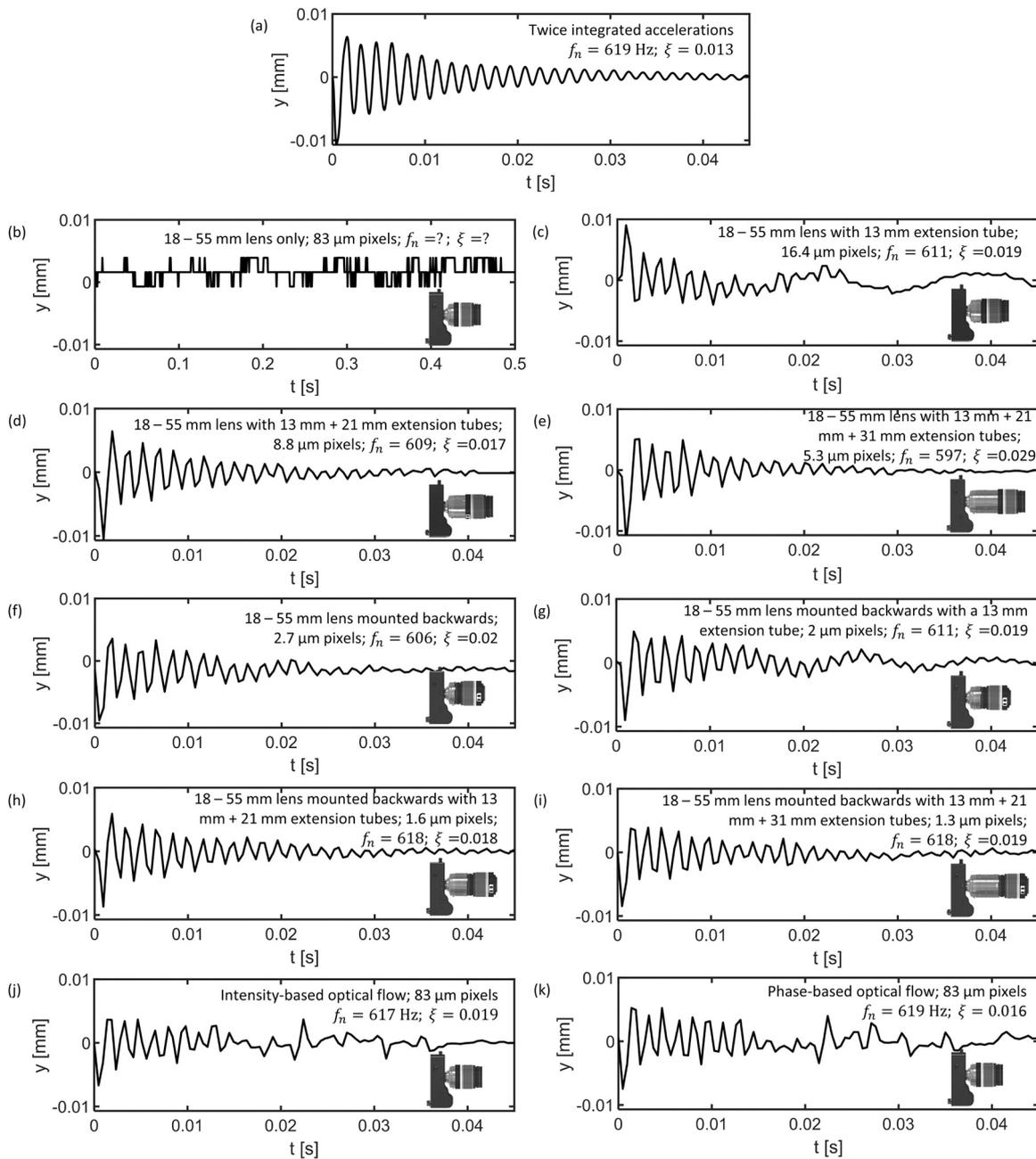


Fig. 6. Response estimated from (a) twice integrated accelerations and video recorded with: (b) 18–55 mm lens, (c) 18–55 mm lens and a 13 mm extension tube, (d) 18–55 mm lens and 13 mm and 21 mm extension tubes, (e) 18–55 mm lens and 13 mm, 21 mm, and 31 mm extension tubes, (f) 18–55 mm lens mounted backwards, (g) 18–55 mm lens mounted backwards and a 13 mm extension tube, (h) 18–55 mm lens mounted backwards and 13 mm and 21 mm extension tubes, (i) 18–55 mm lens mounted backwards and 13 mm, 21 mm, and 31 mm extension tubes, (j) 18–55 mm lens and using the intensity-based scheme, (k) 18–55 mm lens and using the phase-based scheme.

using the standard forward mounted 18–55 mm lens without extension tubes that resulted in 83 μm pixels. All modal parameters and responses are benchmarked against those evaluated from twice integrated accelerations that were measured concurrently while recording the video. To make comparisons meaningful, the input force for all impulse-like excitations with a modal hammer was kept at the same level of ~250 N. And, though the input force was monitored, it was not measured to be synchronized with video. And as such, the eigenvectors if estimated would not be mass normalized. Hence, for comparisons herein, only the natural frequencies and damping ratios are estimated from a run of the standard eigen-system realization algorithm (ERA) [22,23].

Motion estimated from 11 different measurements together with modal parameters evaluated from that registered motion are shown in

Fig. 6. Modal parameters are also tabulated in Table 1. The 11 different responses comprise of one measured using the accelerometer – see Fig. 6(a). Another response is estimated from the edge detection and tracking scheme applied to the video recorded with the 18–55 mm lens mounted in the usual forward fashion – see Fig. 6(b). Three more responses correspond to introducing different sets of extension tubes in between the camera body and the forward mounted lens – see Fig. 6(c–e). Another response corresponds to the lens mounted backwards – see Fig. 6(f). Three more responses correspond to different combinations of extension tubes together with the lens mounted backwards – see Fig. 6(g–i). And two responses correspond to those evaluated using the intensity-based scheme (see Fig. 6(j)) and the phase-based scheme (see Fig. 6(k)) applied to the video recorded with the 18–55 mm mounted in the usual forward fashion.

Table 1
Modal parameters evaluated from response estimated from different methods.

Motion registration method	Schematic	Per-pixel resolution [μm]	Natural frequency [Hz]	Damping ratio
Accelerometer		–	619	0.013
Edge detection and tracking	With 18–55 mm lens	83	–	–
	With 18–55 mm lens + 13 mm extension tube	16.4	611	0.019
	With 18–55 mm lens + 13 mm + 21 mm extension tubes	8.8	609	0.017
	With 18–55 mm lens + 13 mm + 21 mm + 31 mm extension tubes	5.3	597	0.029
	With 18–55 mm lens mounted backwards	2.7	606	0.02
	With 18–55 mm lens mounted backwards + 13 mm extension tube	2	611	0.019
	With 18–55 mm lens mounted backwards + 13 mm + 21 mm extension tubes	1.6	618	0.018
	With 18–55 mm lens mounted backwards + 13 mm + 21 mm + 31 mm extension tubes	1.3	618	0.019
Optical flow	With 18–55 mm lens using the intensity-based scheme	83	617	0.019
	With 18–55 mm lens using the phase-based scheme	83	619	0.016

The response of the end mill is characterized by a single dominant mode at $f_n = 619\text{Hz}$ that has a damping ratio of $\xi = 0.013$. These parameters were obtained from a run of the ERA on twice integrated accelerations that were sampled at a rate of 51.2 kHz using a NI-9234 analog to digital convertor together with a NI-9171 chassis.

For vision measurements with just the 18–55 mm lens with which the per-pixel resolution is $83\mu\text{m}$, no perceptible motion is registered using the edge detection and tracking scheme as can be observed from Fig. 6(b), and as such no modal parameters are estimated. This suggests that the motion has amplitudes less than that is easily resolvable with the per-pixel resolution of $83\mu\text{m}$. Displacements obtained from twice integrated accelerations shown in Fig. 6(a) indeed show the response amplitude for the first few cycles to be as low as $\sim 5\mu\text{m}$. This response appears partially resolvable with the hardware setup with one extension tube that makes the per-pixel resolution $16.4\mu\text{m}$. The natural frequency estimated from this response shown in Fig. 6(c) was evaluated to be 611 Hz, i.e., a difference of $\sim 1.2\%$ from the 'true' natural frequency estimated from accelerations. Though frequencies match up well, there is a large difference in the estimate of the damping ratio. The difference is damping estimates can be attributed to the edge detection and tracking scheme not being strictly a subpixel level motion registration scheme, i.e., despite averaging response over several pixels in the region of interest, since the per-pixel resolution in this case is greater than the response amplitude for the first few cycles, the true decaying nature of the response is not characterized correctly – as is evident from Fig. 6(c).

Results with two extension tubes with which the per-pixel resolution was $8.8\mu\text{m}$ were found to be like those with one extension tube – as can be seen in Fig. 6(d). For the case of three extension tubes sandwiched between the camera body and the lens that

results in a smaller per-pixel resolution of $5.3\mu\text{m}$, results for which are shown in Fig. 6(e), the error in natural frequencies and damping ratios surprisingly increase to become $\sim 3.5\%$ and $\sim 123\%$. Since sandwiching more extension tubes results in less light falling on the image detector even with the DC light intensity being cranked up, the image becomes dull. And, since the edge detection and tracking algorithm works best with sharp intensity gradients across the tool's edge, even though the per-pixel resolution was reduced to $5.3\mu\text{m}$ with the help of three extension tubes, this does not translate to better motion registration.

With the lens being flipped to be mounted backwards using a reverse ring, response estimated with a per-pixel resolution of $2.7\mu\text{m}$ is shown in Fig. 6(f). The natural frequency evaluated from this response differs by $\sim 2\%$ when compared to the benchmark. Damping in this case too is overestimated. This setup results in a smaller per-pixel resolution than with three tubes sandwiched together with the lens mounted forwards. And, since the tool is nearer to the image detector in this case, motion is better resolved. Introduction of one, two and three extension tubes with the reverse ring setup, results for which as shown in Fig. 6(g-i), further reduces the per-pixel resolution to $2\mu\text{m}$, to $1.6\mu\text{m}$, and to $1.3\mu\text{m}$, respectively, and that in turn brings the differences in natural frequencies to be less than 0.2%. Differences in damping estimates remain high, ranging between 40% and 50%. As is obvious from observations from Fig. 6(a-i) and Table 1, use of extension tubes in combination with a backward mounted lens results in better motion registration and modal parameter estimates than the use extension tubes with the lens mounted in the usual forward configuration.

Though the use of the edge detection and tracking scheme was unable to resolve motion with per-pixel amplitudes of $83\mu\text{m}$, when the intensity-based and phase-based subpixel level motion registration schemes are applied to the same video, the motion is easily

resolvable – as is evident from the responses shown in Fig. 6(j-k). The natural frequency evaluated from motion estimated from the intensity-based scheme differs by less than 0.3% compared to the benchmark, and for the case of the phase-based scheme, the natural frequency is estimated correctly, with no error. The difference in damping too evaluated from motion registered with the phase-based scheme differs by 23% in comparison to the benchmark. This is less than the difference as compared to the intensity-based scheme and is also better than all the hardware-based experiments. This is likely due to the phase-based scheme being more robust to changes in illumination and noise.

Conclusions and outlook

This paper discussed methods to resolve subpixel level small motion from video recordings of cutting tools. We demonstrated that it is possible to estimate small motion using hardware-based as well as software-based techniques. In the hardware-based techniques, we showed that it was possible to improve the per-pixel resolution to as less as $1.3\ \mu\text{m}$ using combinations of extension tubes sandwiched between the camera body and a standard lens mounted backwards. This reduction was more than what was possible with different combinations of extension tubes sandwiched between the camera body and the lens mounted in the usual forward fashion. In general, we found that the use of extension tubes significantly reduced the per-pixel resolution from being $83\ \mu\text{m}$ with just a standard lens to a few μm . This reduction made it possible to use the established edge detection and tracking scheme to register small cutting tool motion with small pixel sizes. Amongst the software-based approaches we found the phase-based optical flow scheme to fare better than the intensity-based optical flow scheme in registering subpixel level small motion even with pixel sizes of $83\ \mu\text{m}$. Modal parameters evaluated from motion estimated using both methods agreed with those evaluated from twice integrated accelerations.

Since methods proposed herein proffer new ways to register small oscillatory motion, we believe that our findings will further help leverage the adoption of vision-based modal analysis methods in machine tool systems as an alternative to the more traditional experimental modal analysis procedures. Furthermore, methods proposed herein are equally applicable to vision-based modal analysis for micro tooling systems that will likely vibrate with smaller amplitudes than the macro tool geometry discussed herein. Methods we have proposed can also easily be applied to study material deformation mechanisms during cutting. Methods are also generalized for use across other engineering domains.

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.

Acknowledgements

This work was supported by the Government of India's Science and Engineering Research Board's Core Research Grant number CRG/2020/001010.

References

- [1] Feng, D., Feng, M.Q., 2018, Computer Vision for SHM of Civil Infrastructure: From Dynamic Response Measurement to Damage Detection—A Review. *Engineering Structures*, 156:105–117. <https://doi.org/10.1016/j.engstruct.2017.11.018>.
- [2] Xu, Y., Brownjohn, J.M.W., 2018, Review of Machine-Vision Based Methodologies for Displacement Measurement in Civil Structures. *Journal of Civil Structural Health Monitoring*, 8/1:91–110. <https://doi.org/10.1007/s13349-017-0261-4>.
- [3] Spencer Jr, B.F., et al., 2019, Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, 5:199–222. <https://doi.org/10.1016/j.jeng.2018.11.030>.
- [4] Law, M., et al., 2020, Modal Analysis of Machine Tools Using Visual Vibrometry and Output-only Methods. *CIRP Annals Manufacturing Technology*, 69/1: 357–360. <https://doi.org/10.1016/j.cirp.2020.04.043>.
- [5] Gupta, P., et al., 2021, Vision-based Modal Analysis of Cutting Tools. *CIRP Journal of Manufacturing Science and Technology*, 32:91–107. <https://doi.org/10.1016/j.cirpj.2020.11.012>.
- [6] Gupta, P., Law, M., 2021, Evaluating Tool Point Dynamics Using Smartphone-based Visual Vibrometry. *Procedia CIRP*, 101:250–253. <https://doi.org/10.1016/j.procir.2020.09.196>.
- [7] Huang, H., et al., 2022, Vision-based Vibration Measurement of Machine Tool. *Journal of Advanced Mechanical Design, Systems, and Manufacturing*, 16:1–12. <https://doi.org/10.1299/JAMDSM.2022JAMDSM00014>.
- [8] Law, M., et al., 2022, Modal Parameter Recovery from Temporally Aliased Video Recordings of Cutting Tools. *CIRP Annals Manufacturing Technology*. <https://doi.org/10.1016/j.cirp.2022.03.023>.
- [9] Lambora, R., et al., 2022, Recovering Cutting Tool Modal Parameters from Fractionally Uncorrelated and Potentially Aliased Signals. *CIRP Journal of Manufacturing Science and Technology*, 38:414–426. <https://doi.org/10.1016/j.cirpj.2022.05.014>.
- [10] Reu, P.L., et al., 2015, Camera System Resolution and its Influence on Digital Image Correlation. *Experimental Mechanics*, 55:9–25. <https://doi.org/10.1007/s11340-014-9886-y>.
- [11] Percoco, G., et al., 2015, Preliminary Study on the 3D Digitization of Millimeter Scale Products by Means of Photogrammetry. *Procedia CIRP*, 33:257–262. <https://doi.org/10.1016/j.procir.2015.06.046>.
- [12] Galantucci, L.M., et al., 2015, A Stereo Photogrammetry Scanning Methodology, for Precise and Accurate 3d Digitization of Small Parts With Sub-millimeter Sized Features. *CIRP Annals Manufacturing Technology*, 64:507–510. <https://doi.org/10.1016/j.cirp.2015.04.016>.
- [13] Guerra, M.G., et al., 2019, Measuring Techniques Suitable for Verification and Repairing of Industrial Components: A Comparison Among Optical Systems. *CIRP Journal of Manufacturing Science and Technology*, 27:114–123. <https://doi.org/10.1016/j.cirpj.2019.09.003>.
- [14] HajiRassouliha, A., et al., 2021, The Effect of Camera Settings on Image Noise and Accuracy of Subpixel Image Registration. *Machine Vision and Applications*, 32:1–10. <https://doi.org/10.1007/s00138-021-01215-4>.
- [15] Fu, Y., et al., 2021, Non-contact Optical Dynamic Measurements at Different Ranges: A Review. *Acta Mechanica Sinica*, 37:537–553. <https://doi.org/10.1007/s10409-021-01102-1>.
- [16] Javh, J., et al., 2017, The Subpixel Resolution of Optical-Flow-Based Modal Analysis. *Mechanical Systems and Signal Processing*, 88:89–99. <https://doi.org/10.1016/j.ymssp.2016.11.009>.
- [17] Sutton, M.A., 2009, *Image Correlation for Shape, Motion, and Deformation Measurements: Basic Concepts, Theory and Applications*. SpringerLink. ISBN 978-0-387-78747-3.
- [18] Luan, L., et al., Extracting Full-field Subpixel Structural Displacements from Videos Via Deep Learning. *Journal of Sound and Vibration*, 505:116142. <https://doi.org/10.1016/j.jsv.2021.116142>.
- [19] Fleet, D.J., Jepson, A.D., 1990, Computation of Component Image Velocity from Local Phase Information. *International Journal of Computer Vision*, 5:77–104. <https://doi.org/10.1007/BF00056772>.
- [20] Horn, B.K.P., Schunck, B.G., 1981, Determining Optical Flow. *Artificial Intelligence*, 319–331. [https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2).
- [21] Movellan, J.R., 2008, Tutorial on Gabor Filters. Tutorial Paper. (<https://inc.ucsd.edu/mplab/75/media/jgabor.pdf>).
- [22] Juang, J., Pappa, R., 1985, An Eigensystem Realization Algorithm for Modal Parameter Identification and Model Reduction. *Journal of Guidance Control and Dynamics*, 8:620–627. <https://doi.org/10.2514/3.20031>.
- [23] Gupta, P., et al., 2020, Evaluating Tool Point Dynamics Using Output-only Modal Analysis with Mass-change Methods. *CIRP Journal of Manufacturing Science and Technology*, 31:251–264. <https://doi.org/10.1016/j.cirpj.2020.06.001>.