

VISION-BASED MODAL ANALYSIS OF MACHINE TOOL SYSTEMS: PROGRESS AND PROSPECTS

Mohit Law

Machine Tool Dynamics Laboratory, Department of Mechanical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India; Tel.: +91 5126796897; Fax: +91 5126797408, E-mail: mlaw@iitk.ac.in

Original Manuscript Submitted: 12/25/2023; Final Draft Received: 3/14/2024

Modal analysis of machine tools involves estimating natural frequencies, damping ratios, and mode shapes from the vibratory response of the machine tool. Usually, modal hammers and shakers are used to excite, and accelerometers or laser vibrometers are used to measure the response. Though these procedures have become routine, sometimes the use of accelerometers can result in mass-loading that distorts the response, and though laser vibrometers are non-contact, their use is precluded by their high costs. To counter these issues, vision-based modal analysis methods have emerged as a viable and promising alternative. The spatiotemporal response is estimated by treating every pixel in every frame in the video of the vibrating machine as a motion sensor. Image processing schemes leveraged from developments in allied fields are then used to register motion from video. The method is non-contact, full field, and only needs a camera and post-processing on a computer, and as such, it offers advantages over the traditional measurement methods. Since vision-based methods are potentially paradigm-shifting, this paper reviews the recent progress to contextualize the prospects of the method. The review includes discussions on selection considerations of cameras and acquisition parameters, on using markers and the machine's own features to register motion, on the efficacy of different motion registration schemes, and workarounds for when motion is spatiotemporally aliased. The paper concludes by discussing challenges and prospects related to motion synchronization, measuring speed and time-varying dynamics, and technological trends that may aid the adoption of the method.

KEY WORDS: *machine tools, vibrations, vision, image processing, modal analysis*

1. INTRODUCTION

The cutting process excites the machine tool and makes it vibrate. Since vibrations can damage elements of the machine tool, it is important to characterize how the machine vibrates prior to cutting. Modal analysis helps characterize machine tool vibrations by estimating natural frequencies, damping ratios, and mode shapes from the vibratory response of the machine tool system. Usually, modal hammers and shakers are used to excite, and accelerometers or laser vibrometers are used to measure the response. These procedures have become routine (Okubo et al., 1982; Brown and Allemang, 2007; Iglesias et al., 2022). However, the use of accelerometers can sometimes result in mass-loading, which distorts the response, and though laser vibrometers are non-contact, their use is precluded by their high costs. Shape analysis also requires that the actuator or the response transducer(s) be moved around a measurement grid. This takes time and requires sophisticated multi-channel data acquisition hardware. To ad-

dress these issues, vision-based modal analysis methods have emerged as a viable and promising alternative.

In vision-based measurements, the spatiotemporal response is estimated by treating every pixel in every frame in the video of the vibrating machine as a motion sensor. Image processing schemes leveraged from developments in allied fields are then used to register motion from video. The method is non-contact, full field, and only needs a camera and post-processing on a computer, and as such, it offers advantages over the traditional measurement methods. Since vision-based machine tool vibration measurement methods are potentially paradigm shifting, this paper reviews the recent progress to contextualize the prospects of the method.

The use of image processing and computer vision techniques are not new to machine tools. Prior uses range from monitoring tool wear and inspecting workpiece quality (Kurada and Bradley, 1997; Dutta et al., 2013; Singh et al., 2023), to *in situ* monitoring of machining processes (Guo et al., 2015; Raizada et al., 2024), to monitoring thermal growth (Vogl et al., 2023), to vision-based machine tool metrology (Mori et al., 2023; Verma et al., 2024). There are excellent papers reviewing the progress in some of these areas, though there exists no review of vision-based machine tool vibration measurements. Additionally, there are some seminal review papers on vision-based vibration measurements and modal analysis of civil infrastructure (Feng and Feng, 2018; Spencer et al., 2019; Beberniss and Ehrhardt, 2017; Baqersad et al., 2017), but since civil infrastructure vibrates with lower frequencies and larger amplitudes than elements of machine tools, those methods and their findings are not of direct relevance to machine tools, which can vibrate in frequencies ranging from a few Hz to a few kHz and with small and micrometer level motion.

Vision-based modal analysis of machine tool systems is still in its nascency, with only a few research groups reporting their findings on the topic. Since the method holds much promise, this paper aims to: (i) review the significant progress, and (ii) outline challenges and discuss prospects of the method.

To keep the review focused, only literature in which vision-based methods have been used to estimate machine tool vibratory motion are reviewed. In perusing the related literature, 16 such papers were identified. These date back to 2018, i.e., most are fairly recent. Of these, 11 are authored by the author himself (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Nuhman et al., 2022; Raizada and Law, 2023; Singh and Law, 2023; Rajput et al., 2023; Rajput and Law, 2024; Raizada et al., 2024). The others are authored by groups working in Hungary (Berezvai et al., 2018), Sweden (Yanamadala and Muralidharan, 2021), Japan (Hunag et al., 2022), Poland (Czyzycki et al., 2021), and China (Zheng et al., 2023). Though broadly concerned with estimating vibratory motion, different groups have estimated the motion of different elements of the machine tool system. There are also differences in the hardware used, in acquisition parameters, in using markers and features of the vibrating artifact itself to track motion, and in the choice of the image processing scheme used to register motion. Since each of these influences registered motion and modal parameters extracted from that motion, and since vision-based

methods are fast evolving, a review of the progress made thus far is thought to be timely and necessary.

The remainder of the paper is structured as follows. First, in Section 2, a general overview of vision-based modal analysis is provided. Section 2 also discusses the differences between different approaches. Section 3 details the different methodologies to acquire and register motion. Section 4 discusses representative results. Section 5 outlines some of the challenges that are yet to be addressed and discusses prospects of vision-based modal analysis, including technological trends in hardware. The paper is concluded in Section 6.

2. OVERVIEW OF VISION-BASED MODAL ANALYSIS

This section overviews the general procedure for registering motion and extracting modal parameters from that motion. Figure 1 shows this overview. First, considerations for cameras, illumination, background, and markers are discussed. This is followed by

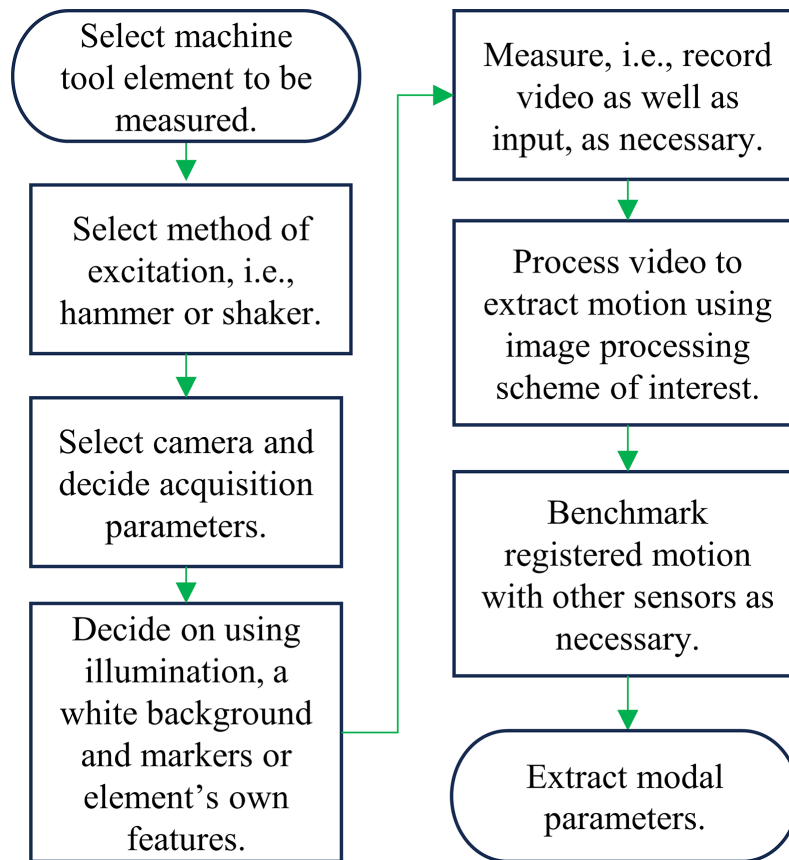


FIG. 1: Overview of the procedure for vision-based modal analysis

discussions on typical experimental setups, including how to excite, how to synchronize motion with inputs, and how to register motion using image processing schemes. Finally, the extraction of modal parameters is discussed.

2.1 Selection Considerations for Cameras

Assuming that the range of natural frequencies of the elements of the machine tool to be measured are known *a priori* along with their expected vibration amplitudes, suitable cameras can be selected. This selection must ensure that the frame rate satisfies the Nyquist criterion, i.e., the camera speed is at least twice the highest natural frequency of interest to be measured. Likewise, the selection of the camera and the lens must also ensure that motion is spatially resolved, i.e., the resolution is sufficient enough such that the motion magnitude spans two or more pixels (Reu et al., 2015). However, since it may not always be possible to select cameras with sufficient speeds and resolutions to ensure motion is spatiotemporally well resolved, methods to register motion in such cases are discussed separately in Section 3.

Different cameras and acquisition parameters used by different groups are listed in Table 1. The table also lists the pixel pitch in μm . As is evident from the table, cameras used include everyday use smartphone cameras recording motion at 30 frames per second (fps), 480 fps, and 960 fps at resolutions ranging from 720×1280 to 1080×1920 pixels (Gupta and Law, 2021; Yanamadala and Muralidharan, 2021; Singh and Law, 2023), medium speed cameras recording motion ranging from 1000 Hz to 5000 Hz with resolutions ranging from 600×800 to 720×1280 (Law et al., 2020, 2022; Huang et al., 2022), respectively, and high-speed cameras recording motion up to 16000 Hz with resolutions in the range of 288×3840 (Berezvai et al., 2018; Czyzycki et al., 2021). The pixel pitch varies from $1.3 \mu\text{m}$ to $197 \mu\text{m}$. These differences are due to the measured artifact in each case being different and vibrating at different frequencies and with different amplitudes, and the camera acquisition parameters (frame rate and resolution) also being different.

2.2 Illumination and Background Considerations

Since most image processing-based motion registration schemes use pixel intensities and their gradients to track if the artifact being recorded is moving, it is important to appropriately illuminate the vibrating element. Ideally, to avoid flickering of alternating current powered lights at supply frequency or its harmonics, such as the case when using fluorescent lamps that use core-coil magnetic ballasts, experiments must be conducted under natural conditions. However, since equipment is almost always housed indoors, DC lights are recommended. If and when possible, these should be with adjustable luminance and gains. A summary of different illumination methods is also presented in Table 1. Furthermore, to minimize the influence of background noise and to increase the contrast of the artifacts being recorded, it is recommended to use whiteboards in the background of the element being recorded.

TABLE 1: Summary of different studies on vision-based vibration measurements of machine tool systems

Ref.	Measured artefact	Measured frequencies	Camera	Frame rate (fps)	Resolution (px × px)	Per-pixel resolution (µm)	illumination	White background?	Markers	Method of excitation	Synchronized inputs?	Image processing method	Benchmarking with	Method of modal parameter extraction
Berezvai et al. (2018)	End mill	Up to 1 kHz	Photron FASTCAM SA5	10,000	—	—	None	No	No	Cutting process	—	<i>Ad hoc</i>	—	FFT
Law et al. (2020)	End mill, boring bar, machine	Up to 1.25 kHz	OnePlus 7T; Photron FASTCAM SA 1.1	480; 5400	720 × 1280; 1024 × 1024	20–50	DC light	Yes	No	Modal hammer	No	Edge	Accelerations	ERA
Gupta et al. (2021)	End mill, boring bar, grooving blade	Up to 1.25 kHz	OnePlus 7T; Photron FASTCAM SA 1.1	480; 5400	720 × 1280; 1024 × 1024	20–37.5	DC light	Yes	No	Modal hammer	No	Edge; Optical flow; DIC	Accelerations	ERA
Gupta and Law (2021)	Grooving blade	Up to 200 Hz	OnePlus 7T	480	720 × 1280	37.5	—	Yes	No	Modal hammer	No	Edge	Accelerations and eddy current sensor	FFT and logarithmic decrement
Yanamadala and Muralidharan (2021)	Beam, machine, robot	Up to 15 Hz	iPhone 11 pro	30	1080 × 1920	—	None	Yes	Yes	Impact	No	OpenCV	Laser vibrometer	FFT
Czyzycki et al. (2021)	Thin wall	Up to 50 Hz	Phantom micro m310	16,000	288 × 3840	—	None	No	No	Cutting process	No	Motion capture software	Laser displacement sensor	FFT
Law et al. (2022)	End mill; Boring bar	Up to 1.4 kHz	Chronos 2.1HD	960; 1000	1080 × 1920	36	DC light	Yes	No	Modal hammer	No	Edge; Optical flow	Accelerations	ERA
Lamboora et al. (2022)	End mill	Up to 1.4 kHz	Chronos 2.1HD	960; 1000	1080 × 1920	35	DC light	Yes	No	Modal hammer	No	Edge; Optical flow	Accelerations	ERA
Nuhman et al. (2022)	End mill	Up to 1 kHz	Chronos 2.1HD	2142	720 × 1280	1.3–83	DC light	Yes	No	Modal hammer	No	Edge; Optical flow	Accelerations	ERA

TABLE 1: (Continued)

Ref.	Measured artefact frequencies	Camera	Frame rate (fps)	Resolution (px × px)	Per-pixel resolution (µm)	Illumination	White background?	Markers	Method of excitation	Synchronized inputs?	Image processing method	Benchmarking with	Method of modal parameter extraction
Huang et al (2022)	Beam, machine	Up to 50 Hz	DITECT, HASEF	1000	600 × 800	3.8	No	Yes	Piezo-electric	Yes	Edge	Capacitive displacement sensor	FFT
Raizada and Law (2023)	Grooving blade; end mill	Up to 1.25 kHz	Chronos 2.1HD	5046; 10488	480 × 800; 240 × 640	84–87	Yes	No	Modal hammer	No	Optical flow; CNN	Accelerations	ERA
Singh and Law (2023)	Machine tool	Up to 100 Hz	Samsung S20	960	720 × 1280	197	Yes	No	Modal hammer	No	Optical flow	Accelerations	ERA
Zheng et al. (2023)	Wire	Up to 150 Hz	Baslar acA720-290	291	480 × 800	—	Yes	No	Vibration exciter	No	Edge	Accelerations	FFT
Rajput and Law (2024)	Grooving blade; end mill	Up to 1 kHz	Chronos 2.1HD	480; 2500	1080 × 1920	83	DC light	Yes	Modal hammer	No	Edge	—	FFT
Rajput et al. (2024)	Grooving blade; End mill	Up to 1 kHz	OnePlus 7T; Chronos 2.1HD	240; 2142	720 × 1280	37.5–53	DC light	Yes	Modal hammer	No	CNN	Accelerations	ERA
Raizada et al. (2024)	Grooving blade; End mill	Up to 1 kHz	OnePlus 7T; Chronos 2.1HD	240; 2142	720 × 1280	36–37.5	DC light	Yes	Modal hammer	No	Edge; Optical flow; DIC; PIV; CNN	Accelerations	ERA

2.3 Tracking Markers or the Machine's Own Features

Markers placed on the object whose motion is being tracked help in sharper intensity gradients that make motion registration with image processing schemes easier. Hence, some studies have used markers (Yanamadala and Muralidharan, 2021; Huang et al., 2022). However, since the artifact being tracked is at times too small to place markers on, such as when tracking the cutting edge of a vibrating tool or the edge of a thin wall, then the features of the objects themselves are tracked to register motion (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Nuhman et al., 2022; Raizada and Law, 2023; Rajput et al., 2023; Rajput and Law, 2024; Raizada et al., 2024; Berezvai et al., 2018; Zheng et al., 2023). A summary of different studies using different features and/or markers is listed in Table 1.

2.4 Typical Experimental Setups

Five representative experimental setups used by different research groups are shown in Fig. 2. The setups show the use of different cameras and light sources. The setup shown in Fig. 2(a) is to record the motion of a vibrating end mill. The setup shown in Fig. 2(b) is to record motion of a slender boring bar. The setup shown in Fig. 2(c) is to record the motion of a thin workpiece wall. The setup shown in Fig. 2(d) is to record the motion of a wire. The setup shown in Fig. 2(e) is to record the motion of a grooving blade using a smartphone. In all of these setups, the camera can only record in-plane motion of the vibrating object. Since these artifacts can also vibrate out-of-plane, the setups are changed and indexed as necessary to also record out-of-plane motion. There are no reports of using stereo vision with two or more cameras to simultaneously record in- and out-of-plane motion. More on this aspect is discussed in Section 5, which outlines the challenges and prospects of vision-based methods.

2.5 Excitation and Synchronizing Response with Input Force

In general, like most modal analyses of machine tool systems (Iglesias et al., 2022), for vision-based modal analysis, the preferred method of exciting the system is to use modal hammers. The transient response is then recorded using cameras. However, some researchers have preferred piezoelectric-based forced excitation (Huang et al., 2022). There is also one report of exciting the system using the cutting process itself and recording the motion of the vibrating tool to detect if the tool vibrations grow to result in large amplitude chatter vibrations (Berezvai et al., 2018). A summary of different excitation methods used by different groups is summarized in Table 1.

In addition to the identification of modal parameters, another aim of modal analysis is to construct frequency response functions (FRF) in terms of their receptances, i.e., the frequency-dependent ratio of displacements to forces. This FRF requires

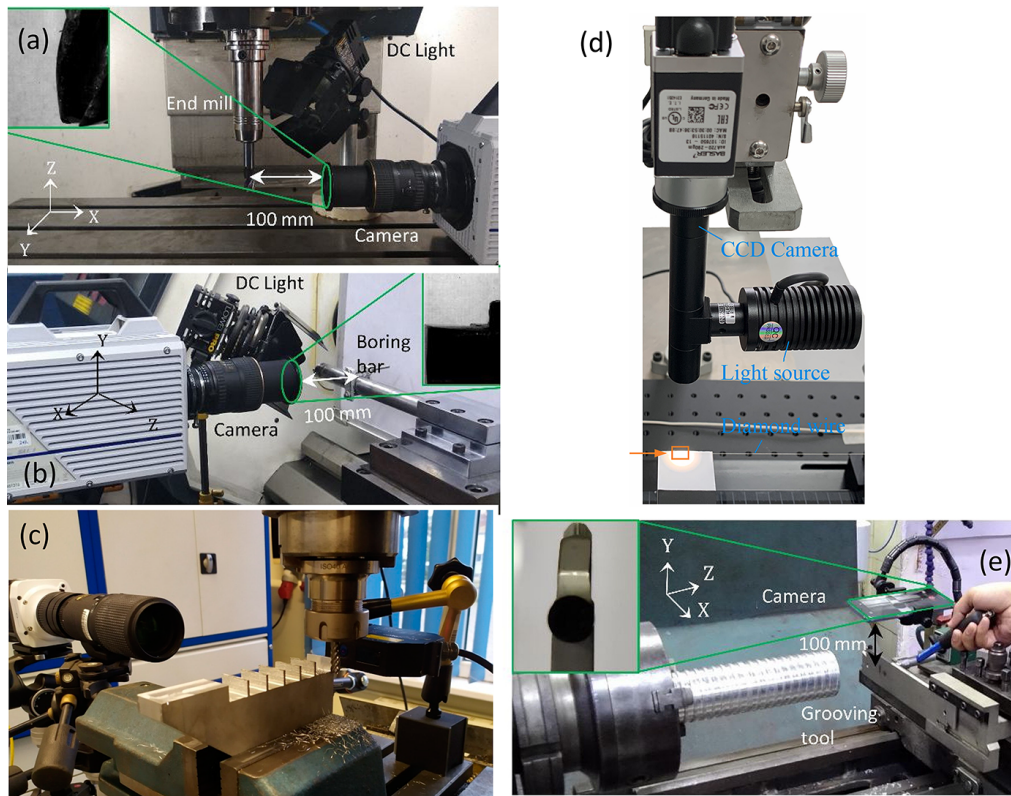


FIG. 2: Four representative experimental setups to record video of machine tool systems. (a) Setup to record video of an end mill (reprinted from Gupta et al. with permission from Elsevier, copyright 2021). (b) Setup to record video of a slender boring bar (reprinted from Gupta et al. with permission from Elsevier, copyright 2021). (c) Setup to record video of a thin wall (reprinted from Czyzycki et al. that was published as open access under the Creative Commons CC BY 4.0 license). (d) Setup to record video of a wire (reprinted from Zheng et al. with permission from Elsevier, copyright 2023). (e) Setup to record video of a grooving blade (reprinted from Gupta et al. with permission from Elsevier, copyright 2021).

that the input excitation force be measured and synchronized with the response. This is routine enough with traditional modal analysis procedures that use hammers and accelerometers in which the acceleration measurement is triggered when the input force (voltage) reaches a triggering threshold. However, since vision-based modal analysis is new, and since not all cameras can be triggered to start recording when the input force (voltage) reaches some threshold, the measured response is usually not synchronized with the input. As such, the construction of FRFs becomes difficult. Workarounds for this have been attempted by some groups using output-only modal analysis methods with structural modifications (Law et al., 2020). Table 1 summarizes which of the research groups attempted synchronization and the method of doing so.

2.6 Image Processing to Register Motion

Once the video of the vibrating artifact has been recorded, to register motion from that video, first the video is decomposed to its individual frames and converted to greyscale. Then, since the field of view is usually larger than the edges of the object whose motion must be tracked, all frames are cropped to retain only the region of interest. These pre-processing steps remain the same for all motion registration schemes used by different research groups. However, the method of motion extraction can differ. The 16 papers reviewed report 20 different image processing schemes to extract motion. These vary from a binary method (Berezvai et al., 2018), to the use of commercial motion capture software (Czyzycki et al., 2021), to the use of one of the many edge detection and tracking schemes (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Nuhman et al., 2022; Raizada et al., 2024; Hunag et al., 2022; Zheng et al., 2023), optical flow-based schemes (Gupta and Law, 2021; Nuhman et al., 2022; Raizada and Law, 2023; Raizada et al., 2024), the use of digital image correlation (DIC) (Gupta and Law, 2021; Raizada et al., 2024), the use of particle image velocimetry (PIV) (Raizada et al., 2024), object tracking using OpenCV's machine learning algorithms (Yanamadala and Muralidharan, 2021), and the use of deep learning based on convolution neural networks (Raizada and Law, 2023; Rajput et al., 2023; Raizada et al., 2024). Each scheme has its merits and demerits, with some being able to register small and subpixel level motion, some being robust to noise and illuminations, while others are more sensitive to acquisition parameters. A detailed review of each of these methods is discussed in Section 3. A summary, however, of which research group preferred which scheme is presented in Table 1.

2.7 Extracting Modal Parameters and Benchmarking

Modal parameters can be extracted from the registered motion using any of the standard system identification techniques. In some cases, since the response is dominated by a single mode, some research groups simply performed a fast Fourier transform (FFT) of the response to obtain the natural frequency of vibration and used the logarithmic decrement method to obtain an estimate of damping (Gupta and Law, 2021; Yanamadala and Muralidharan, 2021; Huang et al., 2023; Zheng et al., 2023). In other cases, when there were multiple modes in the system, researchers preferred the use of the eigensystem realization algorithm (ERA) (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Nuhman et al., 2022; Raizada et al., 2024). The ERA is an established and robust time-domain method of modal parameter extraction. A summary of different preferred modal parameter extraction methods is presented in Table 1.

In all cases, motion and/or modal parameters extracted from that motion were benchmarked with results obtained from using accelerometers (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Nuhman et al.,

2022; Raizada et al., 2024; Zheng et al., 2023), capacitance sensors (Huang et al., 2022), laser sensors (Czyzycki et al., 2021), or laser vibrometers (Yanamadala and Muralidharan, 2021). A summary of different methods to benchmark is included in Table 1.

3. METHODOLOGIES FOR VISION-BASED MODAL ANALYSIS

Having provided an overview of the procedures for vision-based modal analysis, and having discussed the considerations for the selection of hardware, acquisition parameters, motion registration schemes, and modal parameters extraction methods, this section details the key methods to address motion that are potentially spatiotemporal aliased followed by a detailed discussion on the workings and (de)merits of the different image processing schemes.

3.1 Spatiotemporal Aliasing

Natural frequencies of vibration of different elements of the machine tool system range from the tens of Hz to a few kHz. Amplitudes of vibrations at each of these natural frequencies range from a few μm to tens of μm , with further spatial variation of amplitudes for every mode shape. To correctly estimate frequencies and amplitudes, there is a need to use high-speed cameras with high resolutions. Such cameras are expensive. As such, some research groups have used medium- and low-speed cameras (Gupta and Law, 2021; Law et al., 2022; Lambora et al., 2022; Nuhman et al., 2022; Raizada et al., 2024). Regardless of the type of camera, since all cameras trade speed for resolution, it is likely that motion can become temporally aliased, spatially aliased, or worse still, both. Hence, methods to address the potential temporal aliasing problem are first discussed, followed by methods to measure small and subpixel motion that can address the potential spatial aliasing problem.

3.1.1 Addressing Temporal Aliasing

There are three different reported methods to address the temporal aliasing problem. All assume that the video was recorded with the per-pixel resolution being small enough such that the motion was spatially well-resolved. The first method used the folding properties of aliased signals (Law et al., 2022; Lambora et al., 2022). The second used the compressed sensing technique (Rajput and Law, 2024), and the third method used deep learning to up-sample aliased video to rates that respect the Nyquist criterion (Rajput et al., 2023).

The method that used the folding properties of aliased signals found the actual natural frequencies by sampling at two different and fractionally uncorrelated sub-Nyquist rates and using notions of set theory to find intersections between different folded signals. The method was tested for its robustness to noise and was found satisfactory when the signal-to-noise ratio was at least 10. The method was found suit-

able to identify multiple modes. The method also addressed the issue of observations within folds not being unique by recommending additional sampling at rates that were further fractionally uncorrelated to the first two sets. Extensive numerical experiments were performed, and those were used to instruct video recordings of end mills to find true natural frequencies of the end mill from the aliased video. Though the method works, since it requires additional measurements and/or requires down-sampling of the original video at fractionally uncorrelated rates to the original signal, the method is cumbersome.

The second method leveraged developments in the compressed sensing technique (Donoho, 2006; Candes et al., 2006). Compressed sensing enables non-uniform random sampling at sub-Nyquist rates and leverages sparse structures of signals to allow for exact recovery of signals that are not aliased. Since compressed sensing required video to be randomly sampled at the time of acquisition, and since existing camera hardware does not allow for this yet, the method demonstrated modal parameter recovery from motion registered from video that was randomly down-sampled at non-uniform rates from video that was originally properly and uniformly sampled (Rajput and Law, 2024). It was shown that for a machine tool system characterized by a single mode, recovery was exact with as few as $\sim 11\%$ randomly sampled frames of the originally sized video, and that for a system with two modes, $\sim 40\%$ randomly sampled frames of the original video were necessary.

The third method that used deep learning works by generating an arbitrary number of intermediate frames between two consecutively recorded frames from video that may be spatiotemporally aliased to form spatially and temporally coherent video sequences, which estimates motion from that interpolated video. The method was based on the Super SloMo scheme, which is based on a convolution neural network (CNN) architecture (Jiang et al., 2018; Paliwal, 2018) to learn optical flow from a pair of images. Generating intermediate frames between two frames is premised on the assumption that if a pixel in the frame being generated at any time instant is visible, it is most likely also visible in at least one of the two input image pairs that were used to generate the intermediate frame. And since all pixels in every frame of a video sequence are virtual sensors, the notion of a change in light intensity of one pixel from one frame to the next was used to estimate the motion of the artifact being recorded. And, since it is likely that motion from up-sampled video may contain frequency content other than the natural frequencies of vibration, new and simple ways to deduce the real modes from observations from several up-sampled videos were also presented. The method was illustrated with aliased video of end mills and was found to work satisfactorily.

3.1.2 Addressing Spatial Aliasing

Several fixes exist to avoid spatial aliasing when registering small, subpixel-level motion of vibrating cutting tools. These include software-based image processing schemes as well as hardware-based ones.

Within the family of software-based schemes, the simplest is to use intensity- and phase-based optical flow schemes (Nuhman et al., 2022) that are inherently capable of registering subpixel motion. The per-pixel resolution of these methods is governed by the camera's intensity depth in bits (Javh et al., 2017). For example, with an 8-bit intensity depth camera, the pixel intensity can take any value between 0 and 255 ($2^8 = 256$) to result in $\approx 0.004(1/255)$ pixel-displacement resolution. Likewise, with a 12-bit intensity depth camera, the pixel intensity can take any value between 0 and 4095 ($2^{12} = 4096$) to result in $\approx 2.4 \times 10^{-4}$ pixel-displacement resolution. And, since the pixel size is typically of the order of tens of μm , the pixel resolution becomes sub- μm .

Aside from the optical flow-based schemes, small and subpixel-level motion has also been reported to be estimated using the simple edge detection and tracking scheme. Even though the edge method 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 can sometimes register motion smaller than the pixel size (Gupta et al., 2021).

In addition to classical image processing schemes, though the DIC- and PIV-based schemes are capable of estimating small and subpixel-level motion, these schemes fare better when there are markers that can be correlated. Also, since most research on vision-based modal analysis has not used markers, those schemes did not fare well (Gupta et al., 2021; Raizada et al., 2024).

Recent work has also shown that deep learning based on the use of CNNs is capable of estimating motion that may potentially be spatially aliased (Rajput et al., 2023; Raizada et al., 2024). Interestingly, small subpixel-level motion was estimated more easily when the CNN was trained on a publicly available dataset of flying inanimate objects (Raizada et al., 2024), as opposed to training the CNN on videos of tools with small subpixel-level motion (Raizada and Law, 2023).

Other than software-based methods, the other method to avoid the issue of spatial aliasing is to use proper lenses and camera accessories to increase magnification to make the per-pixel size smaller than the expected amplitude of motion, as has been done and reported in Nuhman et al. (2022). In this hardware-based method, clever combinations of extension tubes and spacers between the camera and a reverse-mounted lens help reduce the pixel size. By moving the lens away from the body, the extension tube adjusts the focal point and allows focusing on subjects closer to the camera, thus magnifying the image. Also, since these tubes contain no optical elements, they do not distort the image. Further magnification was achieved by simply reversing the lens, which reverses the optics and results in a change in focal length. Though there is a loss of depth of field with these setups, and though less light falls on the image sensor, the hardware fixes were shown capable of bringing the pixel size down from $83 \mu\text{m}$, obtained with a simple 18–55 mm lens, to $1.3 \mu\text{m}$ with the use of three extension tubes placed in between the camera and the same reverse-mounted lens. A summary of the magnification potential with different combinations of tubes and lenses is reproduced from Nuhman et al. (2022) in Fig. 3.

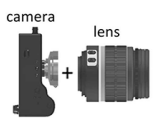


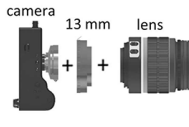


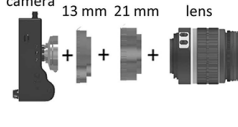

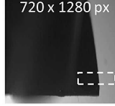
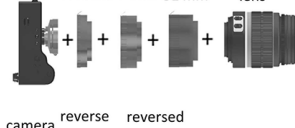

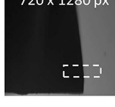
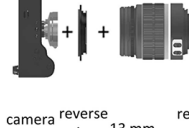


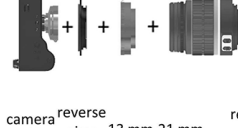


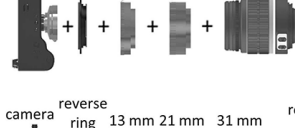


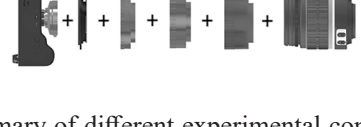


Configuration	Schematic of the setup	Tool	Distance between lens and tool in mm	Intensity of DC light in %	Field of view	Size of the pixel in μm
Camera with 18 - 55 mm lens			115	19		83
Camera with 18 - 55 mm lens and a 13 mm extension tube			55	25		16.4
Camera with 18 - 55 mm lens and a 13 and a 21 mm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

FIG. 3: Summary of different experimental configurations using extension tubes and a reverse mounted lens (reprinted from Nuhman et al. with permission from Elsevier, copyright 2022)

3.2 Detailed Overview of Image Processing Schemes

This section reviews the different image processing schemes used to register the motion of machine tool systems. Discussions include procedures for each, as well as comments on the (de)merits of each scheme.

3.2.1 Edge Detection and Tracking

Of the 16 papers reviewed on vision-based vibration monitoring, 10 are preferred edge detection and tracking schemes. This is likely due to this being amongst the simplest techniques to estimate motion from video. Though this is the preferred method, different groups have used different methods to register edges. Methods can be broadly classified into two types based on the order of derivative used to obtain the gradient. In the search-based types, a first-order derivative is used to compute gradients, and in the zero-crossing method, a second-order derivative is used to compute gradients.

Search-based methods use kernels convolved with the pixel intensities to estimate intensity gradient magnitudes and directions to signal the presence of edges. These are signaled when predefined thresholds for intensities are reached. Usually used kernels are of the Roberts' type, the Sobel type, or the Prewitt type (Gupta et al., 2021). Differences between these are due to their different sizes and the different weights assigned to every element with the kernel. Due to these differences, different kernels fare differently in correctly estimating edges. The choice of appropriate thresholds also changes with different kernels. Since lower thresholds detect more edges and are susceptible to noise, whereas higher thresholds may miss subtle edges, the choice of a single threshold is non-trivial. As such, the preferred edge detection scheme has become the Canny edge detector (Law et al., 2020, 2022; Gupta et al., 2021; Gupta and Law, 2021; Lambora et al., 2022; Raizada et al., 2024; Huang et al., 2022) in which thresholds are used to bind the magnitude of detected edges between upper and lower bounds. The Canny edge detection algorithm also smoothens the image by convolving with a Gaussian filter, thus minimizing the influence of noise on the registered motion.

Different from the search-based methods, the zero-crossing method, on the other hand, involves evaluating the second-order derivative of the image intensity to find regions of zero-crossings that are deemed to be the edges of interest. Within the family of these zero-crossing methods, the only method used is the Laplacian of Gaussian (LoG) method (Gupta et al., 2021; Huang et al., 2022).

Once the edge in every frame is detected, since the edges of interest span several pixels, these are averaged in every frame to result in a scalar value for that frame. That value is stored in an array. The process is repeated for all frames to result in pixel-displacement time series data. This pixel-displacement time series data is then easily converted to response in physical coordinates with the knowledge of the size of the pixel in μm .

An overview of the implementation of the edge detection schemes is shown in Fig. 4, which is borrowed from Gupta et al. (2021). Since that research registered the motion of three different cutting tools, Fig. 4 naturally includes those details. The figure shows the five key steps in registering motion, which include: (1) cropping images to retain regions of interest; (2) detecting the edge in every frame; (3) mapping the pixels to displacements; (4) averaging pixels in every frame; (5) resulting response from video.

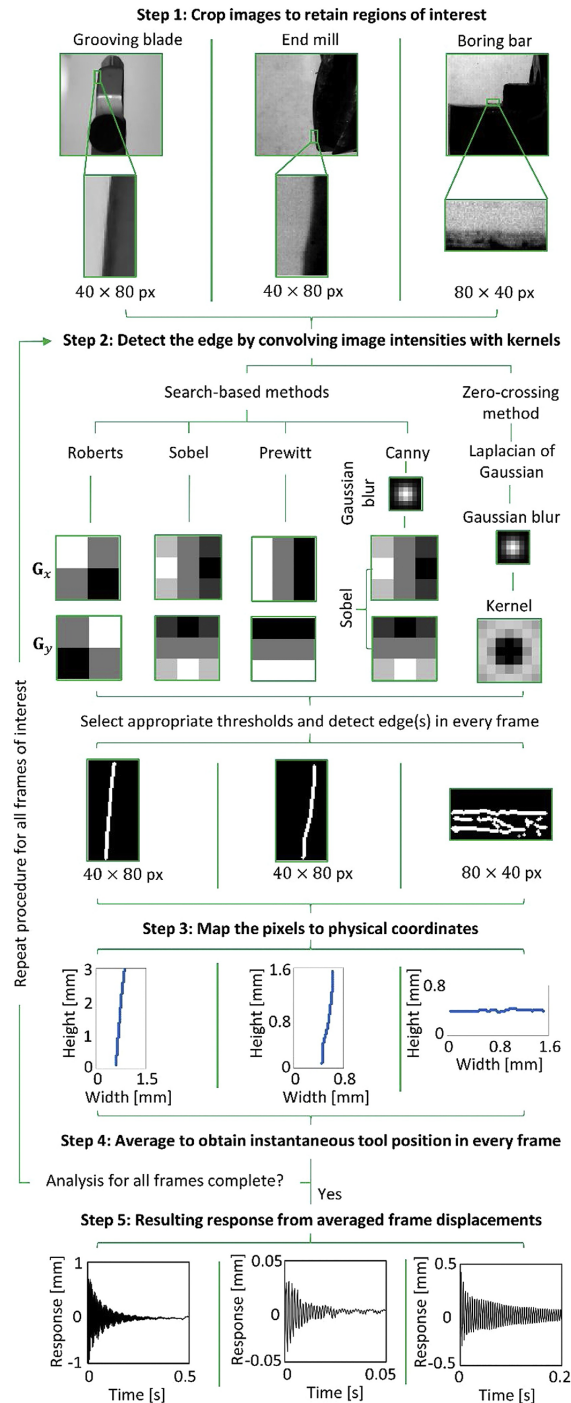


FIG. 4: Overview of edge detection and tracking schemes (reprinted from Gupta et al. with permission from Elsevier, copyright 2021)

The edge detection and tracking scheme remains amongst the simpler techniques for registering tool motion from video. The method, however, is sensitive to thresholds to the size of the Gaussian blur and is also not always a sub-pixel scheme.

3.2.2 Optical Flow: Intensity and Phase-Based

The reported use of optical flow for motion registration of machine tool systems has been less than that of edge-based methods. Only three of the 16 relevant research works have reported using optical flow-based methods (Gupta et al., 2021; Nuhman et al., 2022; Raizada et al., 2024). This is despite the fact that optical flow-based methods are inherently subpixel motion registration capable. Within the family of optical flow methods, the use of intensity-based optical flow and phase-based optical flow have both been reported.

The intensity-based optical flow method is premised on the assumption that there is small motion of the system in a short time interval. As such, the method assumes brightness constancy for every point and pixel within the frame, i.e., the intensity of light for a point on the tool that moves slightly in a short time to another location within the frame is assumed to remain constant. With this assumption, and the use of a Sobel kernel to compute gradients, the method estimates a motion matrix from which a scalar displacement can be computed by averaging the motion matrix. Since the gradient-based optical flow scheme is sensitive to image noise, in its implementation for the estimation motion of machine tool systems (Gupta et al., 2021; Nuhman et al., 2022; Raizada et al., 2024), researchers applied Gaussian blurs like in the case of the Canny edge detector. Moreover, since cutting tool motion was also observed to be sensitive to the regions of interest being evaluated, an expanded implementation was included using an edge detector in combination with optical flow to evaluate the motion of the vibrating tool (Gupta et al., 2021). An overview of the method to estimate motion using the intensity-based optical flow scheme is shown in Fig. 5.

The figure shows the four key steps in registering motion, which include: (1) cropping images to retain regions of interest; (2) convolving a Sobel kernel with the pair of frames of interest; (3) estimating motion by either averaging the motion matrix or by averaging after convolving the motion matrix with an edge detector; (4) repeating steps 1–3 for all frames to obtain response from video. Though this method is more robust than the edge detection and tracking scheme, it is sensitive to the size of the Gaussian blur, the choice of the kernel used to compute image intensity gradients, and the method of averaging full-field motion.

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. No brightness constancy is assumed and neither is the method confined to assuming small motion in small time intervals. The key step in the method involves convolving the image with a complex kernel to obtain local phase and amplitude information to decipher changes in the local phase to track motion over consecutive frames. The complex kernel is an oriented Gabor filter that can be thought of as a sinusoidal signal of a particular fre-

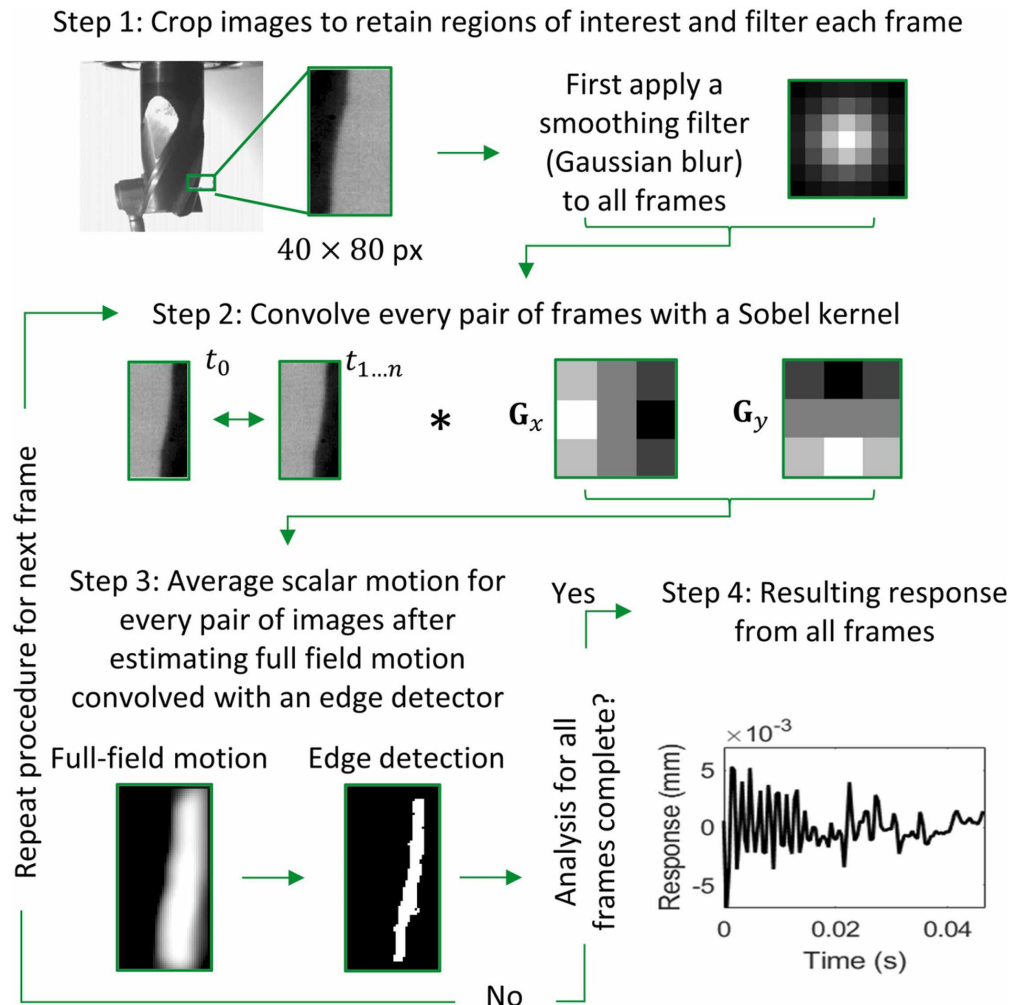


FIG. 5: Overview of the intensity-based optical flow scheme (reprinted from Nuhman et al. with permission from Elsevier, copyright 2022)

quency and orientation, modulated by a Gaussian wave. The parameters that control its shape, size, and performance are its orientation, its wavelength, and its bandwidth. For the Gabor filter to work effectively, it must be tuned for the image sequence of interest. Velocities of the local phase contours thus obtained are integrated to obtain displacements. Since only displacements in regions with sufficient contrast are treated as reliable, implementation also includes thresholding. Also, since the phase-based optical flow is more robust and more accurate, implementation for use in machine tool systems (Nuhman et al., 2022) has preferred filtering every image with a Gaussian blur prior to minimizing the influence of noise. This is done prior to convolving frames with the Gabor filter.

An overview of the five key steps to register machine tool motion using the phase-based scheme is shown in Fig. 6, which is borrowed from Nuhman et al. (2022). The steps are: (1) cropping images to retain regions of interest and applying a Gaussian blur; (2) convolving a Gabor filter to estimate phase contours; (3) convolving phase with a

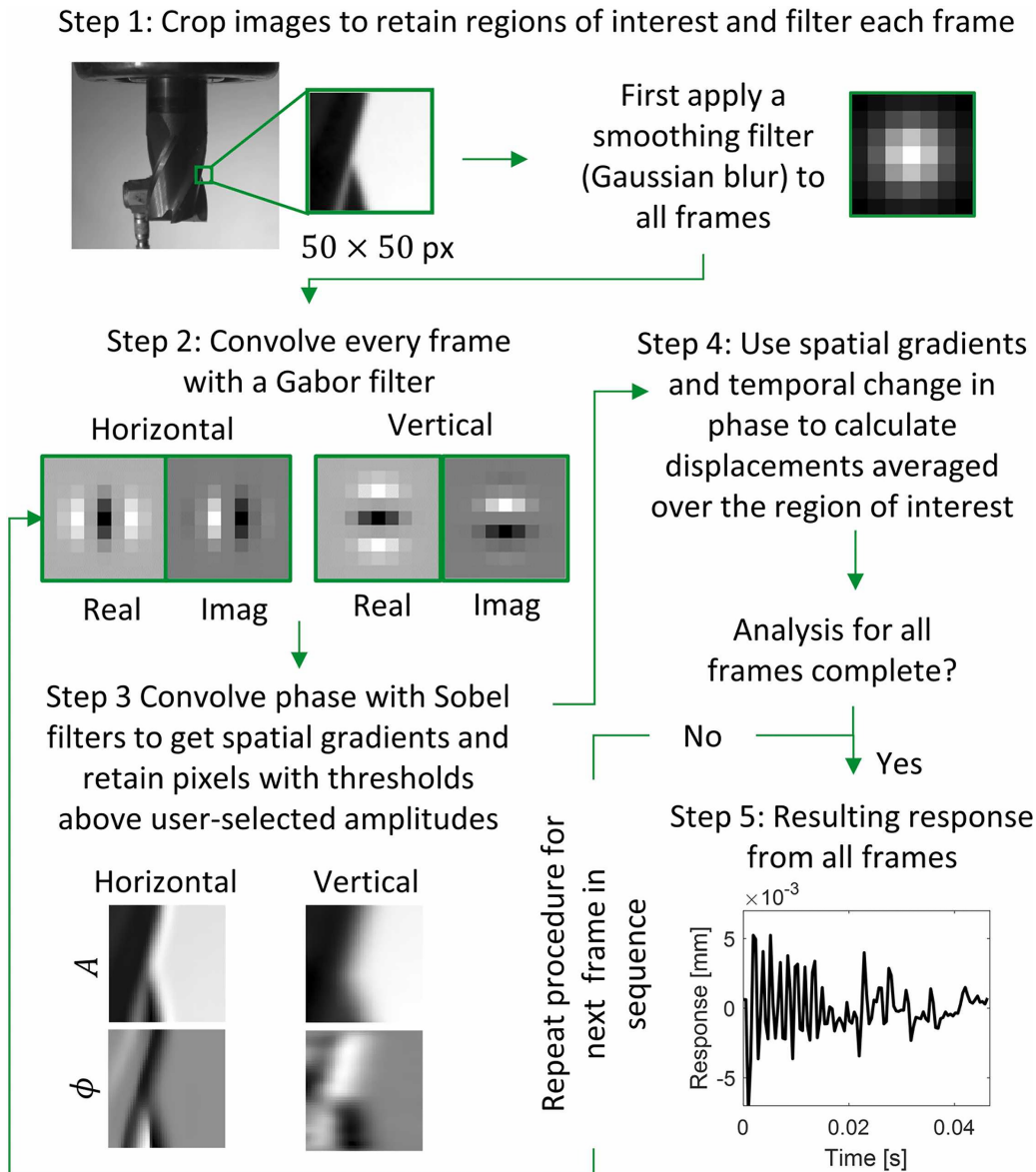


FIG. 6: Overview of the phase-based optical flow scheme (reprinted from Nuhman et al. with permission from Elsevier, copyright 2022)

Sobel kernel and use thresholds; (4) estimating displacements; (5) repeating steps 1–4 for all frames to obtain a response from the video.

3.2.3 Digital Image Correlation

The digital image correlation (DIC) method has been used in two reports by Gupta et al. (2021) and Raizada et al. (2024). The DIC method, like the optical flow-based schemes, is also a subpixel-level motion registration capable scheme. It estimates motion by finding the maximum correlation between pixel subsets in the image of interest and the reference image. For the subset window, the region around the vibrating tool's edges is chosen since that is the region of high contrast, which makes tracking easier. Implementation in machine tool systems is correlated to the tool's own features and not random speckle patterns sprayed on the tool, which is what is usually recommended with DIC. To deal with the noise, every frame was also convolved with a Gaussian kernel.

An overview of the DIC procedure to estimate motion is shown in Fig. 7. The figure is borrowed from Gupta et al. (2021), and as such, includes details from that study. Since implementation by Gupta et al. (2021) was with their own code and not a commercially available code, the procedure to register motion may differ from commercial software packages. There are three main steps. The first involves cropping each frame around the region of interest, followed by a smoothing operation. In the second and main step, a correlation function is evaluated for every subset within every pair of frames, and that is then used to estimate rigid body displacement of the tool between those frames. The third step involves repeating step two for all pairs of frames from the video to finally result in the response of the vibrating artifact.

3.2.4 Particle Image Velocimetry

Particle image velocimetry (PIV) is usually used to measure 2D velocity fields in macroscopic fluid flow by imaging the flow seeded with particles followed by cross-correlating groups of particles, like in DIC, to obtain average displacements (Yamamoto et al., 2017). Since PIV can estimate velocity fields, and since extracting displacements from those velocities is trivial, PIV was also leveraged to estimate displacements of vibrating cutting tools—as reported in Raizada et al. (2024). That report is the only one. In that study, the tool's own features were correlated across frames to estimate displacements. In that sense, the implementation was like that of the DIC scheme. And like DIC, PIV requires a speckle pattern to be sprayed on the object to be tracked to ensure a high degree of correlation is obtainable. However, since that is not possible, the challenges associated with DIC are applicable to PIV-based motion registration too.

As opposed to the case of DIC, in which the researchers wrote their own code (Gupta et al., 2021), the PIV study used PIVLab, an easy to use, GUI-based MATLAB tool to analyze, validate, post process, visualize, and simulate PIV data (Thielicke and Stamhuis, 2014). PIVLab provides nice inbuilt features for image pre-processing and process parameter selection. The Fourier transform correlation with multiple passes was

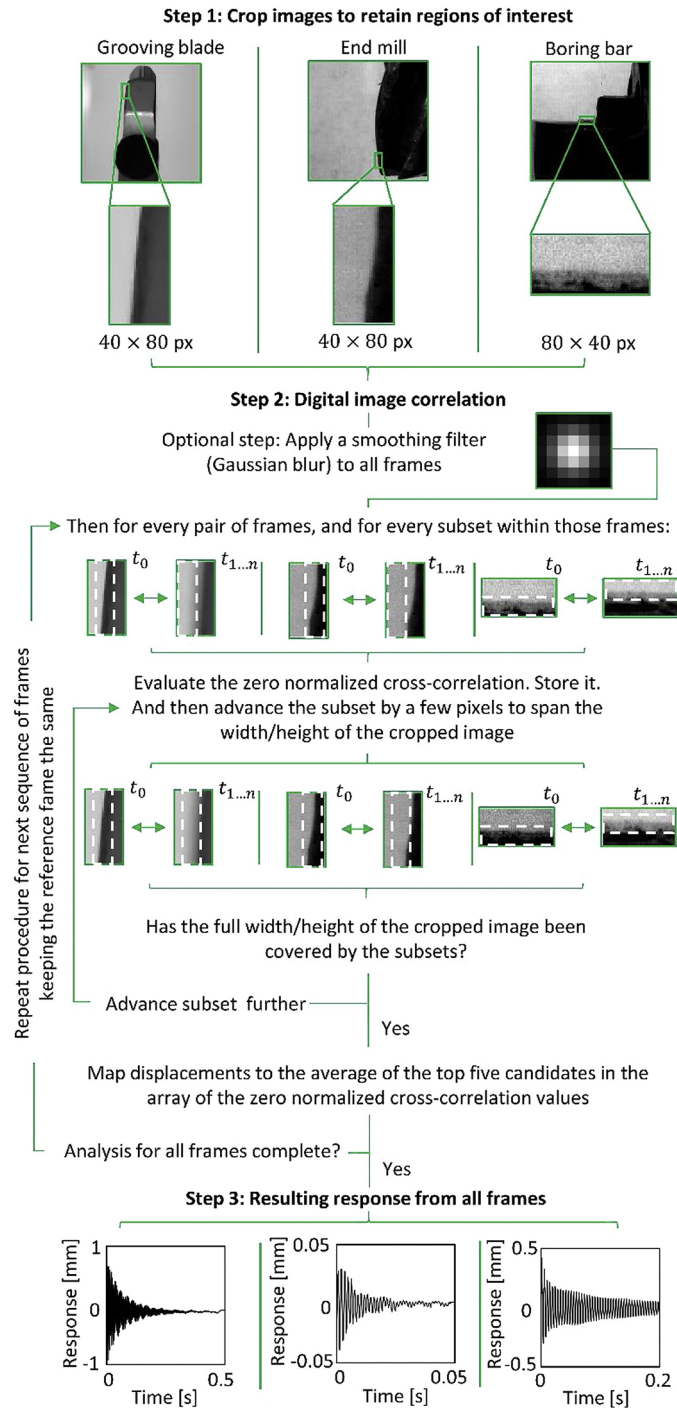


FIG. 7: Overview of the digital image processing scheme (reprinted from Gupta et al. with permission from Elsevier, copyright 2021)

used to estimate motion with PIVLab. The use of multiple passes is said to make motion estimation more robust than the case of DIC. The documentation for PIVLab (Thielicke and Stamhuis, 2014) is robust, and an overview of its use and implementation in the context of estimating motion of vibrating machine tool elements would essentially reduce to providing screenshots of the GUI. This will not add to the scientific discourse of this paper; an overview such as those in Figs. 4–7 is not provided herein, and the reader is instead directed to the help manuals and documentation of PIVLab for its use.

3.2.5 Deep Learning

There are three reports of using deep learning methods to register motion of vibrating machine tool systems. The first such report used the robust and open-source OpenCV computer vision tool (Yanamadala and Muralidharan, 2021). In the use of OpenCV, the researchers investigated the influence of different algorithms for object tracking. Eight such algorithms were tried, and the CSRT algorithm was reported to fare better than the others. Since OpenCV was developed for real-time computer vision applications, the researchers reported that OpenCV was well-suited for tracking motion, even in the presence of noise.

The other two reports used convolutional neural networks (CNN) (Raizada and Law, 2023; Raizada et al., 2024). The CNNs have seen wide-scale adoption across engineering domains for classification, segmentation, and recognition. Unlike classic image processing schemes, CNNs learn what filters to use to estimate flow by training itself on a dataset. In one study using CNNs, the researchers trained the model with videos of a tool vibrating with potentially small and subpixel-level motion, whose motion was recorded in laboratory settings (Raizada and Law, 2023). In the other use of CNNs, open-source data sets were used to train the model (Raizada et al., 2024). The two implementations also differed in their choice of CNN models.

In the implementation in which data was generated in the laboratory, the FlowNet model, (Dosovitskiy et al., 2015) which has become the de facto CNN model, was used. The FlowNet model is a supervised learning model whose model architecture comprises convolution and deconvolution layers with residual connections. The model was trained using videos of a tool vibrating with potentially small and subpixel-level motion. Ground truths for training from these videos were generated using the phase-based optical flow algorithm, which can provide an accurate displacement estimation even for subpixel motion. The method was reported to satisfactorily estimate natural frequencies of vibration. However, there were issues in the estimates of the amplitude of motion, and those errors led to errors in estimating damping.

The other method using CNNs to extract small and subpixel level motion from video used the iterative residual refinement (IRR) scheme (Hur and Roth, 2019), is an improvement over the PWC-Net scheme (Sun et al., 2018) and builds on the standard FlowNet scheme (Dosovitskiy et al., 2015). The IRR scheme has been designed according to simple and well-established principles of an encoder, warping, use of a cost volume, and a decoder. Implementation of the scheme proceeds as outlined in Fig. 8.

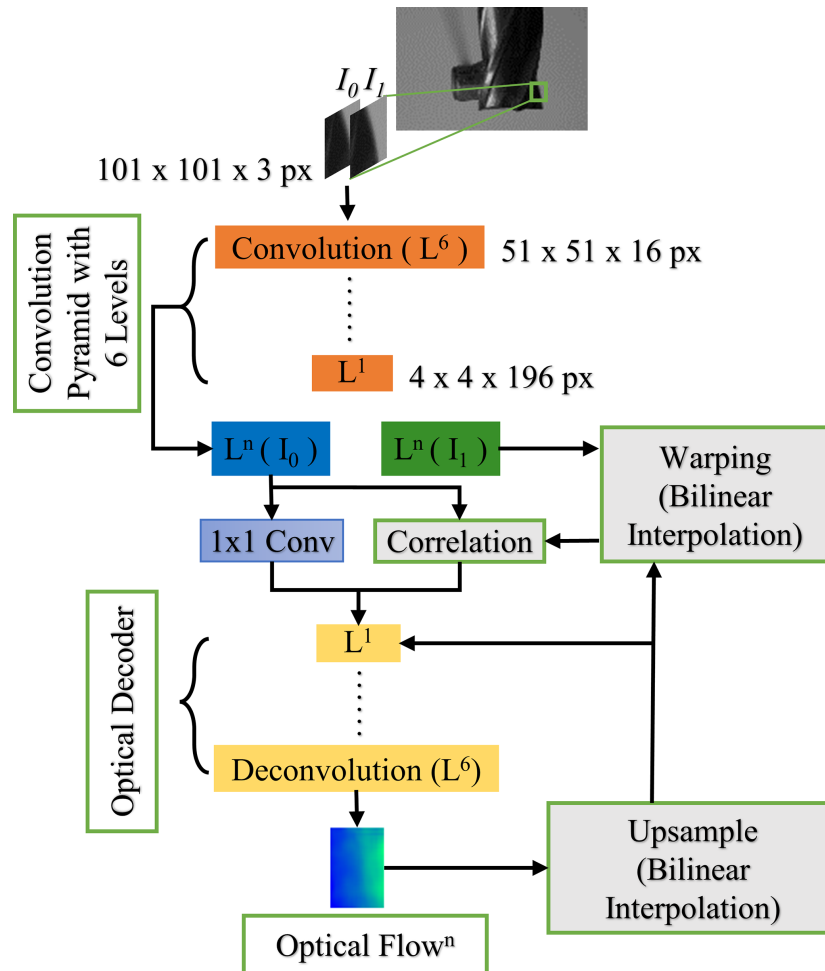


FIG. 8: Overview of the iterative residual refinement deep-learning scheme. Figure reproduced from (Raizada et al.) under a Creative Commons license.

First, a pair of images is passed into the encoder, which is a 6-level convolution pyramid. At each level, filters and kernels are applied to the images for feature extraction. These kernels form the learnable unit. Output from the sixth level of the encoder (level L^1) is passed directly to the decoder, which is also a 6-level deconvolution pyramid for optical flow estimation and is then used in the next level. In the implementation reported in Raizada et al. (2024), the input image was $101 \times 101 \times 3$ pixels. This was encoded to result in an image size of $4 \times 4 \times 196$ pixels, which was again decoded to result in its original size. The optical flow thus obtained was then up-sampled using bilinear interpolation. That up-sampled flow became the current optical flow for that level to warp the features of the second image using bilinear interpolation. The warped features from the second image, combined with features of the first image, were then used to construct

a cost volume using a correlation function. That then became the input to the optical decoder along with the first image and current optical flow from the previous level, resulting in a more refined optical flow. The original PWC-Net incrementally updated the estimation across the pyramid levels with individual decoders for each level. The IRR scheme substituted the multiple decoders with only one shared decoder that iteratively refined the output over all the pyramid levels. Additional details are in Raizada et al. (2024).

In this implementation, the model was trained on a publicly available dataset called FlyingThings3D. Testing was against the ground truth, which in this case was twice integrated accelerations. Besides needing large data for training, the method was found to be robust and capable of predicting small subpixel-level motion.

4. REPRESENTATIVE RESULTS

Since different research groups have used different methods of motion registration and have measured different artifacts, a direct comparison of the different methods is not possible. Hence, as representative cases, the motion of a vibrating end mill for a setup, like in Fig. 2(a), is shown in Fig. 9, and the motion of a vibrating grooving blade for the setup shown in Fig. 2(e) is shown in Fig. 10. These results are borrowed from Raizada et al. (2024). The response shown in Figs. 9 and 10 was obtained using six different image processing schemes. These include the Canny edge detector, intensity- and phase-based optical flow-based schemes, DIC, PIV, and the IRR CNN. Results were benchmarked with twice-integrated accelerations. A normalized root mean square error (NRMSE) was used to quantify differences between motion estimated from vision and twice-integrated accelerations. Since data was sampled at different rates, to estimate the NRMSE, response from the video was up-sampled to the ground truth data acquisition rate using spline interpolation.

From Fig. 9 and the NRMSE shown therein, it is evident that the response obtained with the DIC method is not as well resolved as the other five methods. The PIV method fares best in this case and is closely followed by the other four methods. Though PIV uses the same principle as DIC, lower NRMSE with it is thought to be due to PIVLab handling image noise better than their own implementation of the DIC method. Moreover, the results are likely also better because PIVLab automatically suggests the interrogation window size and allows the user to select multiple interrogation windows in multiple passes. This is different than their (Raizada et al., 2024) implementation of DIC in which only one subset size in two frames is correlated. What is further evident is that for all methods, motion of the order of $\pm 54 \mu\text{m}$ that is averaged over the first four cycles is easily resolved, even when the pixel size is $36 \mu\text{m}$. However, since there is damping in the system, the response decays and becomes magnitudes lesser than the pixel size. Also, since not all methods are good at subpixel level motion registration, the methods that are fare better.

From Fig. 10 it is observable that the CNN-based method fares best in this case, followed by the edge and the PIV schemes, respectively. The DIC method fares better in

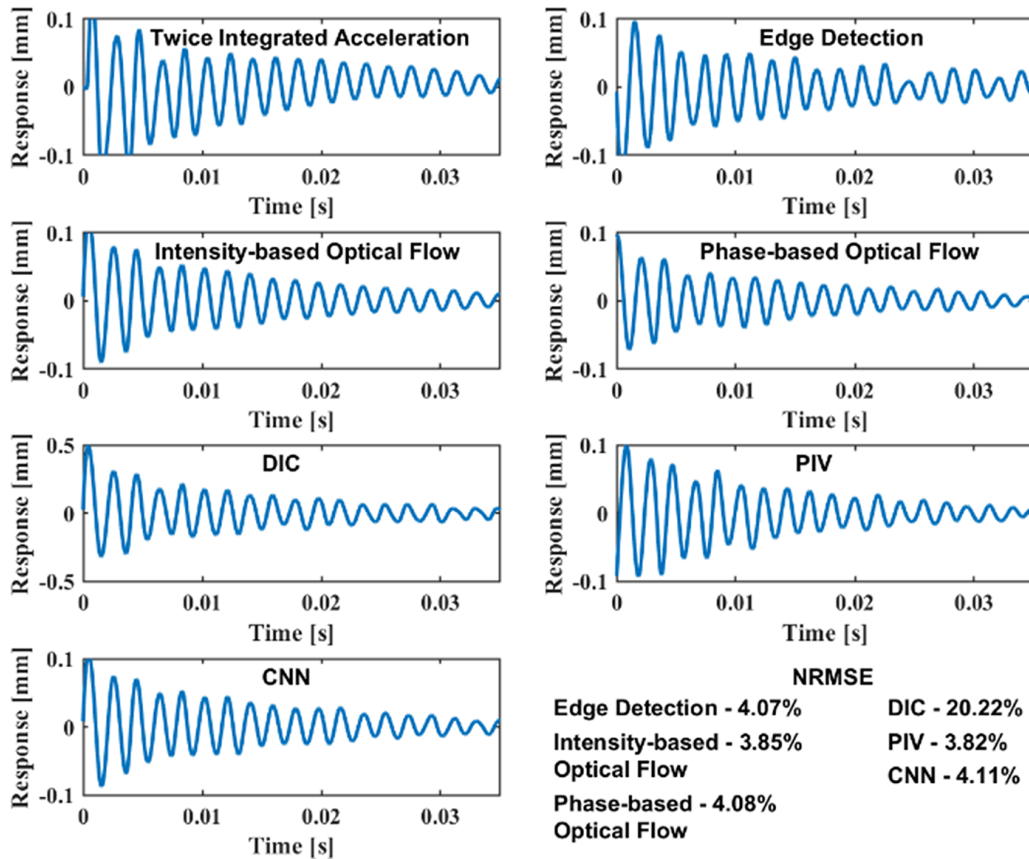


FIG. 9: End mill's motion estimated from the six image registration schemes being benchmarked with ground truth in terms of the NRMSE. Figure reproduced from (Raizada et al.) under a Creative Commons license.

this case than in the case of the end mill, and this is thought to be due to the blade vibrating with larger magnitudes than the end mill. It is thought to be due to the blade having different features and having been acquired with different parameters. The response obtained with the intensity-based optical flow method has the largest NRMSE, followed by the phase-based optical flow method.

These differences between results for the end mill and the blade suggest that the registered motion is sensitive to the magnitude of motion, to the image processing scheme being used, tool features, and to image acquisition parameters. However, since the edge detection scheme, the PIV scheme, and the CNN scheme consistently give low NRMSE, it may be concluded that these schemes are more robust.

Since the ultimate objective of registering motion from video is to extract modal parameters from that video, modal parameters for the response that is shown in Figs. 9 and 10 are listed in Table 2, which is also borrowed from Raizada et al. (2024). Parameters were extracted using the eigensystem realization algorithm. From Table 2, it is evident

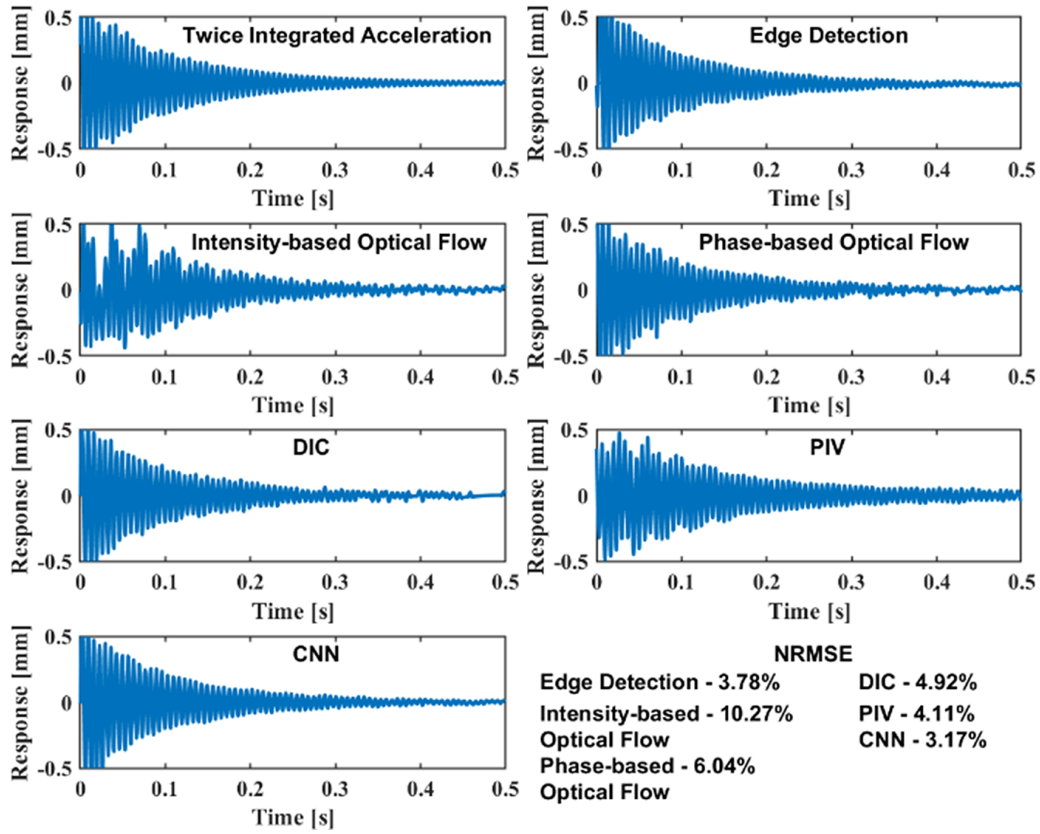


FIG. 10: Grooving blade’s motion estimated from the six image registration schemes being benchmarked with ground truth in terms of the NRMSE. Figure reproduced from (Raizada et al.) under a Creative Commons license.

TABLE 2: Modal parameters for the response shown in Figs. 9 and 10 (adapted from Raizada et al., 2024)

Method	Modal parameters			
	End mill		Grooving blade	
	f [Hz]	ζ [%]	f [Hz]	ζ [%]
Twice integrated accelerations	523	1.80	151	1.08
Edge detection	524	1.91	150	0.98
Intensity-based optical flow	524	1.96	151	0.62
Phase-based optical flow	524	2.05	150	1.03
DIC	523	2.12	150	0.93
PIV	524	2.30	151	0.60
CNN	523	1.88	150	1.00

that all six vision-based schemes could estimate the natural frequencies correctly. However, there are differences in the estimated damping ratios (ζ). Since correct damping is a function of the response being captured accurately, those schemes with larger NMRSE also exhibit higher errors in damping estimates. Since the research done by Raizada et al. (2024) did not measure the input force, that was a case of output-only modal analysis, and hence, the eigenvector was not estimated or reported in Table 2.

Results in Figs. 9 and 10 and Table 2 are representative only. Similar comparisons between edge detection and tracking methods and other deep learning-based tracking methods implemented in OpenCV are available in Yanamadala and Muralidharan, (2021). Likewise, comparisons between the different methods of edge detection are available in Gupta et al. (2021). Those comparisons observed the Canny edge detector to fare best. However, for a different vibrating artifact, Huang et al. (2022) reported that the LoG method fares better than the Canny edge detector.

It is also important to highlight here that aside from the deep learning-based scheme, which may require training and testing which takes time, none of the other methods of motion registration are computationally heavy. Once the video is provided, the time that most schemes take to register motion is of the order of hundreds of milliseconds—for the representative cases discussed in Figs. 9 and 10. These times were for code that was written for research purposes and running on a desktop computer with 16 GB RAM and an i7-8550U CPU @ 1.80 GHz processor. The code was not optimized to make it computationally efficient.

5. CHALLENGES AND PROSPECTS

Having outlined the research progress on vision-based machine tool vibratory motion registration, this section discusses some challenges and future prospects of the method. Discussion on challenges include robustness of motion registration and stereo measurements for simultaneous in- and out-of-plane motion registration. Discussions on the prospects of the method include developments in cameras.

5.1 Challenges

Since motion registered from video is sensitive to image acquisition parameters, to features of the vibrating artifact, to the magnitude of motion, and to the image processing scheme being used, and since the use of vision for vibratory motion registration is still in its infancy, it is difficult to comment on which method is more robust than the others. Only wider adoption of the method to measure different machine tool elements will help hone the method. Moreover, the method has only been tested in laboratory settings. Suitability of the method to measure machinery in shop floor settings within industries also needs to be evaluated. These evaluations must include characterizing the influence of lighting conditions and the presence of coolant, mist, and chips within the machine's region that may distort acquisition. Again, only wider adoption and use of the method will help allay these questions.

Furthermore, there are several other settings that influence measurements, such as the field of view, the sensor size, the aperture, the ISO settings, the analog and digital gains, the magnification, the choice of lens, and other camera-specific settings. Some of these settings need to be adjusted for high-speed recordings, and some other settings need to be tuned for high-resolution recordings. Since some of these settings are confounded with others, a systematic investigation on how image acquisition settings and parameters other than speed and resolution affect measurements is necessary and can be taken up in future investigations.

Another issue about the use of the method is that all reported research on vision-based modal analysis has been restricted to measuring in-plane response only. Since the out-of-plane response is also of interest in machine tool systems, the vision setups need to be changed/oriented/indexed to measure that out-of-plane response. This is much like what also needs to be done in traditional modal analysis procedures that use contact-type single-axis accelerometers. However, for 2D/3D vision-based modal analysis, it is possible to leverage the developments in stereo vision across other disciplines for simultaneous in- and out-of-plane measurements using two or more cameras. The challenges pertaining to using stereo will be to correct for potential distortions due to projections, to ensure stereo matching, to synchronize cameras to each other, and to the input. Since these issues are uncharted, there is tremendous potential for the method.

5.2 Prospects

Rapid developments in camera hardware, especially that of smartphone cameras that can already record video at rates of up to 7680 Hz and with impressive resolutions of 720 px × 720 px (Huawei, 2023), will make equipment for video recording more accessible. Since most image processing schemes to register motion are relatively lightweight, these can be easily integrated within an application on smartphones to make every smartphone a vibration measurement device. Furthermore, since the main time-taking task in visual vibrometry is to record and transfer high-speed and high-resolution video recordings that could be a few gigabytes, and since the use of smartphones would not require the transfer of such large files, it might even become possible to make smartphone aided visual vibrometry a real-time vibration measurement process. This will require the development of dedicated applications, which enterprising engineers well-versed with structural dynamics, image processing, and code could help realize. Such developments are expected to democratize visual vibrometry.

Allied developments in compressed sensing hardware, including developments that will permit non-uniform random acquisition of images at sub-Nyquist rates with cameras, will also make it possible to address the temporal aliasing problem. Such developments, however, are likely to reap fewer benefits than developments in smartphone cameras.

Another revolutionary development that can cause another paradigm shift in vision-based vibration measurements is the advent of event cameras (Event camera, 2023). Event cameras do not capture images using a shutter as conventional (frame) cameras do. Instead, each pixel inside an event camera operates independently and asynchronously, reporting

changes in brightness as they occur, and staying silent otherwise. Since only those pixels whose intensity change are tracked and stored, recording at higher frame rates becomes possible without adversely increasing storage (memory) requirements. Such cameras are already commercially available and can record at rates of up to 1,00,000 Hz and with impressive resolutions. Since there is no research yet that reports on the use of such event cameras for visual vibrometry of machine tools, there are opportunities aplenty.

Besides developments in camera hardware, commercial solutions such as those offered by RDI Technologies (RDI Tech., 2023) have made possible vision-based modal analysis and vibration-based condition monitoring by leveraging developments in image processing-based motion registration and magnification. Such solutions will enable the modal analysis community to adopt vision-methods as a working and more efficient method of modal analysis.

Furthermore, since vision-based measurements are non-contact, they allow for measuring speed-dependent changes in the dynamics of rotating elements of machine tools. Vision-based methods could also be used for *in situ* measurements of the changing dynamics of thin walls as they are machined. Measuring speed- and time-dependent dynamic behavior using vision-based methods remains largely unexplored and offers researchers more opportunities.

6. CONCLUSIONS

The aim of vision-based modal analysis of machine tool systems is to use cameras to record the motion of vibrating elements of the machine tool system and apply image processing schemes to that video to extract motion from that video. The method is non-contact, does not require sophisticated data acquisition hardware, and makes possible full field shape analysis with just one measurement. As such, the method is fast emerging as a viable alternative to traditional experimental modal analysis procedures that use contact-type transducers. Since the method holds promise, this paper discussed the research progress and outlined the prospects.

A systematic review of the research revealed that different research groups have measured different elements with different cameras and acquisition parameters and have also used different methods to register motion from video. It is difficult to comment on which image processing scheme is more robust than the others. Given the developments in camera technology and the advent of new event cameras, together with the maturity of image processing schemes, and since the method holds much promise, a structured round-robin to measure with similar artifacts and similar image processing schemes of researchers active in the area, will likely resolve issues and help hone the method. This may also help uncover other fundamental issues that may be tackled by further research.

ACKNOWLEDGMENT

This work was supported by the Government of India's Science and Engineering Research Board's Core Research (Grant No. CRG/2020/001010).

REFERENCES

- Baqersad, J., Poozesh, P., Niezrecki, C., and Avitabile, P., Photogrammetry and Optical Methods in Structural Dynamics – A Review, *Mech. Syst. Signal Process.*, vol. **86**, pp. 17–34, 2017. DOI: 10.1016/j.ymsp.2016.02.011
- Beberriss, T.J. and Ehrhardt, D.A., High-Speed 3D Digital Image Correlation Vibration Measurement: Recent Advancements and Noted Limitations, *Mech. Syst. Signal Process.*, vol. **86**, pp. 35–48, 2017. DOI: 10.1016/j.ymsp.2016.04.014
- Berezvai, S., Bachrathy, D., and Stepan, G., High-Speed Camera Measurements in the Mechanical Analysis of Machining, *Proc. CIRP*, vol. **77**, pp. 155–58, 2018. DOI: 10.1016/j.procir.2018.08.264
- Brown, D.L. and Allemang, R.J., The Modern Era of Experimental Modal Analysis: One Historical Perspective, *Sound Vib.*, vol. **41**, no. 1, pp. 16–33, 2007.
- Candes, E.J., Romberg, J., and Tao, T., Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information, *IEEE Trans. Inf. Theory*, vol. **52**, no. 2, pp. 489–509, 2006. DOI: 10.1109/tit.2005.862083
- Czyżycki, J., Twardowski, P., and Znojkwicz, N., Analysis of the Displacement of Thin-Walled Workpiece Using a High-Speed Camera during Peripheral Milling of Aluminum Alloys, *Materials*, vol. **14**, no. 16, p. 4771, 2021. DOI: 10.3390/ma14164771
- Donoho, D.L., Compressed Sensing, *IEEE Trans. Inf. Theory*, vol. **52**, no. 4, pp. 1289–1306, 2006. DOI: 10.1109/TIT.2006.871582
- Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., van der Smagt, P., Cremers, D., and Brox, T., FlowNet: Learning Optical Flow with Convolutional Networks, *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Santiago, Chile, pp. 2758–2766, 2015. DOI: 10.1109/iccv.2015.316
- Dutta, S., Pal, S.K., Mukhopadhyay, S., and Sen, R., Application of Digital Image Processing in Tool Condition Monitoring: A Review, *CIRP J. Manuf. Sci. Technol.*, vol. **6**, no. 3, pp. 212–32, 2013. DOI: 10.1016/j.cirpj.2013.02.005
- Feng, D. and Feng, M.Q., Computer Vision for SHM of Civil Infrastructure: From Dynamic Response Measurement to Damage Detection – A Review, *Eng. Struct.*, vol. **156**, pp. 105–17, 2018. DOI: 10.1016/j.engstruct.2017.11.018
- Guo, Y., Dale Compton, W., and Chandrasekar, S., *In Situ* Analysis of Flow Dynamics and Deformation Fields in Cutting and Sliding of Metals, *Proc. R. Soc. A: Math. Phys. Eng. Sci.*, vol. **471**, no. 2178, pp. 20150194–94, 2015. DOI: 10.1098/rspa.2015.0194
- Gupta, P. and Law, M., Evaluating Tool Point Dynamics Using Smartphone-Based Visual Vibrometry, *Proc. CIRP*, vol. **101**, pp. 250–53, 2021. DOI: 10.1016/j.procir.2020.09.196
- Gupta, P., Rajput, H.S., and Law, M., Vision-Based Modal Analysis of Cutting Tools, *CIRP J. Manuf. Sci. Technol.*, vol. **32**, pp. 91–107, 2021. DOI: 10.1016/j.cirpj.2020.11.012
- GSMarena, Huawei Mate 30 Pro, accessed December 16, 2023, from https://www.gsmarena.com/huawei_mate_30_pro-9885.php, 2023.
- Huang, H., Kono, D., and Toyoura, M., Vision-Based Vibration Measurement of Machine Tool, *J. Adv. Mech. Des. Syst. Manuf.*, vol. **16**, no. 1, pp. JAMDSM0014–14, 2022. DOI: 10.1299/jamdsm.2022jamdsm0014
- Hur, J. and Roth, S., Iterative Residual Refinement for Joint Optical Flow and Occlusion Estimation, *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Long Beach, CA, pp. 5754–5763, 2019. DOI: 10.1109/CVPR.2019.00590
- Iglesias, A., Tunç, L., Özşahin, O., Franco, O., Muñoa, J., and Budak, E., Alternative Experimental Methods for Machine Tool Dynamics Identification: A Review, *Mech. Syst. Signal Process.*, vol. **170**, pp. 108837–37, 2022. DOI: 10.1016/j.ymsp.2022.108837
- Javh, J., Slavič, J., and Boltežar, M., The Subpixel Resolution of Optical-Flow-Based Modal Analysis, *Mech. Syst. Signal Process.*, vol. **88**, pp. 89–99, 2017. DOI: 10.1016/j.ymsp.2016.11.009

- Jiang, H., Sun, D., Jampani, V., Yang, M.-H., Learned-Miller, E., and Kautz, J., Super SloMo: High Quality Estimation of Multiple Intermediate Frames for Video Interpolation, *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018. DOI: 10.1109/cvpr.2018.00938
- Kurada, S. and Bradley, C., A Review of Machine Vision Sensors for Tool Condition Monitoring, *Comput. Ind.*, vol. **34**, no. 1, pp. 55–72, 1997. DOI: 10.1016/s0166-3615(96)00075-9
- Lambora, R., Nuhman, A.P., Law, M., and Mukhopadhyay, S., Recovering Cutting Tool Modal Parameters from Fractionally Uncorrelated and Potentially Aliased Signals, *CIRP J. Manuf. Sci. Technol.*, vol. **38**, pp. 414–26, 2022. DOI: 10.1016/j.cirpj.2022.05.014
- Law, M., Gupta, P., and Mukhopadhyay, S., Modal Analysis of Machine Tools Using Visual Vibrometry and Output-Only Methods, *CIRP Ann. - Manuf. Technol.*, vol. **69**, no. 1, pp. 357–360, 2020. DOI: 10.1016/j.cirp.2020.04.043
- Law, M., Lambora, R., Nuhman, A.P., and Mukhopadhyay, S., Modal Parameter Recovery Using Temporally Aliased Video Recordings of Cutting Tools, *CIRP Ann. - Manuf. Technol.*, vol. **71**, no. 1, pp. 329–332, 2022. DOI: 10.1016/j.cirp.2022.03.023
- Mori, K., Kono, D., and Matsubara, A., Vision-Based Volumetric Displacement Measurement with a Self-Illuminating Target, *CIRP Annals*, vol. **72**, no. 1, pp. 305–308, 2023. DOI: 10.1016/j.cirp.2023.04.059
- Nuhman, A.P., Singh, A., Lambora, R., and Law, M., Methods to Estimate Subpixel Level Small Motion from Video of Vibrating Cutting Tools, *CIRP J. Manuf. Sci. Technol.*, vol. **39**, pp. 175–184, 2022. DOI: 10.1016/j.cirpj.2022.08.005
- Okubo, N., Yoshida, Y., and Hoshi, T., Application of Modal Analysis to Machine Tool Structures, *CIRP Annals*, vol. **31**, no. 1, pp. 243–246, 1982. DOI: 10.1016/s0007-8506(07)63306-x
- Paliwal, A., Open-Source Pre-Trained Code for Super SloMo, accessed from <https://github.com/avinash-paliwal/Super-SloMo>, 2018.
- Raizada, V. and Law, M., Obtaining Subpixel Level Cutting Tool Displacements from Video Using a CNN Architecture, *Manuf. Technol. Today*, vol. **22**, no. 3, pp. 14–19, 2023. DOI: 10.58368/mtt.22.3.2023.14-19
- Raizada, V., Rajput, H.S., and Law, M., Comparative Analysis of Image Processing Schemes to Register Motion from Video of Vibrating Cutting Tools, *Procedia CIRP*, vol. **121**, pp. 85–90, 2024.
- Raizada, V., Singh, M., and Law, M., Estimating Material Deformation Characteristics during Orthogonal Cutting using Digital Image Correlation, *Select Proceedings of AIMTDR*, 2023 (to appear).
- Rajput, H.S. and Law, M., A Compressed Sensing Framework to Recover Cutting Tool Modal Parameters from Aliased Video, *Proc. of the 15th Int. Conf. on Vibration Problems*, Doha, Qatar, 2023.
- Rajput, H.S., Raizada, V., and Law, M., Deep Learning-Based Recovery of Cutting Tool Vibrations from Spatiotemporally Aliased Video, *Sadhana*, under review, 2023.
- RDI Technologies, accessed December 25, 2023, from <https://rditechnologies.com/>, 2023.
- Reu, P.L., Sweatt, W.C., Miller, T.M., and Fleming, D., Camera System Resolution and Its Influence on Digital Image Correlation, *Exp. Mech.*, vol. **55**, no. 1, pp. 9–25, 2015. DOI: 10.1007/s11340-014-9886-y
- Singh, S.A., Kumar, A.S., and Desai, K.A., Vision-Based System for Automated Image Dataset Labeling and Dimension Measurements on Shop Floor, *Measurement*, vol. **216**, p. 112980, 2023. DOI: 10.1016/j.measurement.2023.112980.
- Singh, V. and Law, M., Machine Tool Multibody Dynamic Model Updating Using Vision-Based Modal Analysis, *Manuf. Technol. Today*, vol. **22**, no. 2, pp. 23–28, 2023.
- Spencer, B.F., Hoskere, V., and Narazaki, Y., Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring, *Engineering*, vol. **5**, no. 2, pp. 199–222, 2019. DOI: 10.1016/j.eng.2018.11.030
- Sun, D., Yang, X., Liu, M.-Y., and Kautz, J., PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume, *2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, pp. 8934–8943, 2018. DOI: 10.1109/cvpr.2018.00931
- Thielicke, W. and Stamhuis, E.J., PIVlab – Towards User-Friendly, Affordable and Accurate Digital Particle Image Velocimetry in MATLAB, *J. Open Res. Software*, vol. **2**, no. 1, 2014. DOI: 10.5334/jors.bl

- Verma, N., Wahi, P., and Law, M., Vision-based Runout Measurement Method for End Mills, *Select Proc. of AIMTDR*, 2023 (to appear).
- Vogl, G.W., Rexford, A., Li, Z., Landers, R.G., Kinzel, E.C., Alkan Donmez, M., and Chalfoun, J., Vision-Based Thermal Drift Monitoring Method for Machine Tools, *CIRP Annals*, vol. **72**, no. 1, pp. 301–4, 2023. DOI: 10.1016/j.cirp.2023.04.053
- Wikipedia, Event Camera, accessed December 16, 2023, from https://en.wikipedia.org/wiki/Event_camera, 2023.
- Yamamoto, F., Wada, A.I., Iguchi, M., and Ishikawa, M.A., A Review on the Cross-Correlation Methods for PIV, *J. Flow Vis. Image Process.*, vol. **24**, nos. 1–4, pp. 215–228, 2017. DOI: 10.1615/JFlowVisImageProc.v24.i1-4.140
- Yanamadala, H. and Muralidharan, P.K., Comparative Study of Vision Camera-based Vibration Analysis with the Laser Vibrometer Method, accessed December 16, 2023, from <https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1608598&dswid=-3097>, 2021.
- Zheng, J., Zhao, Y., Ge, M., Bi, W., and Ge, P., Machine Vision-Based Transverse Vibration Measurement of Diamond Wire, *Precis. Eng.*, vol. **80**, pp. 115–26, 2022. DOI: 10.1016/j.precisioneng.2022.12.004

