

When Mobile ToF Meets Micro-Vibration: Multi-Point kHz-Frequency Sensing using Laser Speckle

Shangcheng Jin

The Chinese University of Hong Kong
Hong Kong SAR, China
1155206412@link.cuhk.edu.hk

Guoliang Xing

The Chinese University of Hong Kong
Hong Kong SAR, China
glxing@cuhk.edu.hk

Zhiyuan Xie

The Chinese University of Hong Kong
Hong Kong SAR, China
zyxie@cuhk.edu.hk

Zhenyu Yan*

The Chinese University of Hong Kong
Hong Kong SAR, China
zyyan@cuhk.edu.hk

Abstract

Sensing micro-vibration plays a critical role in diverse industrial applications, including structural health monitoring and machinery diagnostics. However, most existing vibration sensing technologies rely on physical contact with the target object or demand complex installation procedures, making them intrusive and costly. This paper presents PocketVib, the first cost-effective, portable smartphone-based laser speckle vibrometry system for multi-point kHz-frequency micro-vibration sensing. PocketVib integrates built-in smartphone sensors, i.e., Time-of-Flight (ToF) depth cameras and rolling shutter cameras, to achieve non-contact micron-level vibration sensing. The design features three key innovations: 1) emitter-imager synchronization, which ensures stable speckle pattern acquisition, 2) a dual-lens enhanced imager that stretches speckle patterns to improve their limited coverage on the rolling shutter camera, and 3) a reference-free algorithm that leverages temporal correlations across rolling shutter image rows for precise vibration extraction. We implement PocketVib on commodity smartphones with a custom hardware kit. Extensive experimental results demonstrate that PocketVib enables fast and accurate vibration frequency estimation on mobile devices, achieving frequency measurements of up to 1 kHz with an average error of only 0.6 Hz at just 30 fps, while supporting simultaneous measurements at seven points. The project page is available at <https://shangchengjin.github.io/PocketVibWeb>.

* Zhenyu Yan is the corresponding author.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

MobiCom '26, Austin, TX, USA

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2505-0/26/10

<https://doi.org/10.1145/3795866.3796680>

CCS Concepts

• Human-centered computing → Ubiquitous and mobile computing systems and tools; • Computing methodologies → Perception.

Keywords

ToF Depth Camera, Rolling Shutter Camera, Smartphone, Vibration Sensing, Laser Speckle

ACM Reference Format:

Shangcheng Jin, Zhiyuan Xie, Guoliang Xing, and Zhenyu Yan*. 2026. When Mobile ToF Meets Micro-Vibration: Multi-Point kHz-Frequency Sensing using Laser Speckle. In *The 32nd Annual International Conference on Mobile Computing and Networking (MobiCom '26)*, October 26–30, 2026, Austin, TX, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3795866.3796680>

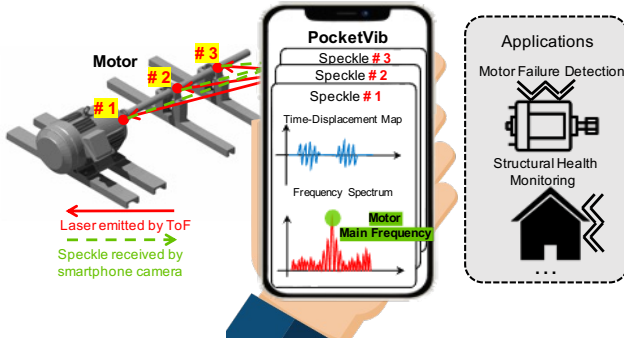
1 Introduction

Sensing micro-vibration is essential in industrial scenarios, from detecting subtle structural tremors to monitoring mechanical oscillations in factories. For example, technicians need to conduct rapid routine inspections of numerous equipment items to identify potential safety hazards at an early stage. Portable vibration meters are widely used for these inspections due to their ease of use, affordability, and mobility [18]. Notably, the global market for portable vibration meters is expected to reach USD 23.11 billion by 2033 [17].

Existing micro-vibration sensing methods can be categorized as contact-based or non-contact-based, as summarized in Table 1. Contact-based methods utilize accelerometers [23, 29, 41] and piezoelectric sensors [37, 62] that are physically attached to the vibrating surface. However, this attachment will alter the object's vibration characteristics and restrict measurements to a single point per sensor. Some methods use wireless technologies like RFID [55, 58] and mmWave [42, 61] for vibration sensing. However, the long wavelength of RFID signals limits their capacity to detect

Table 1: Comparison of existing technologies for micro-vibration sensing.

Sensing Technologies	Type	Micron-level	Frequency Band	Sensing Range	Cost	Multi-Point	Deployment
Accelerometer / Piezoelectric Sensors	Contact	✓	kHz-level	0 m	Low	×	Physical attachment
Wireless Signals (mmWave / RFID)	Non-contact	✓ / ×	kHz-level	0.5–7 m [33]/<1.5 m [55]	Medium	✓(mmWave)	Costly setups and calibration
RGB Camera	Non-contact	✓	kHz-level (high-speed)	0.5–2 m [27]	High	✓	Bulky, environment-dependent
Laser Doppler Vibrometry	Non-contact	✓	MHz-level	0.2–5 m [16]	Very High	×	Bulky hardware
Laser Speckle Vibrometry	Non-contact	✓	kHz-level (high-speed)	Tens of meters [63]	Medium	✓	Bulky hardware
PocketVib (Ours)	Non-contact	✓	kHz-level (at 30 fps)	0.05–2 m	Low	✓	A single smartphone

**Figure 1: Illustration of PocketVib.**

micron-level vibrations. Although mmWave radars offer high accuracy and multi-object sensing via beamforming, they demand costly setups and calibration. Some approaches use cameras for micro-vibration sensing by tracking pixel shifts in videos [24, 26, 27]. However, they require expensive high-speed cameras, good lighting conditions, and textured surfaces. Laser Doppler Vibrometry (LDV), the gold standard for accurate vibration sensing, is extremely expensive (approximately 30,000 USD per unit), limited to single-point measurements, and lacks portability due to its bulky setup [50].

Recently, Laser Speckle Vibrometry (LSV) has emerged as a promising technology for sensing kHz-frequency micro-vibration without physical contact [43, 63]. A typical LSV system contains a laser to generate speckle patterns and a high-speed, defocused camera to capture them. The camera is deliberately defocused to enlarge the speckle patterns, thereby amplifying the subtle shifts caused by the micro-vibration. Furthermore, LSV can monitor vibrations at multiple points simultaneously by emitting multiple laser beams. However, current systems rely on expensive, bulky, and specialized lasers and high-speed cameras. Currently, no existing solution supports contactless, multi-point micro-vibration sensing while remaining easy to deploy for cost-effective routine inspections.

In this paper, we propose PocketVib, the first mobile laser speckle vibrometry system that enables multi-point, kHz-frequency micro-vibration sensing by leveraging built-in sensors in smartphones, i.e., a ToF depth camera (e.g., iPhone LiDAR) and a rolling shutter camera. This novel approach is made possible by three key technologies in modern smartphones: First, an increasing number of smartphones are now

equipped with ToF depth cameras [56, 57], which utilize eye-safe infrared laser sources. These sensors have the potential to replace specialized lasers in speckle-based applications [22]. Second, the built-in auto-focus capability of smartphone cameras enables defocused imaging, which is critical for amplifying speckle shifts caused by micro-vibration. Third, smartphone cameras utilize a rolling shutter mechanism, which functions as row-wise high-sampling-rate sensors for kHz-frequency vibration sensing [27, 38, 43].

However, compared with traditional LSV, the design of PocketVib faces three major challenges. First, the laser source in the ToF camera will be turned on and off periodically. The desynchronization between the ToF laser source and the rolling shutter camera causes unstable speckle patterns, which hinder the reliable extraction of continuous vibration information. Second, due to the limited optical characteristics of smartphone rolling shutter cameras, it is difficult to capture defocused speckle patterns of the desired size for accurate vibration analysis, as required by existing LSV methods [21, 30, 43]. Third, speckle-based vibration analysis typically requires an additional reference frame to compare with the current frame for extracting vibration information.

To address the above challenges, PocketVib introduces three novel key designs, i.e., an emitter-imager synchronization, a dual-lens enhanced imager, and a reference-free micro-vibration extraction. First, we conduct a comprehensive profiling of the operational mechanism of the ToF laser emission and design a precise synchronization mechanism between the ToF camera and the rolling shutter camera. This ensures stable speckle patterns, enabling continuous and reliable acquisition of vibration data. Second, we develop the PocketVib Kit equipped with a cylindrical lens to address the limited coverage of defocused speckle patterns on the rolling shutter camera sensor. The lens stretches the speckle patterns along the rolling shutter direction, allowing more rows of the speckle image to be captured and enhancing the accuracy of vibration analysis. Third, instead of relying on an external reference frame, we leverage the temporal relationship between adjacent rows of the rolling shutter camera to extract micro-vibration information.

We implement PocketVib on two smartphone platforms (Apple and Samsung). The system comprises a customized, lightweight hardware kit that mounts seamlessly to the phone camera and a mobile app that performs on-device data

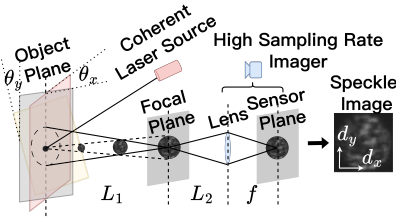


Figure 2: Speckle-based vibrometry.

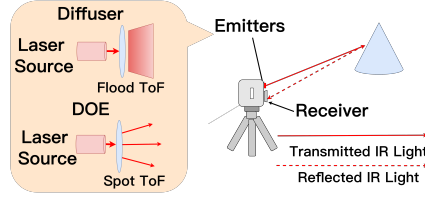


Figure 3: ToF emitter types.

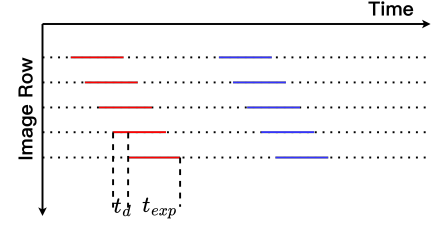


Figure 4: Rolling shutter mechanism.

collection and processing, achieving an end-to-end latency of under 206 ms per frame. We conduct a comprehensive evaluation of PocketVib, covering various vibration amplitudes and frequencies, as well as multi-point scenarios. We also explore the impact of several practical factors that affect the real-world applicability of PocketVib, such as distances, surface materials, hand-held motions, and lighting conditions. Results demonstrate that PocketVib achieves high-precision, micron-level vibration frequency detection up to 1 kHz, with an average error of just 0.6 Hz, and supports simultaneous seven-point measurements. Additionally, PocketVib maintains robust performance at distances up to 70 cm (Setup S1, Apple) and 200 cm (Setup S2, Samsung), across highly reflective surfaces, under varying illumination, and during hand-held use. Finally, we demonstrate PocketVib's real-world applicability by measuring vibrations from various mechanical sources as well as those induced by sound. Our key contributions are summarized as follows:

- To the best of our knowledge, PocketVib is the first cost-effective and portable smartphone-based LSV system for multi-point, kHz-frequency micro-vibration sensing.
- We conduct an in-depth analysis of the spot ToF emission mechanism and design a precise synchronization between the ToF camera and the rolling shutter camera, providing theoretical support for mobile speckle-based applications.
- We introduce a dual-lens enhanced imaging design to extend speckle coverage on rolling shutter sensors and develop a reference-free algorithm that leverages temporal correlations across rolling shutter rows for precise micro-vibration extraction.
- We implement PocketVib on commercial smartphones and evaluate it in real-world scenarios. Extensive experimental results show that PocketVib achieves accurate multi-point vibration frequency estimation up to 1 kHz with an average error of only 0.6 Hz at 30 fps.

2 Background and Motivation

In this section, we provide the technical background of laser speckle vibrometry, systematically analyzing its operation principles, advantages in vibration sensing, and inherent limitations. Building upon these constraints, we subsequently

investigate smartphone-based solutions to overcome the system barriers of conventional LSV.

2.1 Laser Speckle Vibrometry

As a non-contact optical sensing technique, laser speckle vibrometry enables multi-point micro-vibration measurement through the coherent light interference phenomena. Its standard implementation comprises two core components: 1) a **coherent laser source** that generates stable speckle patterns, and 2) a **defocused high-speed imaging system** that captures speckle dynamics [63].

The coherence requirement is essential because when coherent light illuminates a rough surface, scattered waves with varying optical paths create constructive and destructive interference on the detector plane due to their different relative phases, forming speckle patterns with complex bright and dark spots [25]. To maximize sensitivity, LSV systems intentionally deploy a defocusing strategy by moving the focal plane far from the target surface. As illustrated in Fig. 2, this configuration enlarges individual speckles across multiple sensor pixels, amplifying motion-induced intensity variations [64]. Concurrently, high sampling rates (typically ≥ 2 kHz [63]) ensure accurate capture of high-frequency vibrations while maintaining multi-point measurement capability through parallel speckle tracking.

LSV exhibits extremely high sensitivity to micro-motions, especially tilt motion [63]. According to geometric optics in Fig. 2, the relationship between tilt motion (θ_x, θ_y) and speckle pattern shifts (d_x, d_y) can be derived as follows [30]:

$$(d_x^{tilt}, d_y^{tilt}) = \left(\frac{fL_1 \tan \theta_x}{L_2}, \frac{fL_1 \tan \theta_y}{L_2} \right), \quad (1)$$

where f denotes focal length, L_1 represents object-to-focal-plane distance, and L_2 is focal-plane-to-lens separation.

Despite their precision, conventional LSV systems suffer from significant implementation constraints: 1) dependency on specialized laser sources (Class 3B), 2) bulky optical components such as specialized lenses and cameras unsuitable for field deployment, and 3) restricted operation in controlled laboratory environments. These limitations motivate our exploration of smartphone-based alternatives leveraging consumer-grade hardware.

2.2 Coherent Laser Source: Time-of-Flight Depth Cameras

Current smartphones integrate coherent light sources via their built-in ToF depth cameras, particularly those implementing spot illumination architectures. ToF cameras measure distance by calculating the round-trip time of an artificial light signal (typically emitted by Vertical-Cavity Surface-Emitting Lasers, VCSELs). Unlike conventional flood ToF systems that homogenize the illumination energy via a diffuser, spot ToF preserves beam coherence using diffractive optical elements (DOEs) to generate structured light patterns (Fig. 3). Spot ToF technology has been widely adopted in consumer devices. For instance, Apple's iPhone Pro series (12-16 generations) introduces two spot ToF systems with distinct illumination matrices. Similarly, Samsung [8, 44] and STMicroelectronics [13, 34] have also advanced spot ToF technology into mass production. In this paper, we leverage Apple's commercially available spot ToF cameras (LiDAR) as an example to demonstrate that spot ToF emitters have the potential to replace specialized lasers in LSV systems.

2.3 High-Sampling-Rate Defocused Imager: Rolling Shutter Cameras

Although spot ToF emitters offer a viable alternative to specialized lasers in LSV systems, their native receivers fall short of meeting the requirements of LSV systems due to several critical limitations. First, these receivers are low-frame-rate global shutter cameras designed to accurately calculate the time of flight, which is inadequate for kHz-frequency vibration detection. Second, they lack auto-focus modules that are essential for capturing defocused imaging to enhance micro-motion sensitivity.

Modern smartphone cameras feature rolling shutter mechanisms and auto-focus systems, making them particularly well-suited as high-sampling-rate defocused imagers for LSV systems. Fig. 4 illustrates the operating principle of a rolling shutter camera. The exposure of adjacent rows is staggered by a time interval t_d , and each row is exposed for a duration of t_{exp} . This design enables asynchronous information capture across rows, resulting in an equivalent sampling rate of $1/t_d$. However, the effective sampling rate is also limited by t_{exp} . When the vibration frequency exceeds $1/t_{exp}$, the prolonged exposure time within a single row can generate motion blur artifacts, thereby hindering accurate vibration extraction [27]. Consequently, the system's effective sampling rate is governed by $\min(1/t_d, 1/t_{exp})$. The time interval t_d can be measured through camera calibration [40]. A recent study [38] measured t_d values for smartphones from various manufacturers like Google, Samsung, and Apple, typically ranging from 8 to 20 microseconds. As a result, rolling shutter cameras achieve equivalent sampling rates

between 50 kHz and 125 kHz. Meanwhile, the exposure time t_{exp} can be easily configured in camera settings, with a minimum value of $1/8000$ s available on smartphone cameras. Therefore, the overall sampling rate for vibration detection is limited to 8000 Hz, still significantly exceeding the standard video frame rate of 30 fps.

3 Feasibility and Challenges

3.1 A Feasibility Study

We conduct a feasibility study to evaluate the feasibility of smartphone-based LSV using iPhone's LiDAR and rolling shutter cameras. To investigate the relationship between object motion and the resulting speckle dynamics, we use the iPhone's LiDAR to project laser spots onto an optomechanical platform with reflective patches. This platform provides precise control (0.1° tilt resolution, 0.1 mm transverse precision). To enable infrared speckle capture, we perform hardware modifications on a Samsung S20 Ultra wide-angle camera by removing its IR-cut filter following established methodologies [9]. The imaging system operates at 4K resolution with 30 fps frame rate, $1/2000$ s exposure time, and ISO 3200 sensitivity. To maximize speckle defocus, we adjust the focus parameter to one of the two extremes: minimum focus (focus = 0.0) or maximum focus (focus = 1.0). Under both settings, we capture speckle patterns at distances ranging from 5 to 400 cm and analyze the resulting speckle diameters. Due to the instability of the captured speckle patterns, we capture short video sequences for reliable analysis (Fig. 5a).

For this camera, a focus setting of 1.0 is optimal for distances under 10 cm, while a setting of 0.0 provides better defocus for distances over 10 cm (Fig. 5b). We also calculate the average intensity of the speckle patterns at these defocused settings. Since larger defocus speckles generally correspond to higher motion sensitivity in the above experiments [64], we further analyze the motion sensitivity of speckle patterns at $d = 20$ cm, where the diameter of the maximum defocus speckle patterns reaches its minimum. As illustrated in Figs. 5c and 5d, a linear relationship exists between mechanical displacement and speckle shift, with higher sensitivity to tilt motion. For tilt micro-motion, the displacement of the observation point can be approximated as $\Delta x \approx l\theta$, where l is the distance from the observation point to the rotation axis, and θ is the angle displacement in radians. For instance, a $1 \mu\text{m}$ displacement at $l = 5.73$ mm corresponds to a 0.01° angular change, resulting in speckle shifts of 3.33 pixels. By substituting Δx into Eq. 1, we further establish the relationship between speckle shift and the actual displacement, given by $\Delta x = \frac{s l L_2 d^{\text{tilt}}}{f L_1}$ (where s is the pixel size). In summary, the preliminary results show that it is promising to capture micron-level motion using the iPhone's LiDAR and the rolling shutter camera.

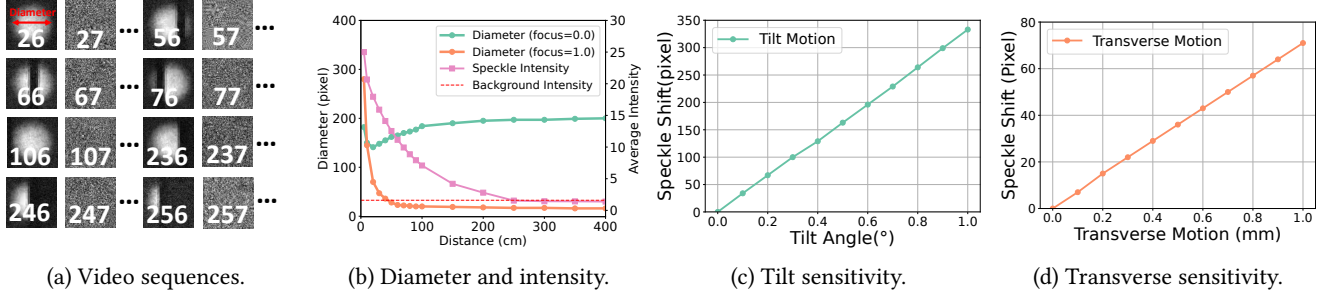


Figure 5: (a) Captured unstable speckle patterns (the numbers in the frames represent the frame index); (b) Speckle diameter and intensity at distances of 5–400 cm under two focus settings (0.0, 1.0); (c) (d) Motion sensitivity of defocused speckle patterns at $d = 20$ cm for tilt and transverse motion, respectively.

3.2 Challenges

While our feasibility study shows promising results for micro-motion sensing, several limitations hinder the application of these two sensors for kHz-frequency micro-vibration sensing. There exist three technical challenges.

First, the observed speckle patterns in the feasibility study are significantly unstable. The speckle images in Fig. 5a show three types of unstable speckle patterns caused by two factors. The first is rolling shutter distortions. The row-wise exposure mechanism of rolling shutter cameras introduces spatial artifacts (see frames 56, 66, 76 in Fig. 5a). Sensor rows exposed during the laser’s “on” phases capture bright bands, while those exposed during the “off” phases record dark bands. These dark bands cause incomplete speckle patterns by obscuring the vibration-induced speckle information. The second is temporal misalignment. The mismatch between the laser emission cycle and the camera frame rate causes two distinct artifacts across frames: flickering speckle patterns (see frames 26~27 in Fig. 5a) and “scrolling” bright-dark bands (see frames 56~76 in Fig. 5a).

Second, in our feasibility study, the maximum diameter of defocused speckles is approximately 300 pixels, which is much smaller than the 2160 pixel rows of the rolling shutter sensor. This limitation mainly arises from the restricted optical characteristics of smartphone cameras [12, 19]. With a row scanning speed of $10\mu\text{s}$, this speckle size corresponds to only 300 rows per frame, yielding a duty cycle of merely 9.1% (far below the device’s maximum of 65.4%). Such a low-duty cycle significantly limits the accurate vibration analysis.

Third, speckle-based vibration analysis requires an additional reference frame to allow differential comparison with the current frame for vibration extraction. Existing approaches capture this frame when the object is stationary [63], which is not always available. Other methods employ an additional high-end camera to obtain the reference [43], which is impractical with our single-smartphone setup.

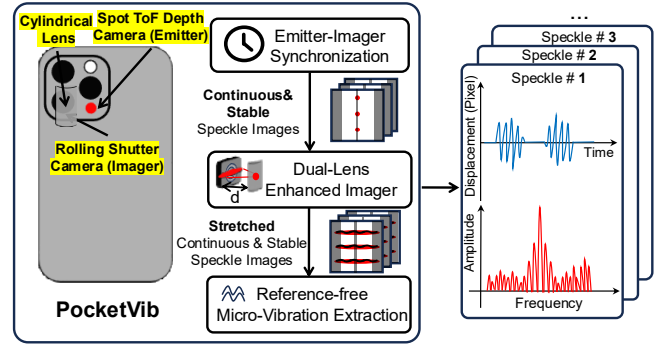


Figure 6: System overview of PocketVib.

4 Design of PocketVib

4.1 System Overview

In this paper, we propose PocketVib, the first smartphone-based laser speckle vibrometry (LSV) system that enables multi-point, kHz-frequency micro-vibration sensing using a spot ToF depth camera and a rolling-shutter camera. Fig. 6 illustrates the system design. First, we develop an emitter-imager synchronization module that can synchronize the rolling shutter imager with the ToF emitter to capture continuous and stable vibration-related speckle images. Second, we design a dual-lens enhanced imager that stretches defocused speckle patterns along the rolling shutter direction, increasing their coverage across sensor rows. This enables the camera to sample more data per frame, supporting accurate vibration analysis. Last, we adopt a reference-free method to extract micro-vibration information by exploiting the temporal relationship across rolling shutter rows, eliminating the need for additional reference frames.

Application Scenarios. PocketVib aims to enable contactless, multi-point kHz-frequency micro-vibration sensing on mobile devices like smartphones. In rotating machinery, faults such as imbalance, misalignment, and bearing defects typically produce sub-kHz fault frequencies. However, they

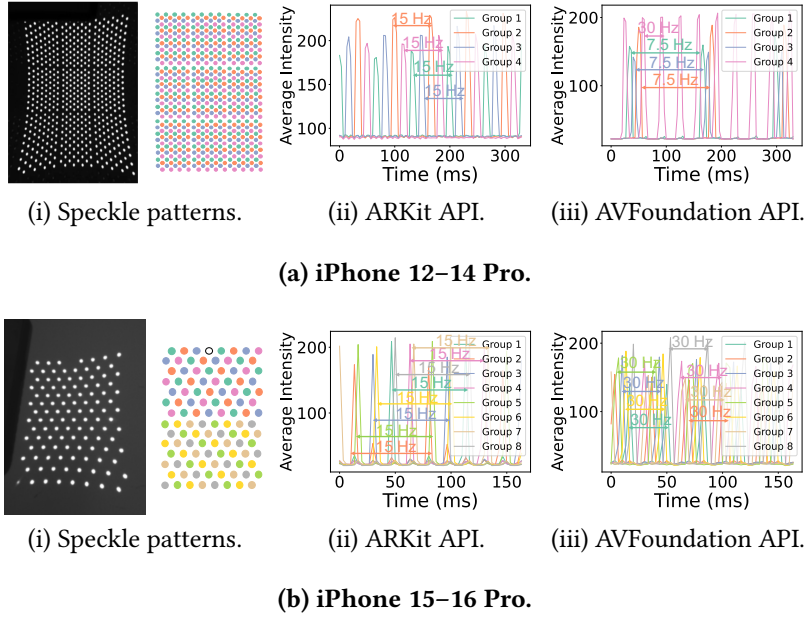


Figure 7: Profiling LiDAR dot matrix's emission characteristics on two smartphones.

can generate impulses that excite structural resonances at 1–2 kHz [20]. PocketVib's kHz-level sensing capability is highly effective for this diagnosis. For practical use, users or automated guided vehicles (AGVs) can simply hold a smartphone to access the testing object at sub-meter ranges, which is sufficient for routine inspection tasks in most industrial environments.

4.2 Emitter-Imager Synchronization

To enable continuous and stable acquisition of vibration-related speckle images, we first profile the emission mechanism of the iPhone's LiDAR systems. This step is crucial for understanding the causes of unstable speckle patterns and enabling precise synchronization between the spot ToF emitter and the rolling shutter imager. Specifically, we use a high-speed global shutter infrared camera¹ to observe the speckle patterns generated by two Apple smartphone LiDAR systems (spot ToF, see § 2.2). These LiDAR systems are activated using two APIs: ARKit [6] and AVFoundation [10]. Fig. 7 shows the distinct speckle patterns generated by the two LiDAR systems. By analyzing the intensity of each speckle over time, we observed that the speckle patterns follow a group activation mechanism, where only one group is active at any given moment, and each group has its own activation cycle. Furthermore, the activation cycle varies depending on

¹The Intel RealSense D435i [11] was used, operating at 300 fps with an exposure time of 2.95 ms.

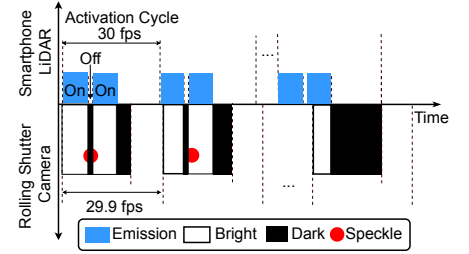


Figure 8: Understanding incomplete, flickering, and scrolling speckle patterns.

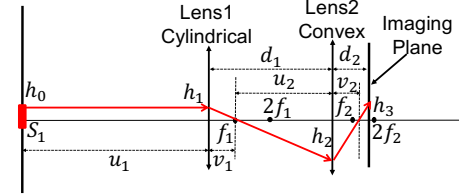


Figure 9: Optical model of Dual-Lens Enhanced Imager.

the API used: ARKit limits each group to operate at 15 Hz, whereas AVFoundation supports group operation at 30 Hz. Prior work [4] notes that each cycle contains tens of thousands of laser pulses, emitted in two phases (Fig. 8).

Now, we synchronize the rolling-shutter imager with the LiDAR emitter. Let f_o and f_i represent the imager frame rate and the LiDAR group activation cycle, respectively. If $f_o = f_i/N$, where $N \in \mathbb{N}^*$, speckle patterns can be captured continuously without flickering due to the consistent energy harvesting across frames. Many smartphone cameras natively support 30 fps, matching the AVFoundation-driven LiDAR cycle. Then, we illustrate rolling-shutter sampling of these speckle patterns under short exposure times (Fig. 8). The row-wise exposure mechanism introduces rolling shutter distortions, resulting in incomplete speckle patterns. Specifically, rows exposed during the laser's "on" phases capture bright bands, while those exposed during the "off" phases record dark bands. There are two types of dark bands: the first arises from the dead time between two emission phases, and the second is caused by the idle state between activation cycles. However, achieving perfect synchronization is non-trivial. The slight frequency mismatches between the LiDAR emitter (30 Hz) and the rolling shutter imager (29.9 Hz) cause bright-dark bands to drift across frames. By eliminating flickering patterns and analyzing the causes of incomplete and scrolling patterns, we can identify continuous, stable speckle regions for downstream processing.

4.3 Dual-Lens Enhanced Imager

Since the defocusing capability of smartphone cameras is limited [12, 19], the observed defocused speckle patterns only occupy a few rolling shutter rows, resulting in a low-duty cycle for vibration analysis. To address this limitation, we design a dual-lens enhanced imager by introducing a cylindrical lens in front of smartphone cameras to stretch defocused speckle patterns. As illustrated in Fig. 9, for the sake of discussion, we model the rolling shutter camera as a system consisting of a convex lens (Lens 2, focal length f_2) and an imaging plane, where the distance between Lens 2 and the imaging plane is d_2 . Then we place a cylindrical lens (Lens 1, focal length f_1) at a distance d_1 from Lens 2. To quantify the stretching effect, we use ray tracing to study the characteristics of the dual-lens optical system. Consider a parallel ray originating from light source S_1 (at distance $u_1 = \infty$ from Lens 1) at height h_0 . It maintains its height ($h_1 = h_0$) at Lens 1. After refraction through Lens 1, the ray passes through the focal point f_1 (image distance $v_1 = f_1$) and intersects Lens 2 at height h_2 , which can be expressed as $h_2 = h_0(1 - d_1/f_1)$. Subsequent refraction through Lens 2 produces the image distance v_2 and the final image height h_3 . The v_2 and h_3 can be expressed as: $1/(d_1 - v_1) + 1/v_2 = 1/f_2$ and $h_3 = h_2(1 - d_2/v_2)$. By substituting h_2 and v_2 , the final expression for h_3 is derived as:

$$h_3 = \frac{h_0}{f_1 f_2} [-d_2 f_2 + (d_2 - f_2)(d_1 - f_1)], (d_1 \neq v_1). \quad (2)$$

Since Lens 1 only magnifies along one axis, we now consider the other axis, where Lens 1 has no effect. The height h_4 is $h_0(1 - d_2/f_2)$. The stretch ratio of the dual-lens imaging system is quantified by:

$$\frac{h_3}{h_4} = \frac{f_2^2}{f_1 d_2 - f_1 f_2} + (1 - \frac{d_1}{f_1} + \frac{f_2}{f_1}). \quad (3)$$

Let us consider a special case, where $d_1 = f_1$. In this case, the expression for h_3 needs to be modified, and we have the following equation:

$$\frac{h'_3}{h_4} = \frac{h_0(1 - (d_1 + d_2)/f_1)}{h_0(1 - d_2/f_2)} = \frac{f_2^2}{f_1 d_2 - f_1 f_2} + \frac{f_2}{f_1}. \quad (4)$$

This result is consistent with substituting $d_1 = f_1$ into Eq. 3.

For smartphone cameras, d_2 is typically constrained to the range $f_2 < d_2 < 2f_2$ [52]. According to Eq. 3, the stretch ratio h_3/h_4 decreases as d_2 increases within this interval. To achieve the stretching effect for all focus settings, it is sufficient to guarantee that $h_3/h_4 > 1$ when d_2 approaches $2f_2$ (as h_3/h_4 approaches positive infinity, when d_2 approaches f_2). From this analysis, we have $h_3/h_4 = 1 + (2f_2 - d_1)/f_1 > 1$. For a given smartphone camera (f_2 is fixed), we can ensure the stretching effect by simply keeping $d_1 \leq 2f_2$. Additionally,

smaller values of d_1 and f_1 can further enhance the stretching effect while maintaining the system's portability.

4.4 Reference-free Micro-vibration Extraction

We begin by formulating the rationale for exploiting the temporal relationship across rolling shutter rows to extract micro-vibration information. Let $I_y(x)$ denote the intensity map of the y th row. Assuming a stationary reference frame $I^{ref}(x)$, we have $I_y(x - d_y) = I_y^{ref}(x)$ and $I_{y+1}(x - d_{y+1}) = I_{y+1}^{ref}(x)$, where d_y and d_{y+1} represent the speckle shifts for y th and $y + 1$ th row, respectively. Since the start times of two adjacent rows are very close and their spatial intensity distributions are nearly identical, we can assume: $I_y^{ref}(x) = I_{y+1}^{ref}(x)$. Substituting this into the earlier equations, we obtain the relationship $I_y(x - d_y) = I_{y+1}(x - d_{y+1})$. Rearranging this expression yields $I_y(x) = I_{y+1}(x - (d_{y+1} - d_y))$. According to Eq. 1, the speckle shifts (d_y, d_{y+1}) are linearly related to the motion of the object. Let $\Delta x_{y+1} = d_{y+1} - d_y$. The term Δx_{y+1} is also linearly related to the motion of the object. The final expression can be written as:

$$I_y(x) = I_{y+1}(x - \Delta x_{y+1}) \quad (5)$$

This formulation enables the extraction of one-dimensional vibration signals perpendicular to the rolling shutter direction and eliminates the need for an external reference. Prior work [54] further shows that even when the vibration direction is not perfectly perpendicular, the captured motion component still retains valuable vibration information.

Consider a set of N frames, each containing K speckles. For each frame, we develop a dynamic region filtering method to extract vibration-related speckle regions. Specifically, we compute the average intensity along both columns and rows, followed by smoothing to reduce noise. For column-wise processing, we use peak detection algorithms [1] to identify peaks and determine their intervals using full width at half maximum. For row-wise processing, we use a simpler statistical approach, with intervals defined by the mean and standard deviation (i.e., $mean + 0.5\sigma$). The final speckle regions are obtained by intersecting the column-wise and row-wise intervals. An example is shown in Fig. 10. For each identified speckle region, we estimate inter-row displacement Δx_{y+1} using sub-pixel normalized cross-correlation (NCC) [45]. Specifically, we compute the NCC between adjacent rows and determine the sub-pixel peak location of the correlation function via quadratic interpolation. The resulting peak index (Δx_{y+1}) is then converted to the time domain. To address the small dark band observed between two bright bands, we apply auto-regressive interpolation [32]. To ensure temporal continuity, the rolling shutter dead-time intervals

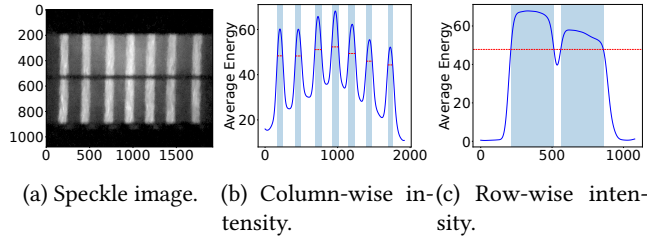


Figure 10: Dynamic speckle region detection. Red lines in (b) and (c) indicate adaptive thresholds.

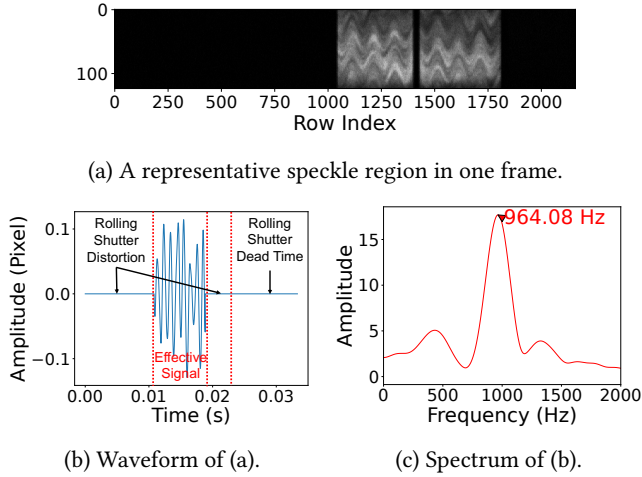


Figure 11: An example of vibration signal extraction and frequency analysis.

are padded with zeros before processing the next frame. Finally, the extracted time-domain signals are denoised using a low-pass filter to enhance signal quality (i.e., 2000 Hz cut-off, corresponding to the default 1/2000 s shutter speed). To further analyze these signals, we apply the Fast Fourier Transform (FFT) to obtain the frequency spectrum, which is widely used in industrial applications for identifying key vibration characteristics such as resonance, imbalance, and structural faults. Fig. 11 shows an example of the processing results², including a representative speckle region, the extracted waveform, and the corresponding frequency spectrum with a dominant vibration component at 964.08 Hz.

Although we have derived the relationship between pixel displacement maps and physical vibration amplitude, given by $\Delta x = \frac{slL_2 d^{t_{tilt}}}{fL_1}$ (see § 3), the key parameters (l, L_1) require frequent manual measurement, making PocketVib impractical for amplitude measurement in mobile routine inspections.

²The speckle image used in this example is obtained from § 6 and corresponds to a 964.90 Hz vibration.

5 Implementation of PocketVib

The implementation of PocketVib consists of a customized hardware module, the PocketVib Kit, and a mobile app for on-device vibration analysis.

The PocketVib Kit integrates a cylindrical lens, a slit, an IR pass filter, and a mounting bracket. S1 in Fig. 12 shows the prototype attached to the iPhone 12 Pro. As discussed in § 4.3, a short-focal-length cylindrical lens mounted directly on the rolling shutter camera enables sufficient spatial stretching. We use a 10 mm focal length cylindrical lens (~30 USD), placed in direct contact with the camera lens. Since this stretching makes the spots along the rolling shutter direction in the LiDAR laser dot matrix overlap with each other, we use a small slit on the LiDAR emitter, allowing only a single row of speckles to exist within the camera’s field of view (FOV). Note that the number of speckles in the row corresponds to the number of simultaneous measurements. The LiDAR system in the 12–14 Pro models supports multiple simultaneous spots per row, while the 15–16 Pro activates only one spot per row. To capture the infrared-defocused speckle images required by LSV systems, we manually remove the IR-cut filter [49] of the smartphone’s camera with auto-focus capability. This is a well-known operation for infrared photography, used in both professional and consumer-grade imaging systems [9, 22]. In this paper, we modify the iPhone 12 Pro’s telephoto camera and the Samsung Galaxy S20 Ultra’s wide-angle camera. Lastly, an IR-pass filter is attached to the modified lens to block visible light interference. The total cost of PocketVib Kit is ~33 USD.

The PocketVib App enables continuous, reliable acquisition of vibration-related speckle images and supports on-device vibration analysis. We implement this app on the iOS platform because the camera systems on the iPhone 12–16 Pro series already support our system, and Android smartphones such as Samsung will support it in the future (see § 2.2). In particular, the app builds on the AVFoundation API [10] and the AVCaptureMultiCamSession API [3] to implement the emitter-imager synchronization in § 4.2 by controlling the iPhone’s LiDAR and configuring the rolling shutter camera parameters, including ISO, shutter speed, focus, and frame rate (30 fps). Due to the AVCaptureMultiCamSession API restrictions, direct raw image access is unavailable [3], so we capture RGB frames and convert them to grayscale (intensity) for subsequent processing. The app’s data processing module builds on the BeeWare framework [14] for the reference-free micro-vibration extraction in § 4.4. Finally, the app outputs the time-displacement maps and corresponding frequency spectra for multiple speckles.

6 Evaluation

6.1 Experimental Setup

Hardware Setup. We implement PocketVib on both Apple (S1) and Samsung (S2) platforms. The S1 setup uses the iPhone 12 Pro’s hardware for both laser projection and speckle patterns capture. As discussed in § 2.2, Samsung’s current smartphone camera system does not include the spot ToF depth camera, but will include one in the future. Therefore, for the S2 setup, we use the iPhone 12 Pro’s spot ToF camera for laser projection and the Samsung S20 Ultra’s wide-angle camera for speckle patterns capture.

Vibration Sources and Ground Truth. We use a series of tuning forks with their vibration frequency ranging from 30 Hz to 965 Hz to generate precise single-frequency vibrations, covering the typical range found in industrial and daily-life scenarios, which is also the range commonly measured by hand-held vibrometers [15]. The vibration direction of the tuning forks is set to be orthogonal to the rolling shutter direction. We use a professional laser displacement sensor to collect the ground truth vibration. Its accuracy is $0.1 \mu\text{m}$, and the required measuring distance is $30 \pm 4 \text{ mm}$.

Evaluation Metrics. To evaluate the performance of PocketVib, we adopt frequency error as the primary metric to assess the accuracy of micro-vibration reconstruction. We will also include the vibration amplitude information measured by the laser displacement sensor to provide additional context for micro-vibration measurements.

Camera Parameters. Both S1 and S2 operate with a default shutter speed of $1/2000 \text{ s}$ (corresponding to 2000 Hz effective sampling rate [27, 43]) at 30 fps and maximum ISO values (S1: 2203, S2: 3200) to compensate for underexposure caused by fast shutter speed. The pixel size (s) are approximately $1.0 \mu\text{m}$ for S1 and $2.4 \mu\text{m}$ for S2. A summary of the camera parameters is provided in Table 2.

Working Distance. Due to the limited power of ToF emitters (Class I [51]) and underexposure caused by fast shutter speed, it is challenging to detect speckle patterns beyond a certain range. Additionally, the target surface’s materials and the camera’s optical characteristics further influence the system’s ability to capture speckle patterns. In our feasibility study, with the use of reflective patches, S2 is able to sense speckle patterns in the range of 5~200 cm. We repeat this feasibility study on S1 by capturing speckle patterns at distances of 5~400 cm under two focus settings (0.0 and 1.0) and analyzing the resulting speckle diameter and intensity. The results are shown in the Fig. 15. S1 could detect speckle patterns only within a shorter range of 5~100 cm. To demonstrate the robustness of PocketVib, we evaluate the system within the effective working ranges of both setups (S1: 5~100 cm, S2: 5~200 cm). The default distances are set to 5 cm for S1 and 20 cm for S2.

Table 2: Camera parameters of S1 and S2.

Setup	f (mm)	t_d (μs)	Resolution	Shutter Speed	ISO	s (μm)
S1	52	11.4	1920×1080	$1/2000$	2203	1.0
S2	28	10.4	3920×2160	$1/2000$	3200	2.4

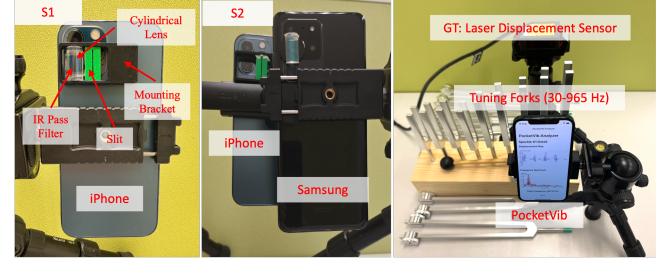


Figure 12: Experimental setup.

6.2 Evaluation on Vibration Measurement

6.2.1 Performance under Different Amplitudes. To evaluate PocketVib’s vibration detection performance across varying amplitudes, we use a 288.86 Hz tuning fork coupled to a bone-conduction speaker, which allows for precise amplitude control through volume adjustment. Since attaching the speaker alters the tuning fork’s natural frequency, we first measure its actual frequency and average amplitude across different volume levels (Fig. 13a). The results show that the frequency remains stable at 288.00 Hz, while the vibration amplitude is controllably varied between $0.2 \sim 2 \mu\text{m}$. This micron-level range falls within the lower end of amplitudes observed in real-world industrial monitoring scenarios [33, 39]. Larger amplitudes (corresponding to greater speckle motion) are inherently easier to detect. For each amplitude setting, we repeat the vibration measurement 10 times. The results are summarized in Fig. 13b. Except for S1, which fails to estimate the frequency at the lowest average amplitude of $0.2 \mu\text{m}$, all other configurations achieve frequency errors within 0.2 Hz.

6.2.2 Performance under Different Frequencies. To evaluate PocketVib’s vibration detection performance across a wide frequency range, we conduct experiments using 13 tuning forks with frequencies spanning from 30 Hz to 965 Hz. For each tuning fork, we perform 10 repeated vibration measurements. The vibrations are induced by manually striking the tuning fork with a rubber rod. The detection results are shown in Fig. 13c. The average frequency errors remain below 0.6 Hz across the entire frequency range for both setups. We also report the average vibration amplitudes for each experiment, which consistently remain at the micron level.

6.2.3 Performance on Multi-point Vibration Measurements. We evaluate the multi-point measurement capability of PocketVib (S1) by projecting a row of laser spots onto different

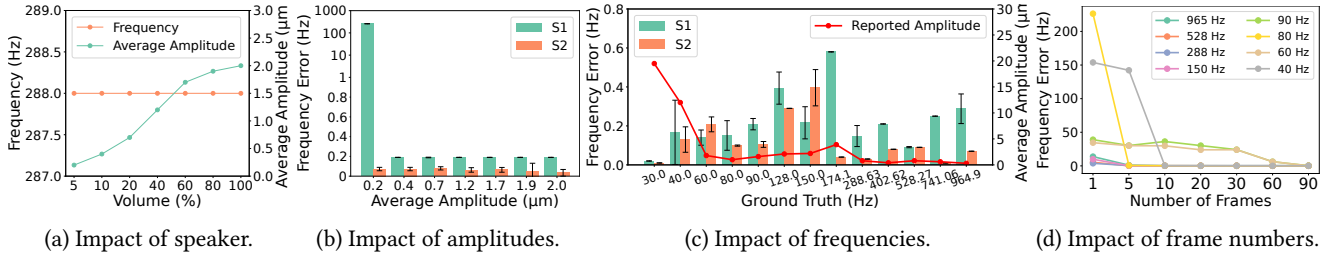


Figure 13: Performance under varying amplitudes (a, b), frequencies (c), and frame numbers (d).

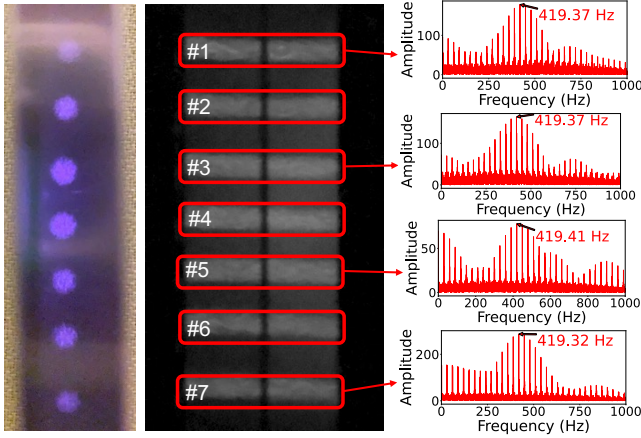


Figure 14: Multi-point vibration measurements.

positions of a 418.96 Hz tuning fork. The tuning fork is struck manually, with an average vibration amplitude of $1.0 \mu m$. As shown in Fig. 14, S1’s rolling shutter camera simultaneously captures up to seven distinct speckles within its FOV, and the frequency spectra of speckles 1, 3, 5, and 7 are also displayed. The average frequency error is 0.39 Hz, with a maximum error of approximately 0.45 Hz. Notably, sidebands around the dominant component are also visible in the frequency spectra, which is a well-known artifact in camera-based vibration measurement systems and will be further analyzed in the following section.

6.2.4 Performance using Different Frame Numbers. In the previous experiments, each frequency analysis is based on three seconds of speckle images, from which time displacement maps are extracted and analyzed via the FFT. Since each frame captures a segment of the vibration signal, the required number of frames for accurate vibration analysis varies with frequency. To examine this relationship, we evaluate eight representative frequencies, with results shown in Fig. 13d. High-frequency vibrations (e.g., 965 Hz) require fewer frames for analysis, as each frame captures multiple vibration cycles. In contrast, low-frequency vibrations (e.g., 40 Hz) require more frames to ensure sufficient sampling

of complete cycles. Additionally, frequencies that are not harmonically related to the camera’s frame rate (30 Hz) can be reliably estimated with 10–20 frames due to significant inter-frame differences. However, harmonic frequencies (e.g., 60 Hz, 90 Hz) require more frames for accurate estimation due to the small inter-frame differences. Notably, the periodically non-uniform sampling, where each frame captures a segment of the vibration signal, introduces sidebands around the dominant component in the frequency spectrum (see Fig. 14). Mathematically, this sampling process is equivalent to convolving the original signal’s spectrum with a periodic impulse train corresponding to the camera’s frame rate. Nonetheless, the original signal’s spectral features remain preserved, ensuring the validity of the vibration analysis.

6.3 Impact of Practical Factors

We evaluate the impact of several practical factors that affect the real-world applicability of PocketVib. In the following experiments, we measure the vibration of a 288.86 Hz tuning fork driven by a bone-conduction speaker at 100% volume (average amplitude: $2 \mu m$). Each measurement is repeated 10 times. The default hardware setup is S1.

6.3.1 Impact of Distances. The practical working distance of PocketVib is influenced by the reflectivity of the target surface. The selected tuning fork is made of aluminum alloy, a material commonly found in industrial applications, and features a highly reflective surface [35]. This makes it suitable for investigating PocketVib’s capability at various distances (S1: 5~100 cm, S2: 5~200 cm). As shown in Fig. 16, S1 successfully captures micro-vibration up to 70 cm, while S2 achieves accurate measurements up to 200 cm, both with near-zero frequency error. Beyond these ranges, measurement fails due to insufficient speckle intensity, as evidenced by the intensity variation curves in Figs. 15 and 5b. Notably, S2 outperforms S1 across all distances. As shown in Figs. 15 and 5b, S2 exhibits a larger maximum speckle size at the same distance and a larger physical pixel size (see Table 2). Consequently, S2 achieves higher magnification of the same speckle, thereby providing greater motion sensitivity.

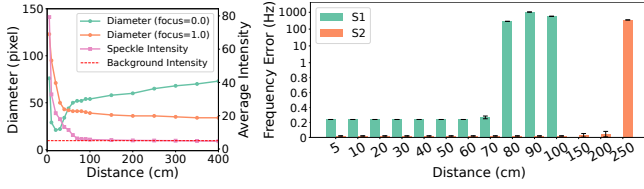
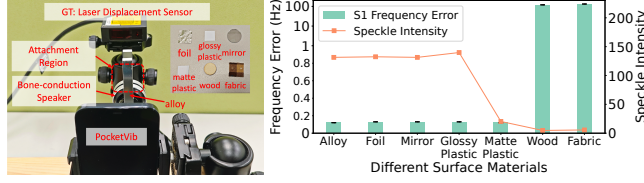


Figure 15: Feasibility study on S1.



(a) Experimental setup. (b) Speckle quality and frequency error.

Figure 17: Impact of surface materials.

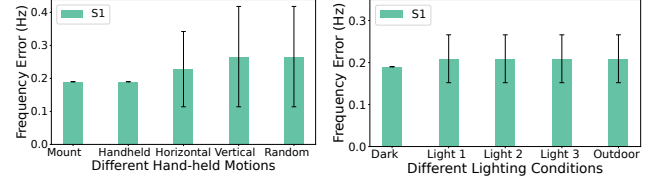
6.3.2 Impact of Surface Materials. The performance of PocketVib is significantly influenced by the optical properties of the target surface. Fig. 17a shows the experimental setup. We select four specular reflective materials, including aluminum alloy, aluminum foil, mirror, and glossy plastic, and three diffuse reflective materials, including matte plastic, wood, and fabric. The thin, uniform patches of these materials are attached to the tuning fork to be driven by its vibrations during testing. The actual resonant frequency and average amplitude are re-measured³. Although specular materials are typically considered smooth, they still generate speckle patterns for sensing, as speckle formation depends on optical roughness, which is determined by surface irregularities on the scale of the wavelength (~ 800 nm) [28]. To assess speckle quality, we adopted the speckle intensity as the metric, which reflects surface reflectivity. Results in Fig. 17b show that when the speckle intensity is too low (e.g., wood and fabric), no valid vibration signals can be extracted, whereas the others succeed. Notably, many common industrial equipment surfaces, including metal casings, machined components, and coated parts, exhibit high reflectivity [35], which supports the formation of strong speckle patterns.

6.3.3 Impact of Hand-held Motions. We evaluate the impact of different hand-held shakes, including static, horizontal, vertical, and random movements. For comparison, we also test PocketVib under a tripod-mounted static setup. Fig. 18a shows that the frequency errors under different types of hand-held motion remain minimal. Compared to the static condition, the average error only increases by 0.05 Hz, with a

³Measured frequency is consistently 288.00 Hz, and average amplitudes (μm) for these materials are: [2.0, 1.8, 1.5, 1.5, 1.45, 0.85, 1.53]

Table 3: System overhead of PocketVib (S1).

	CPU (%)	Memory (MB)	Power (mW)	Latency (ms)
Collection	11.6	8	3.772	33.33
Data Analysis	14.4	0.178	4.355	172.47



(a) Impact of hand-held motions. (b) Impact of lighting conditions.

Figure 18: Impact of hand-held and illumination.

maximum observed increase of 0.2 Hz. This robustness stems from two key factors. First, PocketVib employs a reference-free vibration extraction algorithm that relies solely on temporal relationships across rolling shutter rows. Second, modern smartphones equipped with optical image stabilization (OIS) significantly mitigate the impact of hand-held shakes.

6.3.4 Impact of Light Conditions. We evaluate the impact of various lighting conditions (Fig. 18b), including darkness, increasing indoor light intensity, and outdoor sunlight. The results show that PocketVib demonstrates robust performance across all scenarios, with an average frequency error increase of 0.02 Hz and a maximum error increase of 0.2 Hz compared to dark conditions. This robustness can be attributed to two aspects. First, the spot ToF camera employs an infrared laser (~ 800 nm), enabling reliable operation under a wide range of lighting conditions. Although sunlight includes near-infrared components, their lack of coherence ensures they do not disrupt the speckle patterns generated by the ToF camera. Second, the PocketVib Kit incorporates an IR-pass filter to suppress visible light interference, while the heat-induced infrared radiation ($3\sim 15 \mu\text{m}$) falls outside the detection range of the CMOS sensor ($350\sim 1100$ nm).

6.3.5 System Overhead. We evaluate the system overhead of the data collection and the on-device vibration analysis. Specifically, we capture a one-minute video containing 1,800 frames, and each frame has seven speckles (S1's maximum). Table 3 shows the measurement results of the power consumption, memory usage, and CPU usage for data collection and processing one frame on iPhone 12 Pro using PerfDog software [7]. The results show that PocketVib occupies modest resources on the smartphone, ensuring real-time feasibility on mobile devices.

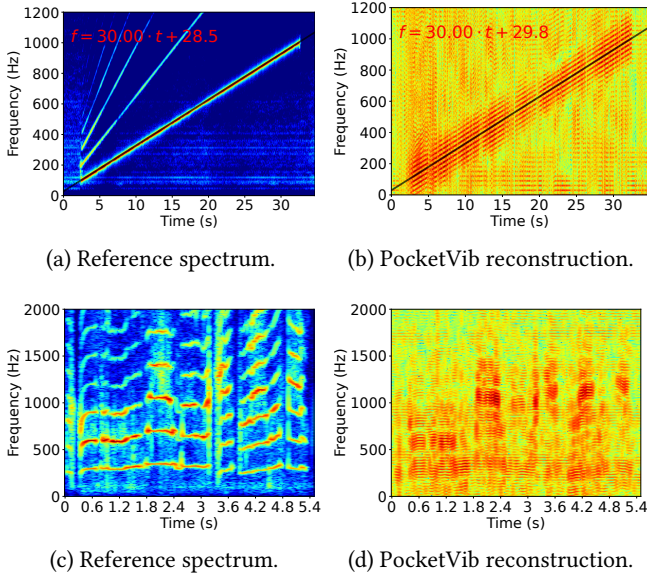


Figure 19: Impact of dynamic frequencies: (a, b) linear chirp; (c, d) child's singing.

6.4 Case Study

6.4.1 Dynamic Vibration Scenarios. In real-world environments, abnormal changes in vibration frequency often indicate equipment failures. To evaluate PocketVib's ability to capture such dynamic changes, we simulate varying vibration frequencies by placing a chip bag near a speaker that plays a chirp signal sweeping from 100 Hz to 1000 Hz over 30 seconds at 60% volume. The observed vibration amplitude varies continuously (0~50 μm), depending on both the speaker volume and the object's response to the acoustic excitation. Figs. 19a and 19b show the spectrograms of ground truth and PocketVib's reconstruction results, respectively. The black lines in both figures represent the ground truth and estimated frequencies, which align closely. We further evaluate PocketVib's dynamic reconstruction capability using a child's singing audio as input. As shown in the Figs. 19c and 19d, PocketVib can reconstruct audio features. This demonstrates PocketVib's ability to sense dynamic and complex vibration patterns. However, since PocketVib captures only a segment of the vibration signal per frame, reconstructing an interpretable audio, comparable to microphone recordings, remains a challenge, and we leave it as future work.

6.4.2 Real-world Vibration Sources. To evaluate the practicality of PocketVib, we measure vibrations from various real-world sources, including a tumble dryer, an industrial motor, a smartphone, and two sonic motors placed side-by-side. The ground truth vibration frequencies and average amplitudes are measured using a laser displacement sensor. Fig. 20 shows the experimental setup, reconstructed time-series, and

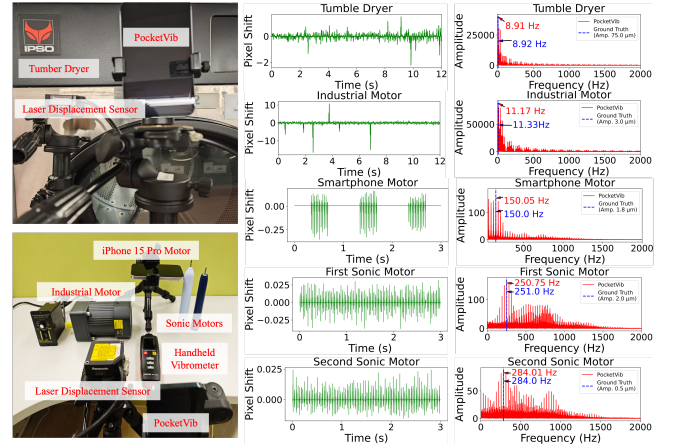


Figure 20: Real-world experimental setups (left), reconstructed time-displacement maps (middle), and corresponding frequency spectra (right).

corresponding frequency spectra (red: PocketVib's results, blue: ground truth). For the first two sources (low vibration frequencies), we collect 12 seconds of speckle images for reliable frequency analysis; for the remaining three higher-frequency sources, only three seconds suffice. The measured frequency errors are within 0.25 Hz, with most below 0.1 Hz, and the sidebands do not affect the main frequency components analysis. Notably, the smartphone's on-off vibration states also exhibit clearly distinguishable signal patterns.

7 Related Work

RF-based Vibrometry. RF-based vibrometry includes RFID-based [36, 55, 58–60] and mmWave-based [33, 42, 47, 48, 61] methods. While RFID-based approaches suffer from limited precision due to their long wavelengths, mmWave-based methods offer micron-level accuracy and support multi-object vibration sensing. However, these systems typically require costly setups and calibration, limiting their suitability for mobile applications. In contrast, PocketVib achieves comparable micron-level multi-point vibration sensing without the need for dedicated hardware or calibration, and can be fully implemented on a single smartphone.

Camera-based Vibrometry. Camera-based vibrometry captures vibrations by tracking pixel shifts in video frames [24, 26, 27]. These methods rely on high-frame-rate cameras, professional lenses, and advanced image processing algorithms, such as phase-based motion magnification [46], to enable high-frequency, micro-vibration measurements. However, this approach is costly and requires good lighting conditions as well as rich surface textures. Recently, MoiréVib [39] proposed a novel approach based on Moiré patterns, enabling micron-level vibration detection without requiring strong

assumptions about the object’s surface. However, MoiréVib relies on pre-designed interference patterns and requires two different frame rate videos to recover vibration signals up to 300 Hz. In contrast, PocketVib leverages the rolling shutter mechanism, enabling the detection of vibration signals up to 1 kHz even using only a single frame.

Speckle-based Vibrometry. Speckle-based vibrometry is an active camera-based technique that utilizes speckle patterns as textures to detect micro-vibrations [21, 30, 31, 43, 54, 63, 65]. Unlike Moiré patterns, speckle patterns are formed by illuminating optical rough surfaces with coherent light [28]. However, existing speckle-based vibrometry systems rely on specialized and bulky hardware. A recent work leverages smartphones’ ToF and camera systems for liquid classification [22]. However, its sampling rate is only 30 Hz, and it cannot reconstruct the object’s vibration frequency up to 1 kHz. In summary, PocketVib is the first to achieve multi-point kHz-frequency micro-vibration sensing using smartphone camera systems, providing a cost-effective solution for mobile high-precision vibration detection.

8 Limitation and Discussion

Detectable Frequency Range. PocketVib supports vibration sensing from 0 to 2000 Hz. The system can achieve up to 2000 Hz ($1/2000$ s shutter speed), with the upper detectable frequency limited by $\min(\frac{1}{t_d}, \frac{1}{t_{exp}})$. However, higher-frequency vibrations often have smaller amplitudes, making them harder to detect. For instance, we successfully measure a 965 Hz tuning fork with an average amplitude of $0.3 \mu m$.

Amplitude Consideration. As discussed in § 4.4, PocketVib does not support amplitude measurement, as it requires manual calibration of key parameters: l (the distance from the observation point to the rotation axis) and L_1 (the object-to-focal-plane distance). While L_1 can be directly measured by the spot ToF camera, determining l necessitates prior knowledge of the rotation axis. Therefore, PocketVib is a portable, calibration-free vibration sensing system for routine inspections, focusing on frequency features.

Extending Working Range. The working range of PocketVib depends on the amount of light reflected from the target surface. An intuitive approach is to increase the laser power; however, this is impractical on mobile devices due to power constraints and eye safety risks. As discussed in § 6.3.1, two feasible alternatives exist: using high-end sensors with larger apertures, or improving the reflectivity of the target surface. Another option is to increase the exposure time of rolling shutter cameras. For instance, increasing the exposure from $1/2000$ s to $1/500$ s can theoretically double the sensing range (the inverse-square law), but limits maximum detectable frequency below 500 Hz.

Hardware Modification. The current PocketVib prototype requires hardware modification to camera modules. However, we emphasize that this hardware modification is limited solely to the camera module itself. No modifications were made to the enclosure of the phone or the lens, so the device’s seal and water resistance remain unchanged. Moreover, users can directly attach an external IR-cut filter to the camera lens to restore color imaging functionality.

Practical Deployment of PocketVib. We note that some smartphones from manufacturers like Huawei [2], Honor [5], and Nokia [53] have included dedicated monochrome cameras that can receive near-infrared (NIR) light emitted by ToF. There are also emerging RGB-IR (Bayer-array) camera modules to capture both visible and infrared light simultaneously for better imaging quality. These sensors can receive NIR light without requiring additional hardware modification. Therefore, in the future, companies can deploy PocketVib along with other apps on pre-modified smartphones customized for industrial vibration monitoring.

The Cylindrical Lens. The PocketVib kit includes a cylindrical lens to stretch the speckle patterns. We explore the feasibility of eliminating this component in future designs. First, certain smartphones support attachable lenses with advanced optics, which can provide sufficient defocusing. Second, cameras with longer focal lengths are capable of achieving effective defocusing. However, current smartphones limit focal length due to size constraints; future devices may overcome this and eliminate the need for cylindrical lenses.

9 Conclusion

In this paper, we present PocketVib, the first mobile laser speckle vibrometry system that enables multi-point, kHz-frequency micro-vibration sensing by leveraging built-in sensors in smartphones, i.e., a spot ToF depth camera and a rolling shutter camera. PocketVib overcomes the limitations of existing vibration measurement methods, such as the need for physical attachment and complex setups, providing a cost-effective and portable solution for industrial routine inspections. Extensive experiments demonstrate that PocketVib accurately estimates multi-point vibration frequencies up to 1 kHz, with an average error of 0.6 Hz at 30 fps.

Acknowledgment

This paper is supported in part by the Research Grants Council (RGC) of Hong Kong under General Research Fund No. 14207123 and No. 14213525. We are grateful to the shepherd and anonymous reviewers for their insightful comments and suggestions.

References

- [1] 2008. SciPy API. https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html.
- [2] 2018. Huawei Unveils the HUAWEI P20 and HUAWEI P20 Pro. <https://www.huawei.com/en/news/2018/3/huawei-p20-p20-pro/>.
- [3] 2019. Introducing Multi-Camera Capture for iOS. <https://developer.apple.com/videos/play/wwdc2019/249/>.
- [4] 2020. A discussion about DTOF technology of Apple iPad Pro. <http://abax-sensing.com/ennewsInfo-3-9.html>.
- [5] 2020. HONOR Magic3 Series. <https://www.honor.com/global/phones/honor-magic3-pro-series/>.
- [6] 2022. ARKit 6. <https://developer.apple.com/augmented-reality/arkit/>.
- [7] 2022. PerfDog. <https://perfdog.qq.com/>.
- [8] 2023. ISOCELL Vizion 63D. <https://semiconductor.samsung.com/news-events/news/samsung-unveils-two-new-isocell-vizion-sensors-tailored-for-robotics-and-xr-applications/>.
- [9] 2023. Night Vision Camera. <https://zh.ifixit.com/Guide/iPhone+DIY+Night+Vision+Camera+and+Massive+External+Battery+Pack/147886>.
- [10] 2024. AVFoundation. <https://developer.apple.com/av-foundation/>.
- [11] 2024. Depth Camera D435i. <https://www.intelrealsense.com/depth-camera-d435i/>.
- [12] 2024. Limitations of smartphone cameras. <https://pixelcraft.photo.blog/2024/05/10/qa-limitations-of-smartphone-cameras-i/>.
- [13] 2024. VD55H1. <https://www.st.com/en/imaging-and-photonics-solutions/vd55h1.html>.
- [14] 2025. Beeware. <https://beeware.org/>.
- [15] 2025. Fluke 805 Vibration Meter. <https://www.fluke.com/en/product/mechanical-maintenance/vibration-analysis/fluke-805>.
- [16] 2025. Laser Doppler Vibrometer. <https://wavelength-oe.com/laser-doppler-vibrometer/>.
- [17] 2025. Portable Vibration Meters Market. <https://www.linkedin.com/pulse/portable-vibration-meters-market-size-type-application-lgsze>.
- [18] 2025. Vibration Monitoring Market (2025-2030). <https://www.grandviewresearch.com/industry-analysis/vibration-monitoring-market-report>.
- [19] 2025. Why depth of field is so important. <https://www.better-digital-photo-tips.com/depth-of-field.html>.
- [20] 2026. Typical bearing defects and spectral identification. <https://power-mi.com/content/typical-bearing-defects-and-spectral-identification>.
- [21] Sammy Apsel, Vika Tarle, Michal Yemini, Zeev Zalevsky, and Nisan Ozana. 2024. Rolling-shutter laser speckle analysis in bio-photonics. In *Biomedical Spectroscopy, Microscopy, and Imaging III*, Vol. 13006. SPIE, 163–166.
- [22] Justin Chan, Ananditha Raghunath, Kelly E Michaelsen, and Shyam-nath Gollakota. 2022. Testing a drop of liquid using smartphone LiDAR. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.
- [23] Subimal Bikash Chaudhury, Mainak Sengupta, and Kaushik Mukherjee. 2014. Vibration monitoring of rotating machines using MEMS accelerometer. *International journal of scientific engineering and research* 2, 9 (2014), 5–11.
- [24] Justin G Chen, Abe Davis, Neal Wadhwa, Frédo Durand, William T Freeman, and Oral Büyüköztürk. 2017. Video camera-based vibration measurement for civil infrastructure applications. *Journal of Infrastructure Systems* 23, 3 (2017), B4016013.
- [25] John Christopher Dainty. 2013. *Laser speckle and related phenomena*. Vol. 9. Springer science & business Media.
- [26] Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. 2015. Visual vibrometry: Estimating material properties from small motion in video. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5335–5343.
- [27] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. 2014. The visual microphone: Passive recovery of sound from video. *ACM Trans. Graph.* (2014).
- [28] Joseph W Goodman. 2007. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers.
- [29] Lixing He, Haozheng Hou, Shuyao Shi, Xian Shuai, and Zhenyu Yan. 2023. Towards bone-conducted vibration speech enhancement on head-mounted wearables. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 14–27.
- [30] Huanhuan Hong, Jiajia Liang, Liza Deng, Wei Guo, and Xiaozhong Wang. 2023. Extreme detectable vibration frequency limited by rolling shutter camera imaging of laser speckles. *Optics Letters* 48, 15 (2023), 3837–3840.
- [31] Xueyuan Huang, Wei Guo, Rui Yu, and Xiaozhong Wang. 2022. Real-time high sensibility vibration detection based on phase correlation of line speckle patterns. *Optics & Laser Technology* 148 (2022), 107759.
- [32] A. Janssen, R. Veldhuis, and L. Vries. 1986. Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34, 2 (1986), 317–330.
- [33] Chengkun Jiang, Junchen Guo, Yuan He, Meng Jin, Shuai Li, and Yunhao Liu. 2020. mmVib: micrometer-level vibration measurement with mmwave radar. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–13.
- [34] Min-Sun Keel, Daeyun Kim, Yeomyung Kim, Myunghan Bae, Myoungoh Ki, Bumsik Chung, Sooho Son, Hoyong Lee, Heeyoung Jo, Seung-Chul Shin, et al. 2021. 7.1 A 4-tap 3.5 μm 1.2 Mpixel indirect time-of-flight CMOS image sensor with peak current mitigation and multi-user interference cancellation. In *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, Vol. 64. IEEE, 106–108.
- [35] Par Kierkegaard. 1996. Reflection properties of machined metal surfaces. *Optical Engineering* 35, 3 (1996), 845–857.
- [36] Ping Li, Zhenlin An, Lei Yang, and Panlong Yang. 2019. Towards physical-layer vibration sensing with rfids. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 892–900.
- [37] Jian Liu, Chen Wang, Yingying Chen, and Nitesh Saxena. 2017. VibWrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 73–87.
- [38] Yan Long, Pirouz Naghavi, Blas Kojusner, Kevin Butler, Sara Rampazzi, and Kevin Fu. 2023. Side eye: Characterizing the limits of pov acoustic eavesdropping from smartphone cameras with rolling shutters and movable lenses. In *2023 IEEE Symposium on Security and Privacy (SP)*. IEEE, 1857–1874.
- [39] Jingyi Ning, Zhihao Yan, Zhaowei Wu, Lei Xie, Chuyu Wang, Yingying Chen, Baoli Ye, and Sanglu Lu. 2024. MoiréVib: Micron-level Vibration Detection based on Moiré Pattern. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 1393–1407.
- [40] Luc Oth, Paul Furgale, Laurent Kneip, and Roland Siegwart. 2013. Rolling shutter camera calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1360–1367.
- [41] Alessandro Sabato, Christopher Niezrecki, and Giancarlo Fortino. 2016. Wireless MEMS-based accelerometer sensor boards for structural vibration monitoring: A review. *IEEE Sensors Journal* 17, 2 (2016), 226–235.
- [42] Hailan Shanbhag, Sohrab Madani, Akhil Isanaka, Deepak Nair, Saurabh Gupta, and Haitham Hassanieh. 2023. Contactless material identification with millimeter wave vibrometry. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 475–488.

- [43] Mark Sheinin, Dorian Chan, Matthew O'Toole, and Srinivasa G Narasimhan. 2022. Dual-shutter optical vibration sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16324–16333.
- [44] Seung-chul Shin, Myeonggyun Kye, Il-pyeong Hwang, Taemin An, Kyu-Min Kyung, Duhyeon Kwak, Hakbeom Jang, Hogen Kim, Jaeh An, Sunhwa Lee, et al. 2021. Indirect-ToF system optimization for sensing range enhancement with patterned light source and adaptive binning. In *Proc. Int. Image Sensor Workshop (IISW)*. 142–145.
- [45] Qi Tian and Michael N Huhns. 1986. Algorithms for subpixel registration. *Computer Vision, Graphics, and Image Processing* 35, 2 (1986), 220–233.
- [46] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. 2013. Phase-Based Video Motion Processing. *ACM Trans. Graph. (Proceedings SIGGRAPH 2013)* 32, 4 (2013).
- [47] Chao Wang, Feng Lin, Hao Yan, Tong Wu, Wenyao Xu, and Kui Ren. 2024. {VibSpeech}: Exploring practical wideband eavesdropping via bandlimited signal of vibration-based side channel. In *33rd USENIX security symposium (USENIX Security 24)*. 3997–4014.
- [48] Teng Wei and Xinyu Zhang. 2015. mtrack: High-precision passive tracking using millimeter wave radios. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 117–129.
- [49] Wikipedia. 2025. Infrared cut-off filter. https://en.wikipedia.org/wiki/Infrared_cut-off_filter.
- [50] Wikipedia. 2025. Laser Doppler vibrometer. https://en.wikipedia.org/wiki/Laser_Doppler_vibrometer.
- [51] Wikipedia. 2025. Laser safety. https://en.wikipedia.org/wiki/Laser_safety.
- [52] Wikipedia. 2026. Lens. <https://en.wikipedia.org/wiki/Lens>.
- [53] Wikipedia. 2026. Nokia 9 PureView. https://en.wikipedia.org/wiki/Nokia_9_PureView.
- [54] Nan Wu and Shinichiro Haruyama. 2021. The 20k samples-per-second real time detection of acoustic vibration based on displacement estimation of one-dimensional laser speckle images. *Sensors* 21, 9 (2021), 2938.
- [55] Binbin Xie, Jie Xiong, Xiaojiang Chen, and Dingyi Fang. 2020. Exploring commodity rfid for contactless sub-millimeter vibration sensing. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 15–27.
- [56] Zhiyuan Xie, Xiaomin Ouyang, Xiaoming Liu, and Guoliang Xing. 2021. UltraDepth: Exposing high-resolution texture from depth cameras. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 302–315.
- [57] Zhiyuan Xie, Xiaomin Ouyang, Li Pan, Wenrui Lu, Guoliang Xing, and Xiaoming Liu. 2023. Mozart: A mobile tof system for sensing in the dark through phase manipulation. In *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*. 163–176.
- [58] Lei Yang, Yao Li, Qiongzhen Lin, Huanyu Jia, Xiang-Yang Li, and Yunhao Liu. 2017. Tagbeat: Sensing mechanical vibration period with cots rfid systems. *IEEE/ACM transactions on networking* 25, 6 (2017), 3823–3835.
- [59] Lei Yang, Yao Li, Qiongzhen Lin, Xiang-Yang Li, and Yunhao Liu. 2016. Making sense of mechanical vibration period with sub-millisecond accuracy using backscatter signals. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. 16–28.
- [60] Panlong Yang, Yuanhao Feng, Jie Xiong, Ziyang Chen, and Xiang-Yang Li. 2020. Rf-ear: Contactless multi-device vibration sensing and identification using cots rfid. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 297–306.
- [61] Yanni Yang, Huafeng Xu, Qianyi Chen, Jiannong Cao, and Yanwen Wang. 2023. Multi-Vib: Precise multi-point vibration monitoring using mmWave radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (2023), 1–26.
- [62] Yang Yu, Xiangju Qin, Shabir Hussain, Weiyan Hou, and Torben Weis. 2022. Pedestrian counting based on piezoelectric vibration sensor. *Applied Sciences* 12, 4 (2022), 1920.
- [63] Zeev Zalevsky, Yevgeny Beiderman, Israel Margalit, Shimshon Gingold, Mina Teicher, Vicente Mico, and Javier Garcia. 2009. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Opt. Express* 17, 24 (2009), 21566–21580.
- [64] Tianyuan Zhang, Mark Sheinin, Dorian Chan, Mark Rau, Matthew O'Toole, and Srinivasa G Narasimhan. 2023. Analyzing physical impacts using transient surface wave imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4339–4348.
- [65] Yang Zhang, Gierad Laput, and Chris Harrison. 2018. Vibrosight: Long-range vibrometry for smart environment sensing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 225–236.