Computational Speckle Pattern Interferometry

Shengxi Wu[1] Sophia Yang[2] Dorian Chan[1] Matthew O'Toole[1]
[1]Carnegie Mellon University [2]University of Toronto

Abstract

Visually imperceptible surface deformations encode rich information about a scene, from the mechanical properties of an object to the acoustic vibrations present in the surrounding environment. Optical interferometric techniques can reveal these subtle changes, typically by capturing a sequence of measurements to perform temporal phase shifting. In this paper, we introduce Computational Speckle Pattern Interferometry (CSPI), a novel single-shot approach to estimating per-pixel displacement and motion. Our key insight is that the image formation model for speckle pattern interferometry can be decomposed into spatial and temporal factors, each represented as a vector. After calibrating for the spatial term, we recover the scene dynamics using a reconstruction algorithm modeled after the classic Horn-Schunck method for estimating optical flow. Unlike traditional interferometric methods, CSPI requires no precision instrumentation to perform phase stepping. We demonstrate its effectiveness by measuring per-pixel displacements and motions at sub-micrometer scales, visualizing high-frequency vibrations of a tuning fork and a Chladni plate, and recovering sound indirectly from these vibrations.

1. Introduction

Subtle surface deformations—whether from vibration or mechanical stress—reveal critical information about how objects interact with their environment. Capturing these minute changes provides insights into material and structural properties [7, 21, 31], manufacturing and design defects [11, 28], human health and behavior [14, 15, 22, 30, 32], sound and acoustics [2, 17, 21, 33], and other phenomena that are invisible to the naked eye.

How can one measure these microscale displacements? Current solutions span contact-based mechanical methods and non-contact optical techniques [3, 21, 28], all of which involve trade-offs between sensitivity, speed, and ease of use. Passive methods observe intensity variations to recover the subtle, often subpixel, changes that occur in image space [3, 25]. Active methods (*e.g.*, approaches based on coherent illumination) can optically enhance sensitivity to this displacement. For example, by imaging laser spots on an object with a defocused camera, tiny tilts of the object surface can be captured with sufficient sensitivity to resolve sound [2, 21, 30], human heartbeats [30], and material information [11, 31]. There are interferometers that make use of the Doppler effect to measure translational displacements at even finer resolution, and such devices are now widely used in industry and science thanks to their robustness and stability [27]. However, these aforementioned setups often only measure deformation at *sparse* locations on the object surface. Capturing a full, dense image of surface deformation would then require scanning [17, 32], resulting in prohibitively long capture times for any transient event of interest. Alternatively, prior work analyzes nonlinear changes in speckle patterns to recover dense displacement maps [19, 20], although relating the observations to quantitative displacement measurements remains challenging.

A classical approach called *electronic speckle pattern interferometry* (ESPI) allows for dense, full-field measurement of surface deformations [28]. In its standard formulation, an interferometric system captures and compares two laser speckle images before and after deformation. Changes in the observed intensity indicate deformation at the corresponding pixels. However, this imaging technique is not without its limitations. First, extracting surface deforma-

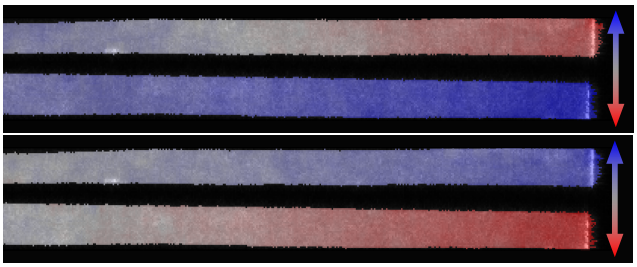


Figure 1. **Recovered displacement of a vibrating tuning fork.** We unwrap the phase to recover per-pixel displacement. The top and the bottom prong move in equal and opposite directions, showing the fundamental resonance mode. The tip of the tuning fork is on the right-hand side of the image.

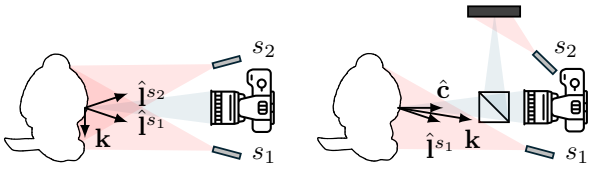
tion estimates from speckle images is an inherently noisy process, which can require heavy post-processing to account for sensor noise and the speckle itself [9]. While estimating both speckle phase and amplitude at every pixel improves measurement accuracy, it typically requires precise phase stepping or wavefront sensing [10, 28]. Such approaches often necessitate complex calibration, specialized hardware, and multiple captures per measurement, reducing real-world practicality. Second, ESPI often assumes the object is static, or frames are captured instantaneously. Unfortunately, in dynamic settings, limited laser power often necessitate longer exposures. While there is work that uses long exposures to analyze periodic motion in both ESPI [13, 26] and time-averaged holography [16], there is little work exploring more general motions.

To address these limitations, we revisit this classical approach and propose what we call **Computational Speckle Pattern Interferometry (CSPI)**: a modern computational framework for dynamic ESPI. Our key insight is that we can express the speckle image formation model as the inner product between two vectors representing speckle and displacement parameters. Armed with this observation, our method first factorizes a speckle video into per-pixel speckle and coarse displacement parameter matrices, much akin to established techniques in uncalibrated photometric stereo [29]; this entirely eliminates the need for manual phase stepping or wavefront modulation. Next, with inspiration from optical flow algorithms, our novel reconstruction procedure employs these parameter matrices to robustly estimate per-pixel displacement from a single photo. Finally, we show that this matrix analysis accounts for finite exposure times, and unlocks per-frame motion estimates.

To validate our findings, we build an interferometric setup, with which we measure surface deformations across a wide number of settings. We verify that our CSPI formulation indeed recovers displacement and velocity in a controlled setting on a translation stage. To demonstrate the applicability of CSPI for mechanical analysis, we use it to capture the spatially varying amplitude of vibrational modes on a tuning fork and a Chladni plate. In the context of acoustics, we apply CSPI for acoustic recovery from the vibrations on a bag of chips. In the supplement, we also use our CSPI formulation to estimate differential displacement (*e.g.*, tilt) via shearography, an approach analogous to ESPI.

Contributions. Our contributions include:

- A factored representation of the image formation model used in ESPI, capable of modeling the effect of displacement and motion on speckle interferograms.
- A calibration procedure that eliminates the need for manual phase stepping or wavefront modulation.
- Reconstruction procedures for extracting per-pixel displacement and motion information from a single photo, based on classic optical flow reconstruction procedures.
- Demonstrations on various scenes, showcasing the ability to capture high-frequency vibrations and subtle deformations with high sensitivity and spatial resolution.



(a) In-plane arrangement (b) Out-of-plane arrangement

Figure 2. **Arrangements for speckle pattern interferometry.**

Limitations. Our method requires active illumination, which may not be feasible in all scenarios. Additionally, while our approach reliably captures small displacements, it may be less effective for large displacements that exceed the phase wrapping limits inherent in interferometric techniques. Our method also performs best when there is minimal environmental vibration, which may require vibration isolation equipment. When there is too much movement or when the speckle decorrelates, re-calibration is required. We also impose a prior in our recovery step where that requires our scene object to displace smoothly across space.

2. Speckle Pattern Interferometry

There are several optical arrangements used in ESPI, including those shown in Fig. 2. In this section, we provide an overview of their image formation model.

Illuminating a target with coherent laser light forms a complex-valued field ψ_j incident on camera pixel j. Suppose that the region observed by pixel j is displaced according to a (small) 3D vector $\mathbf{\Delta x}_k$ at frame k; for simplicity, we will assume $\mathbf{\Delta x}_k$ is constant across space for the time being. This displacement alters the optical path length and introduces a phase shift to the resulting field:

$$\psi_j e^{-i\frac{2\pi}{\lambda}(\hat{\mathbf{l}}_j \cdot \mathbf{\Delta x}_k)} e^{-i\frac{2\pi}{\lambda}(\hat{\mathbf{c}}_j \cdot \mathbf{\Delta x}_k)}, \tag{1}$$

where λ is the wavelength of light, $\hat{\mathbf{l}}_j$ is the unit-length illumination direction from the perspective of the scene point observed at pixel j, and $\hat{\mathbf{c}}_j$ is the unit-length viewing direction. This expression accounts for the change in optical path length that occurs between the light source and target, as well as between the target and camera. However, because cameras only measure the intensity of the field $|\psi_j|^2$, any change in phase is not directly observable.

The goal of speckle pattern interferometry is to detect this phase change through interference. Fig. 2 highlights two interferometric setups used to infer (a) in-plane displacement or (b) out-of-plane displacement. We also dis-

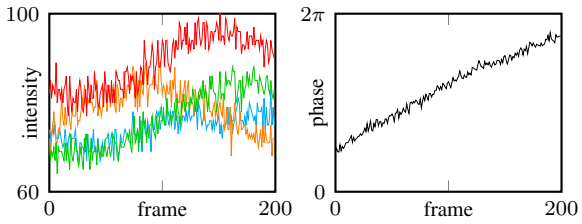


Figure 3. **Real measurements of an object subject to linear translation. (Left)** Raw intensity measurements at four neighboring pixels. Assuming linear motion, the intensity at each pixel varies sinusoidally in accordance with Eq. (2). **(Right)** Recovered phase at a particular pixel; note that we process each frame individually.

cuss the optical arrangement for shearography in the supplement, used to measure spatial derivatives of displacement (*e.g.*, tilt).

2.1. Optical arrangements

In-plane. Suppose two mutually coherent light sources illuminate a target from two different directions $\hat{\mathbf{l}}^{s_1}$ and $\hat{\mathbf{l}}^{s_2}$, producing two fields governed by Eq. (1). The intensity observed at every pixel is given by summing the two fields, and calculating the square of its magnitude:

$$l_{jk} = \underbrace{2|\psi_j^{s_1}\psi_j^{s_2\,*}|}_{a_j}\cos\Big(\underbrace{\angle\psi_j^{s_1}\psi_j^{s_2\,*}}_{b_j} - \phi_k\Big) + \underbrace{|\psi_j^{s_1}|^2 + |\psi_j^{s_2}|^2}_{c_j}$$

$$\text{where} \quad \phi_k = \frac{2\pi}{\lambda}\mathbf{k}\cdot\boldsymbol{\Delta}\mathbf{x}_k \quad \text{and} \quad \mathbf{k} = \hat{\mathbf{l}}^{s_1} - \hat{\mathbf{l}}^{s_2}. \quad (2)$$

Note that the intensity l_{jk} at pixel j depends on the displacement vector $\boldsymbol{\Delta}\mathbf{x}_k$ at frame k. Moreover, the image formation model is most sensitive to displacements along $\mathbf{k}$ and is insensitive to displacements in the orthogonal directions. The vectors $\hat{\mathbf{l}}^{s_1}$ and $\hat{\mathbf{l}}^{s_2}$ also have important implications on the sensitivity of the system. For example, increasing the angle between the two vectors increases the magnitude $\|\mathbf{k}\|$, which increases sensitivity to small displacements. We can also change the orientation of the vector $\mathbf{k}$ by changing the positions of the two light sources. For instance, when placed on either side of the camera, the system becomes sensitive to in-plane (lateral) motion. Fig. 3 provides an example of measured intensity variations observed at four pixels in response to linear in-plane motion.

Out-of-plane. The out-of-plane arrangement uses a single coherent light source to illuminate the target, and interferes the resulting field with a (stationary) reference field unaffected by the target's motion. The resulting interference pattern is still governed by Eq. (2), with the exception that $\mathbf{k} = \hat{\mathbf{l}}^{s_1} + \hat{\mathbf{c}}$. For a light source positioned close to the camera's optical axis, the measurements become sensitive to out-of-plane motion.

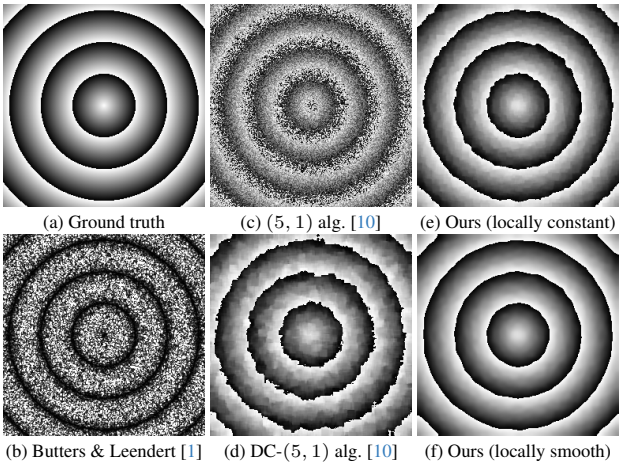


Figure 4. **Simulation of single-image methods for estimating displacement. (a)** Ground truth displacement represented as wrapped phase. **(b)** Traditional ESPI approach captures two images, and computes the difference between two speckle images. Each "fringe" occurs when the phase change is an integer multiple of 2π. **(c)** Kao *et al.* [10] proposed a per-pixel phase estimate that suffers from ambiguities. **(d)** These ambiguities are addressed using their correlation-based procedure, but still suffers from low phase map quality. **(e)** We produce higher-quality displacement measurements, by solving a linear system that assumes locally constant phasors. **(f)** We further refine this solution by assuming smooth phasors and optimizing with Eq. (12).

2.2. Traditional displacement recovery

In the classical dual-exposure ESPI setting, the differences between the measurements before and after deformation can be written as [1]:

$$l_{jk} - l_j = a_j\cos(b_j - \phi_k) - a_j\cos(b_j) \quad (3)$$

$$= 2a_j\sin\left(b_j - \frac{\phi_k}{2}\right)\sin\left(\frac{\phi_k}{2}\right), \quad (4)$$

where l_j represents the intensity before deformation and l_{jk} represents the intensity after deformation at frame k. It can be seen that when ϕ_k equals a multiple of 2π, the difference will yield zero, producing a clear "fringe" that can be used for visualization; see Fig. 4(b). However, for other values of ϕ_k, the measurement is contaminated by the speckle parameters a_j and b_j.

Instead of directly comparing intensities, another potential solution is to measure and compare the phase before and after deformation. Perhaps the most straightforward way to do so is to perform a temporal phase shifting, where controlled phase offsets are applied to one of the coherent beams used for interference [28]. This increases the sensitivity tenfold, but largely limits the technique to static

scenes as it requires capturing multiple images of the displaced object.

To keep the approach suitable for dynamic scenes, Kao *et al*. [10] proposed a way to estimate displacement that only requires five phase-shifted images before deformation and a single image after deformation, referred to as a $(5, 1)$ algorithm. However, their approach uses extra optical components to produce displacements with precision on the order of fractions of a wavelength, does not account for motion during exposure, and produces comparatively low-quality phase maps; see Fig. 4(c)-(d).

To sidestep these issues, another approach is to directly analyze temporal intensity variations at each pixel, and reconstruct phase via temporal priors [4, 12]. However, such methodology may impose unreasonable assumptions on the actual underlying motion, *e.g*., narrowband displacement [12] or linear motion across multiple frames [4].

Similar to Kao *et al*. [10], we propose an approach to estimating displacement from a single image after deformation. In addition to being more accurate (see Fig. 4(e)-(f)), our solution also recovers the amount of motion during a camera's exposure and does not require any precision optical equipment for controlled phase shifting.

3. Factored Representation

A key observation is that the image formation model for speckle pattern interferometry can be decomposed into two components: one that varies spatially and another that varies temporally. As shown in Eq. (2), the in-plane and out-of-plane interferometric arrangements produce intensities governed by:

$$l_{jk} = a_j \cos(b_j - \phi_k) + c_j \,, \tag{5}$$

where a_j, b_j, and c_j vary with pixel j, and ϕ_k is the phase induced by object displacement in frame k. This expression can be rewritten as the inner product of two vectors $\mathbf{v}_j, \mathbf{u}_k \in \mathbb{R}^3$:

$$\begin{aligned} l_{jk} &= a_j \cos(b_j)\cos(\phi_k) + a_j \sin(b_j)\sin(\phi_k) + c_j \\ &= \underbrace{\begin{bmatrix} a_j \cos(b_j) & a_j \sin(b_j) & c_j \end{bmatrix}}_{\mathbf{v}_j} \underbrace{\begin{bmatrix} \cos(\phi_k) & \sin(\phi_k) & 1 \end{bmatrix}^\top}_{\mathbf{u}_k} . \end{aligned} \tag{6}$$

The vector $\mathbf{v}_j$ represents the local speckle-dependent appearance at pixel j, while $\mathbf{u}_k$ represents the effect of displacement within measurement k.

Estimating displacement. Suppose for now that the vector $\mathbf{u}_k$ is already known. To recover the (wrapped) phase shift ϕ_k from $\mathbf{u}_k$, we simply define a complex-valued phasor $\vec{u}_k = u_{k1} + i u_{k2} = \cos(\phi_k) + i \sin(\phi_k)$ and then compute its argument, *i.e*., $\angle \vec{u}_k = \phi_k$. This phasor representation also happens to be useful for analyzing what happens when the phase varies during the camera's exposure.

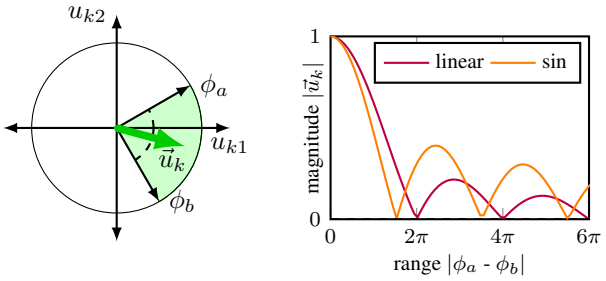


Figure 5. **Estimating motion. (Left)** An object moves between from ϕ_a to ϕ_b during an exposure. The phasor produced in response is an average of unit-length phasors, with a phase that represents the average displacement and a magnitude that encodes motion during the exposure. **(Right)** Simulation of phasor magnitude for linear and sinusoidal motion. When an object is stationary (*i.e*., $\phi_a = \phi_b$), the phasor has unit length. As the object moves, the average of unit-length phasors results in a smaller magnitude. For example, for linear motion, the magnitude becomes zero when averaging phasors uniformly distributed from angles 0 to 2π.

Estimating motion. Until now, we have treated each intensity measurement as an instantaneous snapshot in time. In practice, the object may move during the camera's finite exposure period:

$$l_{jk} = \int \mathbf{v}_j \mathbf{u}_k(t)^\top \, \mathrm{d}t = \mathbf{v}_j \left(\int \mathbf{u}_k(t) \, \mathrm{d}t \right)^\top = \mathbf{v}_j \bar{\mathbf{u}}_k^\top \,. \tag{7}$$

This corresponds to time-averaging multiple speckle patterns, which has the effect of reducing speckle contrast. Despite this, the image formation model can still be factorized, as demonstrated in Eq. (7).

The effect of motion during a camera's exposure can be interpreted naturally in the phasor domain. Specifically, the recovered phasor $\vec{u}_k$ now becomes the temporal average of instantaneous unit-length phasors over time. Assuming the phase changes linearly from ϕ_a to ϕ_b during the exposure (*i.e*., constant-velocity motion), the reconstructed phasor is

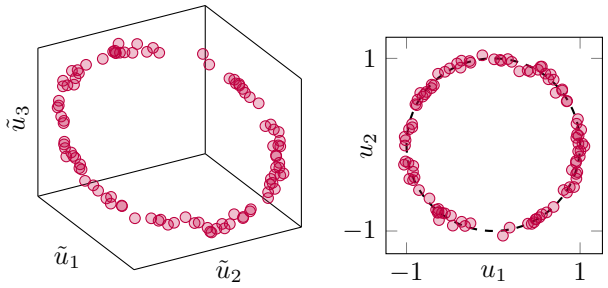
$$\vec{u}_k = \operatorname{sinc}\left(\frac{\phi_a - \phi_b}{2} \right) \exp\left(i \frac{\phi_a + \phi_b}{2} \right) . \tag{8}$$

Its angle,

$$\angle \vec{u}_k = \frac{1}{2}(\phi_a + \phi_b) \,, \tag{9}$$

corresponds to the average phase up to a π flip when $\operatorname{sinc}(\cdot)$ becomes negative, while the magnitude varies with the phase range $|\phi_a - \phi_b|$. Thus, the phasor magnitude provides information about motion during exposure; see Fig. 5.

For motion that oscillates sinusoidally between ϕ_a and ϕ_b an integer number of times during the exposure, the re-



(a) Rows of $\tilde{\mathbf{U}}$ (b) Rows of $\mathbf{U}$

Figure 6. **Estimating unknown phase shifts through factorization. (a)** After constructing matrix $\mathbf{L}$ and computing a rank-3 factorization $\tilde{\mathbf{V}}\tilde{\mathbf{U}}^\top$, the rows of matrix $\tilde{\mathbf{U}}$ represent points that, in the absence of noise, lie on an ellipse in 3D space. **(b)** Our proposed procedure maps these points to the unit circle via a transform $\mathbf{Q}$.

constructed phasor becomes

$$\vec{u}_k = J_0\left(\frac{\phi_a - \phi_b}{2}\right)\exp\left(i\frac{\phi_a + \phi_b}{2}\right), \qquad (10)$$

where $J_0(x)$ is the zero-order Bessel function of the first kind [26]. See supplemental materials for details.

In other words, displacement is encoded primarily in the phasor angle, while motion during exposure is reflected in its magnitude. This interpretation therefore enables recovery of per-pixel displacement and motion, simply by recovering $\vec{u}_k$ and calculating the argument and magnitude of the phasor.

4. Displacement Recovery

In Sec. 4.1, we introduce a method for estimating the speckle-dependent appearance vector $\mathbf{v}_j$ at every pixel j from measurements l_{jk} with *unknown* phase shifts ϕ_k, thereby eliminating the need for specialized hardware. In Sec. 4.2, we explain how to use these calibrated speckle parameters to recover per-pixel displacement and motion measurements from a single image.

4.1. Displacement from unknown phase shifts

We propose a procedure for estimating both $\mathbf{v}_j$ and $\mathbf{u}_k$ from *unknown* phase shifts, inspired by the uncalibrated photometric stereo literature [29] and more recent work in phase-shifting interferometry [5, 6].

We start by capturing a set of speckle images in response to different global phase shifts ϕ_k (*e.g.*, achieved by manually translating one of the light sources by a small, but unknown, amount). Next, we construct a matrix $\mathbf{L}$ of size $J \times K$ with elements given by l_{jk}. Following from Sec. 3, the matrix $\mathbf{L}$ is rank 3, assuming a common phase shift for all pixels and no sensor noise. The matrix can therefore be factorized using the singular value decomposition (SVD) to yield the optimal rank-3 approximation with respect to the Frobenius norm:

$$\mathbf{L} = \mathbf{V}\mathbf{U}^\top, \qquad (11)$$

where the j^{th} row of matrix $\mathbf{V} \in \mathbb{R}^{J\times 3}$ is the vector $\mathbf{v}_j$, and the k^{th} row of matrix $\mathbf{U} \in \mathbb{R}^{K\times 3}$ is the vector $\mathbf{u}_k$.

However, the decomposition is not unique. In particular, for any invertible 3×3 matrix $\mathbf{Q}$, the matrices $\tilde{\mathbf{V}} = \mathbf{V}\mathbf{Q}$ and $\tilde{\mathbf{U}} = \mathbf{U}\mathbf{Q}^{-\top}$ are also a valid factorization. This ambiguity must be resolved to recover meaningful decomposition. To tackle this, there are two important constraints that can be used to resolve matrix $\mathbf{Q}$: $u_{k3} = 1$ and $u_{k1}^2 + u_{k1}^2 = 1$ (*i.e.*, points lie on a circle). The latter constraint assumes that there is no motion during the camera's exposure. In the supplement, we provide details of our proposed method to estimate $\mathbf{Q}$ to compute both $\mathbf{U}$ and $\mathbf{V}$; see Fig. 6 for an example. This procedure recovers the phases ϕ_k up to an unknown offset and sign—both of which are typically non-critical for deformation estimation and can be resolved through additional assumptions.

The factorization requires $J \geq 2$ and $K \geq 5$, which correspond to the minimum number of 2D measurements necessary to estimate the parameters of an ellipse. In practice, larger values for J and K further improve robustness to sensor noise.

4.2. Per-pixel displacement and motion recovery

After calibrating for vector $\mathbf{v}_j$ across all pixels j, we recover *per-pixel* displacement and motion from a single speckle image, producing a vector $\mathbf{u}_j$ at each pixel j (and dropping the subscript k for brevity). Sec. 3 represent a linear constraint that can be used to solve for the vector $\mathbf{u}_j$, or more specifically u_{j1} and u_{j2}. Because a single constraint is insufficient to solve for both elements directly, we further incorporate a spatial smoothness prior on these components. We then formulate and minimize the following global energy functional:

$$\min_{\{u_{j1},u_{j2}\}} \sum_j (\mathbf{v}_j \cdot \mathbf{u}_j - l_j)^2 + \alpha^2 \|\nabla \mathbf{u}_j\|^2 . \qquad (12)$$

Note that this formulation closely resembles the classic Horn-Schunck method for optical flow [8]. To initialize this procedure, we first solve a small linear system that assumes the components u_{j1} and u_{j2} are constant within a small neighborhood. We then run the same iterative scheme used by the Horn-Schunck method to minimize Eq. (12).

5. Implementation

Hardware. Our prototype system consists of a monochrome industrial camera (IDS UI3240CP-NIR) with a 16 mm $f/1.8$ lens, and a fiber-pigtailed Coherent Sapphire FP 548 nm laser that emits up to 300 mW of light. We use a single mode fused fiber optic splitter (Thorlabs

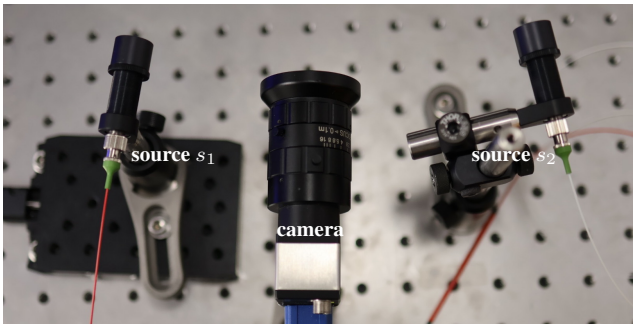


Figure 7. **Prototype setup for estimating in-plane motion.** For each experiment, we reposition the two sources to capture different types of measurements (*e.g.*, horizontal or vertical displacement).

TN532R5A1) to create two mutually coherent sources, and place concave lenses (Thorlabs LC1054) in front of both sources to narrow the emitted cone of light. We mount one of the sources on a linear stage (Thorlabs XR25C), which we manually adjust to produce unknown global phase offsets for the procedure outlined in Sec. 4.1.

Our experiments focus on the in-plane arrangement, where both sources point towards the target object; see Fig. 7 for a photo of our in-plane arrangement. In practice, we found that the out-of-plane arrangement to be *too* sensitive and severely affected by environmental noise, which could be addressed through active vibration isolation.

For certain experiments, a linear stage (Thorlabs XR25P) and motorized actuator (Thorlabs Z825B, KDC101) displaces an object by a known amount and direction. This single-axis translation stage can be controlled with fine step-size control of $29\,\mathrm{nm}$ with travel accuracy within $40\,\mu\mathrm{m}$. The stage also has functionality to query its current position when it is stationary.

Software. We calibrate our system by capturing images with a global change in phase, by manually adjusting the position of one light source using a translation stage. This induces a change in the optical path length from the light source to the target, resulting in a change in phase. The procedure from Sec. 4.1 uses the images to recover vectors $\mathbf{v}_j$ for all pixels j.

After capturing images of a displaced target, we use the procedure outlined in Sec. 4.2 to estimate local phasors $\vec{u}_k$. This requires approximately $0.15\,\mathrm{s}$ in MATLAB on a M1 MacBook Pro to process a 1280×1024 image, and approximately $1\,\mathrm{s}$ for our Python implementation. A parallelized GPU implementation further reduces the runtime to under 5 ms per 1280×1024 image.

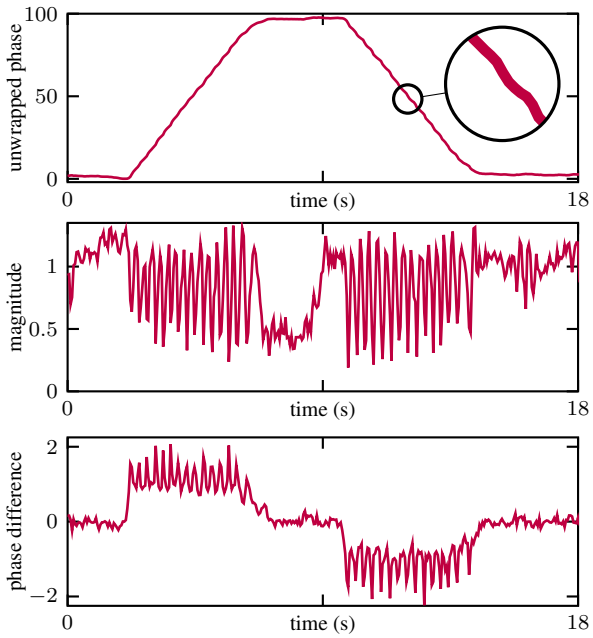
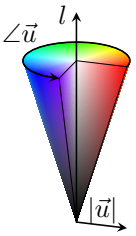


Figure 8. **Recovered phasor of a target moving laterally on a translation-stage motion.** We programmed the stage to move $20\,\mu\mathrm{m}$ in one direction, pause, and move $20\,\mu\mathrm{m}$ in the opposite direction. **(Top)** Unwrapped phase recovered throughout the motion. **(Middle)** Magnitude of the same recovered phasor over time. While the object moves in either direction, we observe that the velocity of our stage is not constant, resulting in fluctuations in the phasor magnitude. **(Bottom)** Finite differencing of unwrapped phase.

Phasor visualization. For Fig. 12 and Fig. 13, we visualize the recovered phasors by using the HSV color space. For these results, hue represents the wrapped phase $\angle\vec{u}$, saturation represents the magnitude $|\vec{u}|$, and lightness represents the intensity of the raw image. For example, for displacements that exceed a phase change of 2π, we observe repeating hues. Also, as the magnitude approaches zero (*e.g.*, pixels experience large motion), the colors desaturate and the phase becomes difficult to discern.



6. Results

6.1. Motorized target

We quantitatively evaluate our method's performance in recovering micrometer scale displacement and motion of a motorized target. The target in this case is a tuning fork painted with diffuse white paint.

In Fig. 8, we recovered the round trip displacement of the target with our camera operating at 20 fps. The mag-

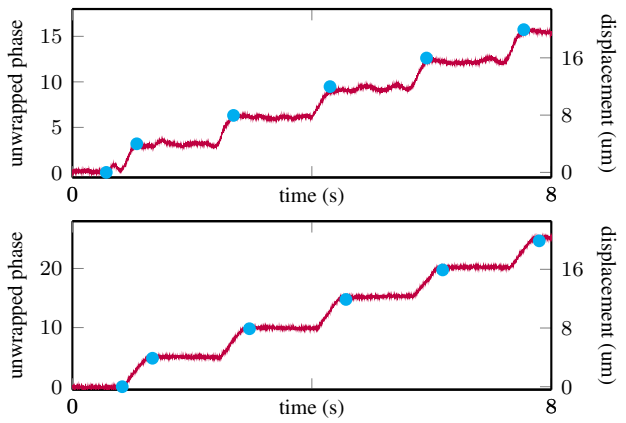


Figure 9. **Stepped in-plane motion.** We measure displacement of a target moving by $4\,\mu\text{m}$ steps at 600 fps. The left y-axis shows the unwrapped phase in radians, and the right y-axis shows the displacement converted into physical units. Ground truth displacements (in cyan) were obtained by querying stage position after each step. **(Top)** The two sources are separated by approximately $5.2\,\text{cm}$, which we use to calculate the mapping from phase to metric displacement. **(Bottom)** The two sources are separated by approximately $9.2\,\text{cm}$, which reduces the influence of sensor noise by increasing the system's sensitivity.

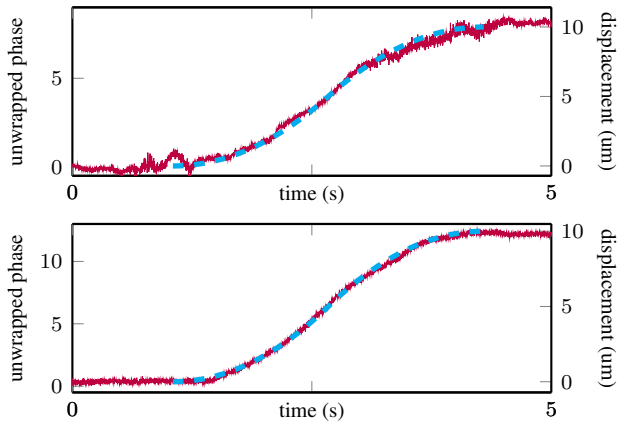


Figure 10. **Continuous in-plane motion.** We generate continuous motion by specifying a max velocity ($10\,\mu\text{m}\,\text{s}^{-1}$), and acceleration ($3.8\,\mu\text{m}\,\text{s}^{-2}$) to move a fixed distance of $10\,\mu\text{m}$. We derive the expected motion (in cyan) from the trapezoidal velocity profile. **(Top)** $5.2\,\text{cm}$ baseline. **(Bottom)** $9.2\,\text{cm}$ baseline.

nitudes of the recovered phasors vary as the stage moves. These observed fluctuations in magnitude reflects the velocity ripples of the translation stage during its motion [23]; we verified this by performing finite differencing on phase-only measurements in the supplement. Phasor magnitude also appears to reduce over a large distance, which we suspect occurs due to the speckle pattern slowly decorrelating.

We also experimented with different laser baselines to control sensitivity in both Fig. 9 and Fig. 10. Given the baseline and the depth of the target, we calculate the expected metric displacement associated with a given phase change. For smaller baselines, the phase changes more slowly with respect to displacement. This reduces phase unwrapping ambiguities, but makes the measurements more susceptible to sensor noise. In contrast, larger baselines increase accuracy, but worsen phase wrapping issues.

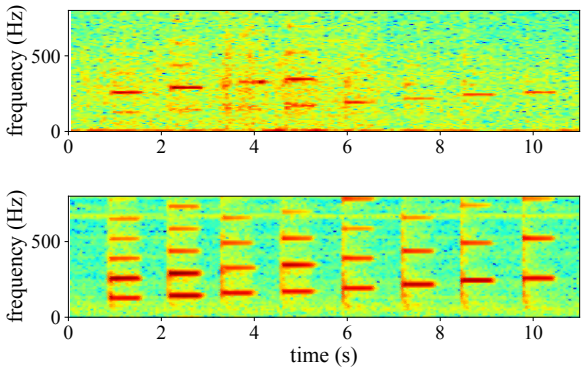


Figure 11. **Spectrogram of a chip bag.** The bag vibrates in response to a speaker playing piano notes from C3 to C4. **(Top)** The spectrogram recovered from in-plane vibrations captures higher-order harmonics, whose frequencies are an integer multiple of the fundamental frequency. **(Bottom)** Microphone Spectrogram.

6.2. Acoustic recovery

We reproduced the visual microphone experiment from Davis *et al*. [3], placing a chip bag in the scene while audio plays on a speaker positioned on its left side, inducing in-plane vibrations to the chip bag surface. We place a small piece of retroreflective tape on the bag to boost the reflected signal, and capture the subtle surface vibration at 1801 fps.

Fig. 11 shows the reconstructed spectrogram of a speaker playing piano notes from C3 to C4. The recovered frequency structure closely matches recording from the microphone, including presence of noticeable second harmonics, demonstrating our method's ability in recovering fine, high-frequency temporal details from subtle surface vibrations induced by sound waves—albeit not to the level of methods using surface tilt to recover audio [21]. In the supplement, we perform shearography to capture surface tilt with our method, and demonstrate the ability to capture higher-quality audio signals as a result.

6.3. Shape deformation

We demonstrated that our method can capture subtle state changes of an object caused by structural deformation. In Fig. 13, a Q-tip touches the upper-right prong of a tuning fork, producing a downward force on the prong. The lasers

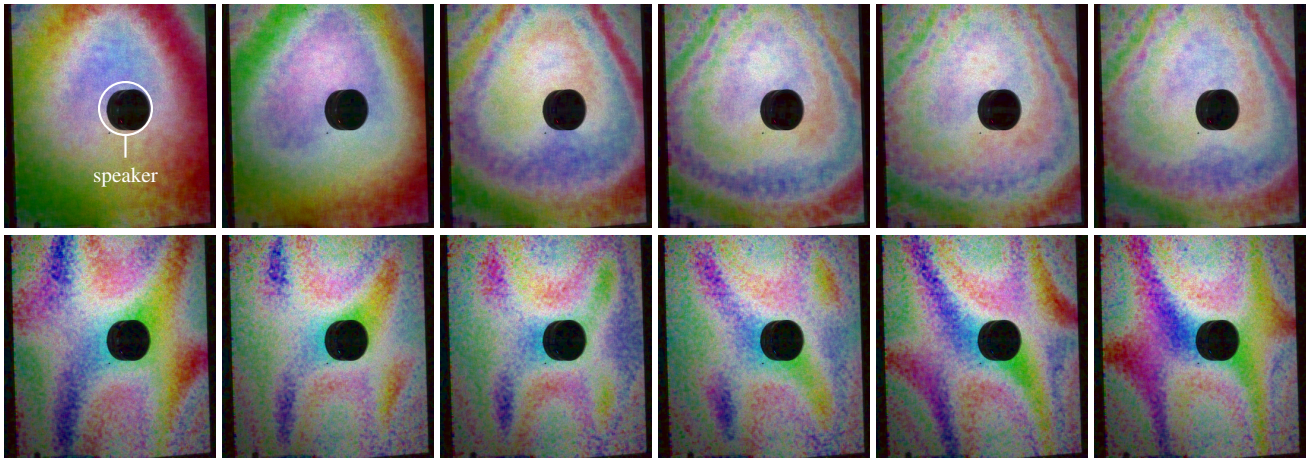


Figure 12. **Chladni patterns.** Vibrational modes of a plate from a contact speaker vibrating at 134 Hz (**Row 1**) and 150 Hz (**Row 2**).

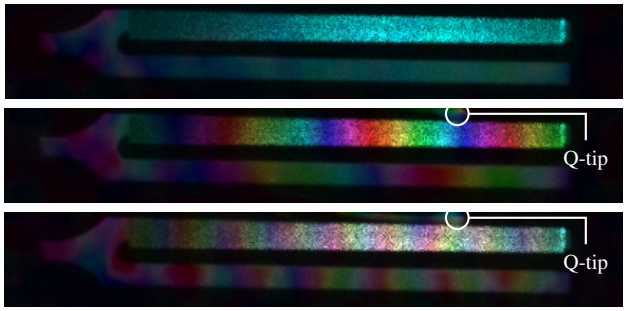


Figure 13. **Deformations created by a Q-tip.** We visualize submicron displacement and motion that occurs to a tuning fork when using a Q-tip to apply **(Top)** no pressure, **(Middle)** light pressure, and **(Bottom)** medium pressure. The top prong displaces more than the bottom prong as evidenced by the number of times the color hues wrap. For the bottom result, the colors desaturate, indicating the presence of motion during the camera's exposure.

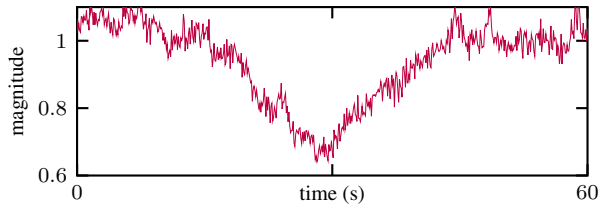


Figure 14. **Recovered phasor magnitude.** A tuning fork is excited by a speaker emitting an audio frequency that increases and then decreases linearly in volume.

are configured in a vertical arrangement such that the system responds to vertical displacement.

6.4. Vibrational modes

A tuning fork with a fundamental of 288 Hz is lightly struck, producing vibrational motion. We record measurements at 600 fps, and place retroreflective tape on the fork to boost the signal in response. We observe that the vibration occurs at approximately 286 Hz after adhering the tape. Fig. 1 provides a visualization of displacement in the vertical direction, after unwrapping the phase. Note how the direction of motion changes as a function of time, and both prongs move in opposite directions. This is consistent with vibrations that occur at the fundamental frequency [18].

We further investigated how vibration amplitude changes with excitation volume in Fig. 14. We place a speaker beneath the tuning fork and play a sound wave that matches the fork's fundamental frequency. Changing the volume up and down induces a change in the phasor magnitudes.

We also demonstrate the ability to capture more complex vibrational modes. In Fig. 12, we mount a bone-conduction speaker to the center of a rigid metal plate, and tilt the plate at an angle relative to the camera's optical axis. When resonated, the plate deforms into Chladni patterns that depend on both the shape of the plate and the frequency [24].

7. Conclusion

We present a new computational framework for estimating displacement and motion from speckle interferograms. We propose a factorization calibration procedure and optical-flow-inspired reconstruction procedure for computing per-pixel displacement and motion information from a single image. We believe our work opens up vibrational analysis to a wide range of applications, including those demanding the full-field capture of transient vibrations.

Acknowledgments. This work was supported by a NSF CAREER award (IIS 2238485) and a NSF-BSF award (IIS 2513219).

References

[1] JN Butters and JA Leendertz. A double exposure technique for speckle pattern interferometry. *Journal of Physics E: Scientific Instruments*, 4(4):277, 1971. 3

[2] Mingxuan Cai, Dekel Galor, Amit Pal Singh Kohli, Jacob L Yates, and Laura Waller. Event2audio: Event-based optical vibration sensing. In *2025 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2025. 1

[3] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J Mysore, Fredo Durand, and William T Freeman. The visual microphone: Passive recovery of sound from video. 2014. 1, 7

[4] Xavier Colonna de Lega and Pierre Jacquot. Deformation measurement with object-induced dynamic phase shifting. *Applied Optics*, 35(25):5115–5121, 1996. 4

[5] Marco A Escobar, Julio C Estrada, and Javier Vargas. Phase-shifting vu factorization for interferometry. *Optics and Lasers in Engineering*, 124:105797, 2020. 5

[6] Julio C Estrada, Victor H Flores, and Javier Vargas. A two steps phase-shifting demodulation method using the vu factorization. *Optics and Lasers in Engineering*, 147:106730, 2021. 5

[7] Berthy T Feng, Alexander C Ogren, Chiara Daraio, and Katherine L Bouman. Visual vibration tomography: Estimating interior material properties from monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16231–16240, 2022. 1

[8] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981. 5

[9] Wenbo Jiang, Tong Ren, and Qianhua Fu. Deep learning in the phase extraction of electronic speckle pattern interferometry. *Electronics*, 13(2):418, 2024. 2

[10] Chih-Cheng Kao, Gym-Bin Yeh, Shu-Sheng Lee, Chih-Kung Lee, Ching-Sang Yang, and Kuang-Chong Wu. Phase-shifting algorithms for electronic speckle pattern interferometry. *Applied optics*, 41(1):46–54, 2002. 2, 3, 4

[11] Matan Kichler, Shai Bagon, and Mark Sheinin. Learning to see inside opaque liquid containers using speckle vibrometry. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9466–9476, 2025. 1

[12] Violeta Dimitrova Madjarova, Hirofumi Kadono, and Satoru Toyooka. Dynamic electronic speckle pattern interferometry (despi) phase analyses with temporal hilbert transform. *Opt. Express*, 11(6):617–623, 2003. 4

[13] Thomas R. Moore and Jacob J. Skubal. Time-averaged electronic speckle pattern interferometry in the presence of ambient motion. part i. theory and experiments. *Applied Optics*, 47(25):4640–4648, 2008. 2

[14] Alexander C Ogren, Berthy T Feng, Jihoon Ahn, Katherine L Bouman, and Chiara Daraio. Visual surface wave elastography: Revealing subsurface physical properties via visible surface waves. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 26446–26455, 2025. 1

[15] Siyou Pei, Pradyumna Chari, Xue Wang, Xiaoying Yang, Achuta Kadambi, and Yang Zhang. Forcesight: Non-contact force sensing with laser speckle imaging. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, pages 1–11, 2022. 1

[16] Pascal Picart, Julien Leval, Denis Mounier, and Samuel Gougeon. Time-averaged digital holography. *Opt. Lett.*, 28 (20):1900–1902, 2003. 2

[17] Mark Rau, Julius O Smith, and Doug L James. Augmenting a single-point laser doppler vibrometer to perform scanning measurements. *The Journal of the Acoustical Society of America*, 151(4):A157–A157, 2022. 1

[18] Daniel A. Russell. Vibrational modes of a tuning fork. https://www.acs.psu.edu/drussell/Demos/TuningFork/fork-modes.html, 2025. Acoustics and Vibration Animations, The Pennsylvania State University. 8

[19] Ryusuke Sagawa, Yusuke Higuchi, Hiroshi Kawasaki, Ryo Furukawa, and Takahiro Ito. Dense pixel-wise micro-motion estimation of object surface by using low dimensional embedding of laser speckle pattern. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2020. 1

[20] Ryusuke Sagawa, Yusuke Higuchi, Ryo Furukawa, and Hiroshi Kawasaki. Acquisition and visualization of micro-vibration of a sound wave in 3d space. *Journal of Robotics and Mechatronics*, 34(5):1024–1032, 2022. 1

[21] Mark Sheinin, Dorian Chan, Matthew O'Toole, and Srinivasa G Narasimhan. Dual-shutter optical vibration sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16324–16333, 2022. 1, 7

[22] Wei Sun, Tuochao Chen, Jiayi Zheng, Zhenyu Lei, Lucy Wang, Benjamin Steeper, Peng He, Matthew Dressa, Feng Tian, and Cheng Zhang. Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser doppler vibrometry. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(3):1–28, 2020. 1

[23] Thorlabs. Z825b manual. https://www.thorlabs.com/_sd.cfm?fileName=17630-D01.pdf&partNumber=Z825B, 2023. Accessed: 2025-11-13. 7

[24] PH Tuan, CP Wen, PY Chiang, YT Yu, Hsing-Chin Liang, Kai-Feng Huang, and Yung-Fu Chen. Exploring the resonant vibration of thin plates: Reconstruction of chladni patterns and determination of resonant wave numbers. *The Journal of the Acoustical Society of America*, 137(4):2113–2123, 2015. 8

[25] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (ToG)*, 32(4):1–10, 2013. 1

[26] Wei-Chung Wang, Chi-Hung Hwang, and Shu-Yu Lin. Vibration measurement by the time-averaged electronic speckle pattern interferometry methods. *Applied optics*, 35 (22):4502–4509, 1996. 2, 5

[27] Wikipedia contributors. Laser doppler vibrometer — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Laser_Doppler_vibrometer, 2025. 1

[28] Lianxiang Yang, Xin Xie, Lianqing Zhu, Sijin Wu, and Yonghong Wang. Review of electronic speckle pattern interferometry (espi) for three dimensional displacement measurement. *Chinese Journal of Mechanical Engineering*, 27 (1):1–13, 2014. 1, 2, 3

[29] Alan Yuille and Daniel Snow. Shape and albedo from multiple images using integrability. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 158–164. IEEE, 1997. 2, 5

[30] Zeev Zalevsky, Yevgeny Beiderman, Israel Margalit, Shimshon Gingold, Mina Teicher, Vicente Mico, and Javier Garcia. Simultaneous remote extraction of multiple speech sources and heart beats from secondary speckles pattern. *Optics express*, 17(24):21566–21580, 2009. 1

[31] Tianyuan Zhang, Mark Sheinin, Dorian Chan, Mark Rau, Matthew O'Toole, and Srinivasa G Narasimhan. Analyzing physical impacts using transient surface wave imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4339–4348, 2023. 1

[32] Yang Zhang, Gierad Laput, and Chris Harrison. Vibrosight: Long-range vibrometry for smart environment sensing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pages 225–236, 2018. 1

[33] Xinyu Zhou, Peiqi Duan, Yeliduosi Xiaokaiti, Chao Xu, and Boxin Shi. Event-based visual vibrometry. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 24666–24676, 2025. 1