Analyzing Physical Impacts using Transient Surface Wave Imaging: Supplementary Materials

Tianyuan Zhang[1]   Mark Sheinin[1]   Dorian Chan[1]   Mark Rau[2]
Matthew O'Toole[1]   Srinivasa Narasimhan[1]
[1]Carnegie Mellon University   [2]Stanford University

1. Camera Calibration

For a well-aligned optical system, the relation between the image-plane shifts (d_x, d_y) and the surface tilts (θ_x, θ_y) is described in Eq. (1) of the main paper. However, there might be misalignment between the object plane axis and camera frame axis. And this misalignment is hard to avoid when we are using anisotropic material with unknown anisotropy axis. (see Fig. 2). Therefore, we compensate for this misalignment by adding an appropriate rotation matrix to the measured shifts in Eq. (1)

$$(\theta_x, \theta_y) = (h_x, h_y) \odot \mathcal{R}(\alpha_z)(d_x, d_y), \qquad (17)$$

where

$$\mathcal{R}(\alpha_z) \equiv \begin{bmatrix} \cos \alpha_z & -\sin \alpha_z \\ \sin \alpha_z & \cos \alpha_z \end{bmatrix}, \qquad (18)$$

and α_z denotes the misalignment angle shown in Fig. 2. Note that because our prototype views the speckle through a cylindrical lens, the conversion factor (h_x, h_y) for isotropic materials is not equal with $h_y > h_x$ [3].

As described by Sheinin *et al.*, the conversion factors (h_x, h_y) can be manually calibrated by measuring known tilts using a goniometer [3]. [1] We found this approach to be time-consuming and thus, for most experiments, we developed a more efficient approach described in the next section.

Calibrating the tilts ratio using an isotropic material. Let us rewrite Eq. (17) in terms of the optical scaling ratio s

$$(\theta_x, \theta_y) = h_x(1, s) \odot \mathcal{R}(\alpha_z)(d_x, d_y), \qquad (19)$$

where

$$s \equiv h_y/h_x. \qquad (20)$$

---

[1]For example, in the whiteboard experiment, calibration using a goniometer yielded h_x to 514 pixels/degree and h_y to 2809 pixels/degree.

Figure 1. Materials used in our experiments.

Using Eq. (19) we can rewrite Eqs. (16) of the main paper as

$$\begin{aligned}
\hat{\boldsymbol{h}}^*(\boldsymbol{x}_n, t) &\approx \frac{\boldsymbol{\theta}_n(t) \odot (1, m^2)}{\|\boldsymbol{\theta}_n(t) \odot (1, m^2)\|} \\
&= \frac{h_x(1, sm^2) \odot \mathcal{R}(\alpha_z)(d_{n,x}(t), d_{n,y}(t))}{\|h_x(1, sm^2) \odot \mathcal{R}(\alpha_z)(d_{n,x}(t), d_{n,y}(t))\|} \\
&= \frac{(1, sm^2) \odot \mathcal{R}(\alpha_z)(d_{n,x}(t), d_{n,y}(t))}{\|(1, sm^2) \odot \mathcal{R}(\alpha_z)(d_{n,x}(t), d_{n,y}(t))\|}.
\end{aligned} \qquad (21)$$

Eq. (21) shows that to compute $\hat{\boldsymbol{h}}^*$ (and similarly $\hat{\boldsymbol{g}}^*$ by setting $m = 1$), we only need to calibrate the relative scaling s and the rotation α_z.

We estimate s and α_z by measuring the vibrations produced by knocking at known locations $\boldsymbol{x}_s$ on an isotropic surface ($m = 1$). Because isotropic materials have circular wavefronts, the expected tilt ratios at each measured point $\boldsymbol{x}_n$ are known and can be used to recover s. For example, if $\alpha_z = 0$, a point $\boldsymbol{x}_n$ located at an angle of $45°$ from $\boldsymbol{x}_s$ should experience the same tilt in both axes $\theta_y/\theta_x = 1$, thus

(a) speckle photography **(b)** camera misalignment

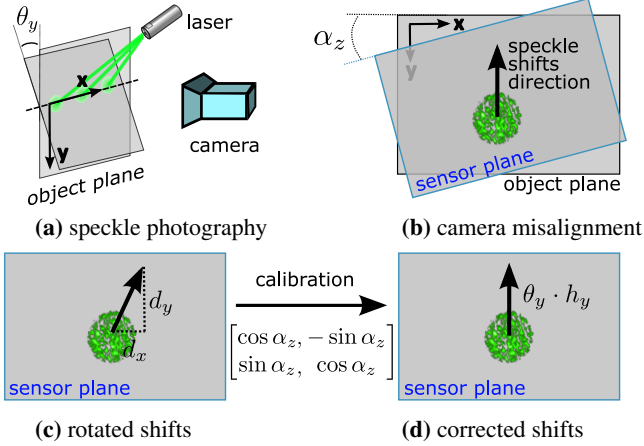**(c)** rotated shifts **(d)** corrected shifts

Figure 2. Speckle shifts in sensor plane depends linearly on the surface tilts. In camera calibration, an misalignment due to camera angles results in rotated shifts in the sensor plane, as described in Eq. 17.

from Eq. (17), $s = d_x/d_y$.

Let r_n denote the image-domain shifts ratio at point $\boldsymbol{x}_n$:

$$r_n \equiv d_{n,y}/d_{n,x} \qquad (22)$$

Then, for each measurement point $\boldsymbol{x}_n$, we compute the image-domain ratio median during the stable time interval:

$$r_n^{\mathrm{med}} = \mathrm{median}\left(\{r_n(t), t \in T^{\mathrm{c}}(\boldsymbol{x}_n)\}\right). \qquad (23)$$

As mentioned, in isotropic surfaces, the gradient direction is known and should coincide with the source direction during the stable time interval

$$\arctan\left(r_n^{\mathrm{med}}\right) = \arctan\left(\frac{y_s - y_n}{x_s - x_n}/s\right) - \alpha_z. \qquad (24)$$

We then solve a least squares optimization problem using the uncalibrated gradients from each measurement point to estimate the relative scaling: s^*, and camera roll angle α_z^*:

$$s^*, \alpha_z^* = \underset{k,\alpha}{\arg\min} \sum_n \left[\arctan(r_n^{\mathrm{med}}) - \arctan(\frac{y_s - y_n}{x_s - x_n}/s) + \alpha\right]^2. \qquad (25)$$

To get a robust calibration, we typically repeated three knocks at five known positions $\boldsymbol{x}_s$. In most experiments, the calibrated α_z was less than 5 degree and s was between 3 and 6.

2. Localization algorithm

Algorithm 1 provides pseudocode for the localization algorithm described in Section 4 of the main paper. There are two hyper-parameters for finding the stable time interval $T^{\mathrm{c}}(\boldsymbol{x}_n)$: Δt and τ. Parameter Δt is the length of the

stable time interval, while τ adjust the threshold for detecting the arrival time of the first wavefront.

There are two additional hyper-parameters for the backprojection step: β which controls the angular size of the 2D cone for backprojection and σ which controls the strength of the distance-dependent weighting. We typically set Δt= 1.5ms, τ=10, β=6°, σ=5.

We experimented with filtered backprojection, namely, applying a filtering function $f(\cdot)$ in line 16 in Algorithm 1. Specifically, we set $f(C(\boldsymbol{x},t')) \leftarrow \exp(C(\boldsymbol{x},t'))$. We found that this filtering had negligible effect in most experiments and was only beneficial for signals having high SNR. Filtered backprojection was only used in the ping-pong experiment. If only the relative scaling term s and misalignment angle α_z are calibrated, line 11,12 can be replaced with Eq.(21) to compute the normalized source direction.

In line 2, we compute σ_n^2 using the signal at static frames (*i.e.* with no vibrations). Additionally, we use different hyperparameters β, σ for signals with different SNRs. For measurements with high SNR, we select smaller β and larger σ values, which produces a sharper voting map. For low-SNR signals, we use a larger β and smaller σ to regularize the reconstructions.

For certain materials, we observed multiple stable time intervals. For example, two stable time intervals occurred in birch, plywood, and porcelain. The first stable interval starts directly after the first wave arrives and lasts for a very short duration. Then, a second interval begins ~0.5ms later and lasts for a longer time period. We found that the second, longer, stable time interval provides better localization accuracy and hence used it for to yield the localization results in these materials.

3. Impact source localization experiments

In this section we provide additional details about the source localization experiments.

3.1. Localization results on additional materials

In Fig. 8 of the main paper, we showed localization results on two materials: an isotropic whiteboard and an anisotropic plywood board. Here, we provide more localization results on eight other boards: medium-density fiberboard, particle board, glass, gypsum panel, two types of plywood, porcelain and PVC panel. We show the recovery in Fig. 3. The boards on the first row are made of isotropic materials. From right to left, the average localization error is 1.88 cm, 1.15 cm, 1.61 cm, and 1.86cm, respectively. The second row shows results for all anisotropic materials, with average errors of 1.57 cm, 1.39 cm, 2.97 cm, and 2.09 cm, from left to right.
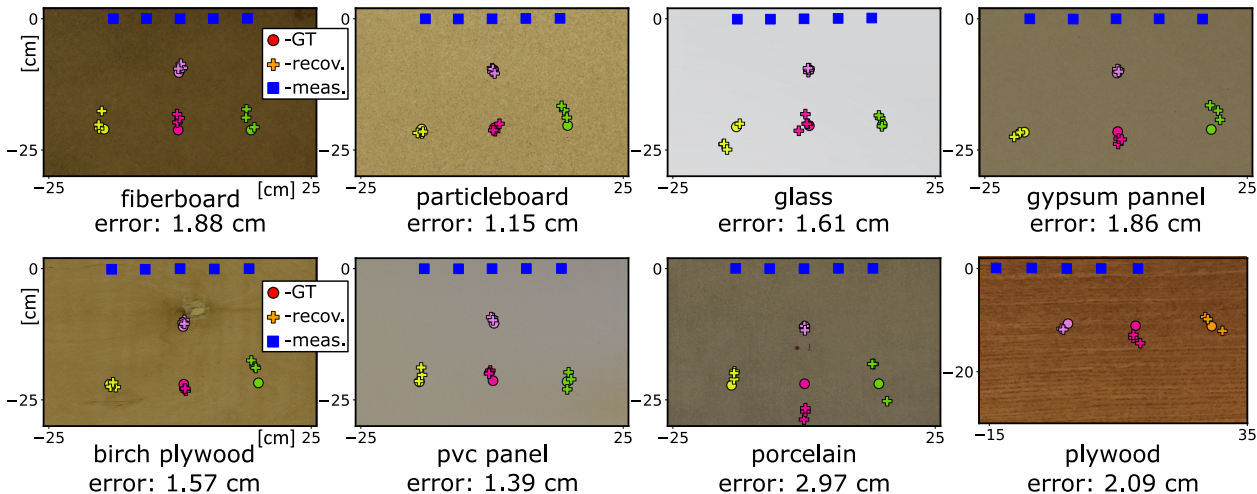
Figure 3. Extra localization results on eight boards. For each board, we knocked at three to four places. In each place, we repeated 3-4 times. The top row shows results for isotropic materials. From right to left, the average localization error is 1.88 cm, 1.15 cm, 1.61 cm, and 1.86 cm, respectively. The bottom row shows results for anisotropic materials, with average errors as 1.57 cm, 1.39 cm, 2.97 cm, and 2.09 cm, from left to right.

3.2. Synthesizing non-collinear measurements configurations

In Figs. 10 and 13 of the main paper, we showed synthesized measurement point configuration that our current camera prototype does not support. Here, detail on how we made these measurements. Our current prototype can only project a set of collinear laser points. To simulate simultaneous measurements using the configurations Figs. 10 and 13, we repeating the impact two times while moving the camera. Specifically, we set up the imaging system to image the first set of five points, created an impact (knock on the board or step on the floor), and recorded the signals. Then, we moved our imaging system to capture the second points set, re-calibrated the camera, and captured the signals while repeating the impact at the same location.

3.3. Additional performance analysis

Localization error vs. point distance Triangulation localization error increases quadratically with target distance [1]. Our backprojection algorithm, which performs triangulation via backprojection, suffers from the same limitation. We test this empirically by recovering a set of points located 20cm, 30cm, 50cm, 70cm, and 100cm away from our (central) marker. We repeated 18, 18, 25, 10, and 10 measurements at each distance, respectively. We show the relation between average recovery error and target distance in Fig. 4(a). The orange dotted line denotes a quadratic fit.

Localization error vs. point angles With a collinear measurement setup, triangulation localization error in-

(a) error vs. distance (b) error vs. angle
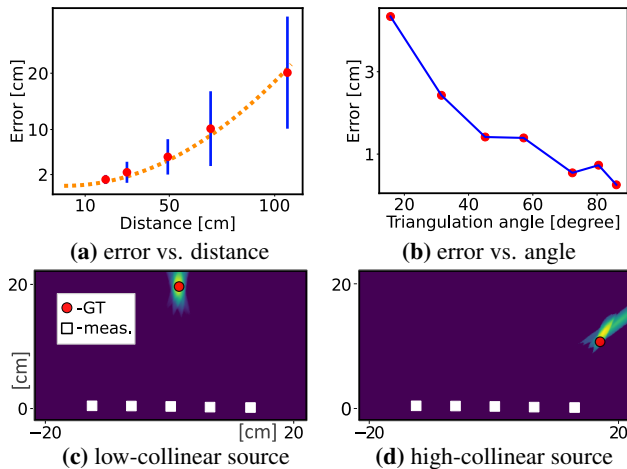

(c) low-collinear source (d) high-collinear source

Figure 4. Analysis on localization error. **(a)** Error vs. distance. The red points denote the average error of multiple recoveries at the same distance. The vertical blue bars denote the standard deviation of the repeated measurements per distance. The orange dotted curve is a fitted quadratic function. **(b)** Error vs. angle between the source point and the line defined by the measurement points. **(c)-(d)** Backprojection voting maps for low-collinear and higher-collinear sources, respectively. The localization accuracy degrades as the source-point angle becomes collinear with the measurement points' line.

creases dramatically when target points lie near collinear angles. We mentioned this limitation in Fig. 8 of the main paper, and here we show a more detailed study. We plot the relation between localization error and the angle between the source point and the line defined by the measurement

Algorithm 1: Source localization with backprojection.

Input: Measured tilts $\{\boldsymbol{\theta}_n(t)\}$;
Measurement point locations $\{\boldsymbol{x}_n\}^N$;
Material calibration term m;
Hyperparameters: $\beta, \sigma, \tau, \Delta t$;
Filtering function $f(\cdot)$

Output: Voting map $C(\boldsymbol{x})$, and prediction of impact location $\boldsymbol{x}^*$

```
/* step-1, compute stable time intervals */
```
1 **for** $n = 1 \to N$ **do**
```
     /* compute variance over time       */
```
2 $\quad \sigma_n^2 \leftarrow \text{Var}\left(\{\|\boldsymbol{\theta}_n(t)\|_2\}_t\right)$
```
     /* get arrival time of vibration wave */
```
3 $\quad t_n^{\text{start}} \leftarrow \min_t\{t, \|\boldsymbol{\theta}_n(t)\|_2 > \tau \cdot \sigma_n\}$
```
     /* Δt as hyperparameter             */
```
4 $\quad T(\boldsymbol{x}_n) \leftarrow [t_n^{\text{start}}, t_n^{\text{start}} + \Delta t]$
5 **end**
```
   /* step-2, run backprojection         */
```
6 $C(\boldsymbol{x}) \leftarrow \mathbf{0}$ `// initialization`
7 **for** $t' = 0 \to \Delta t$ **do**
8 $\quad C(\boldsymbol{x}, t') \leftarrow \mathbf{0}$ `// initialization`
9 $\quad$ **for** $n = 1 \to N$ **do**
10 $\quad\quad t \leftarrow t_n^{\text{start}} + t'$
```
         /* compute source direction      */
```
11 $\quad\quad \hat{\boldsymbol{h}} \leftarrow \boldsymbol{\theta}_n(t) \odot (1, m^2)$
12 $\quad\quad \hat{\boldsymbol{h}} \leftarrow \frac{\hat{\boldsymbol{h}}}{\|\hat{\boldsymbol{h}}\|_2}$
```
         /* cast a weighted cone, Eq.(9)  */
```
13 $\quad\quad C_n(\boldsymbol{x}, t') \leftarrow \text{ConeBP}(\boldsymbol{x}_n, \hat{\boldsymbol{h}}, \beta, \sigma)$
```
         /* sum up over all measured points */
```
14 $\quad\quad C(\boldsymbol{x}, t') \leftarrow C(\boldsymbol{x}, t') + C_n(\boldsymbol{x}, t)$
15 $\quad$ **end**
```
     /* accumulate over time             */
```
16 $\quad C(\boldsymbol{x}) \leftarrow C(\boldsymbol{x}) + f\left(C(\boldsymbol{x}, t')\right)$
17 **end**
18 $\boldsymbol{x}^* = \arg\max_{\boldsymbol{x}} C(\boldsymbol{x})$

(a) measured vibrations (both impacts)

(b) first impact recovery   **(c)** second impact recovery
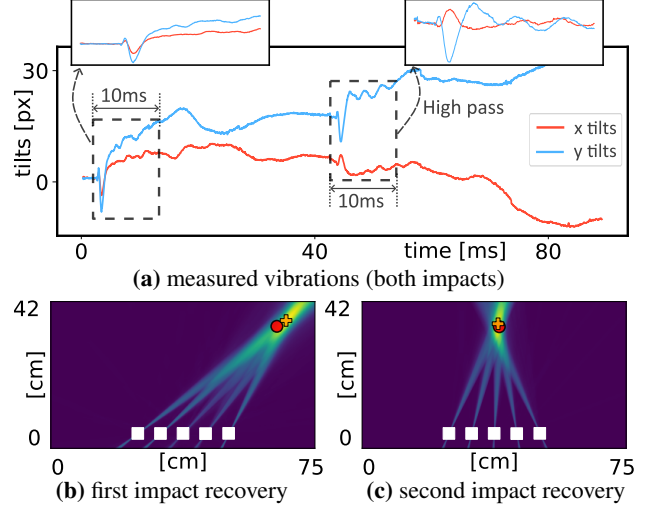
Figure 5. Recovering two temporally adjacent impacts. In this experiment we attempt to manually create two simultaneous knocks at two known surface points. **(a)** shows measured tilts for two knocks within 50 ms. High-frequency information in the measured tilts decays rapidly after waves from each knock arrives. Thus, high-pass filtering can remove most of the energy left from the vibration fields of the first knock when the second knock happens. **Bottom row** shows recovery for the first and second knock.

points in Fig. 4(b). We show the backprojection heatmap of two examples with different uncertainty levels.

Recovering temporally adjacent impacts The localization method described in Sections 3-5 of the main paper assumes a single impact source. However, we attempted to test impact recovery for multiple simultaneous impacts. Despite our attempts, we found it very difficult to manually simultaneously knock at two surface locations at once.

Fig. 5(a) shows the measured signals from one such attempt to knock on a whiteboard at two different (known) points simultaneously. While the knocks appeared simultaneous to the knocking individual, an examination of the measured signals revealed a 40-50 ms time difference be-

tween the two knocks. Fig. 5(b)-(c) show that this short imperceptible time interval between the two impacts is enough to "unmix" the vibration signals between the two knocks and accurately recover both locations.

Specifically, the transient vibrations of the first impact last about 5 ms during which we localize the first impact before the second impact start. Then, when the vibrations from the second impact affect affect the measured points, the high-frequency vibrations from the first impact had mostly subsided, allowing its recovery.

4. Measuring material anisotropy

We can measure material anisotropy by fitting its level-set shape with an ellipse with anisotropy factor m, as described in Eq. (14) in the main paper. To derive m, we need to calibrate the optical scaling term s first, since these two scaling terms are coupled as shown in Eq. 21. After calibrating the optical scaling term s, using either a goniometer or an isotropic material, we can use a similar procedure shown in Eq. 25 to compute the material anisotropy and misalignment angle jointly.

We showed results for three materials in the main paper, in Fig. (11). We provide the calibrated results for another five materials in Fig. 6. The top row shows results for three isotropic materials, while the bottom row shows three anisotropic materials.
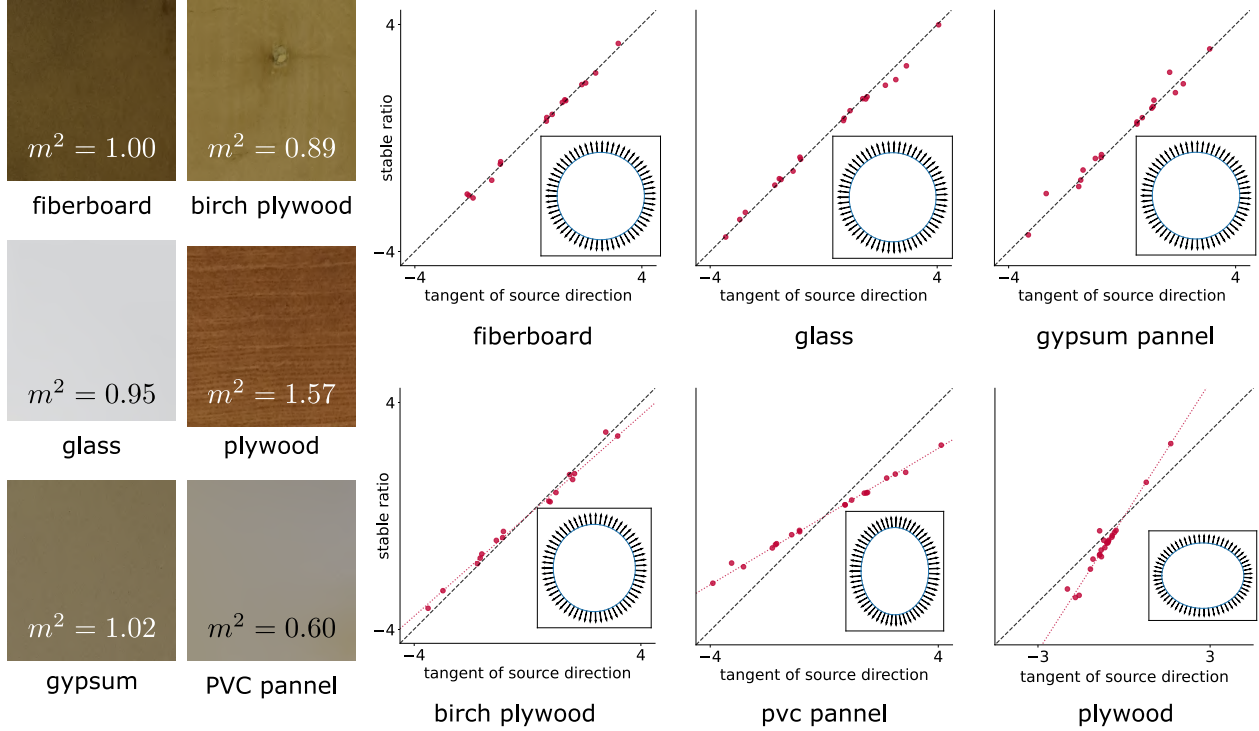
Figure 6. Transient vibration analysis for different materials. **Top plots row:** Results for isotropic materials. The surface gradient direction θ_y/θ_x were measured by knocking at known surface positions. **Bottom plots row:** Results for anisotropic materials. The ellipse corresponds to the approximated level-set shapes.

5. Futher analysis of impact force experiment

Fig.12(a) of the main paper shows that the peak tilts magnitude $\|\boldsymbol{\theta}(t)\|_2^{\mathrm{peak}}$ has a square root relation to the ping-pong ball's dropping height

$$\|\boldsymbol{\theta}(t)\|_2^{\mathrm{peak}} \propto \sqrt{h}, \tag{26}$$

where h is the dropping height. Neglecting air resistance, conservation of energy dictates that

$$mgh = 1/2mV^2, \tag{27}$$

where m is the ball's mass and g is gravitational acceleration. Therefore, the relation between V and h is also

$$V \propto \sqrt{h}. \tag{28}$$

The peak impacting force is linearly proportional to the impact velocity of a ping-pong ball [2]. Thus, Eqs. 29 show that the peak magnitude of the measured tilts is linear to the ball's peak impacting force

$$\|\boldsymbol{\theta}(t)\|_2^{\mathrm{peak}} \propto f^{\mathrm{peak}}, \tag{29}$$

suggesting that the transient vibrations can be used to extract the impacting object's force.

References

[1] Subhodev Das and Narendra Ahuja. Performance analysis of stereo, vergence, and focus as depth cues for active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(12):1213–1219, 1995. 3

[2] Mont Hubbard and WJ Stronge. Bounce of hollow balls on flat surfaces. *Sports Engineering*, 4(2):49–61, 2001. 5

[3] 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