Visual Vibration Tomography: Estimating Interior Material Properties from Monocular Video

BERTHY FENG, California Institute of Technology, USA

ALEXANDER C. OGREN, California Institute of Technology, USA

CHIARA DARAIO, California Institute of Technology, USA

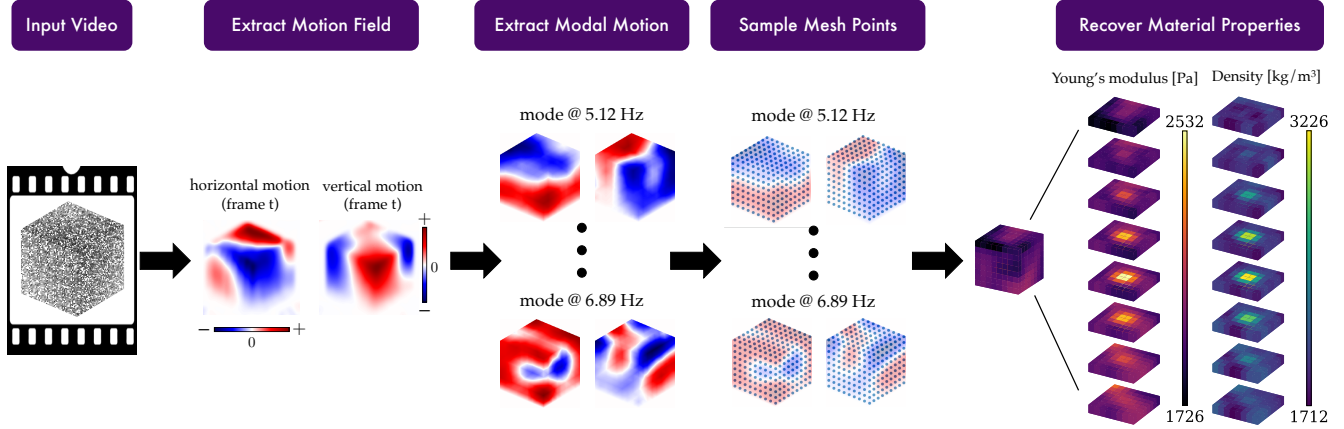KATHERINE L. BOUMAN, California Institute of Technology, USA

Fig. 1. Method overview. We start with an input video showing vibration of the object of interest. We extract a motion field for each frame and then decompose this motion into *image-space modes*. An image-space mode is a horizontal and vertical displacement field that occurs periodically at a certain frequency throughout the video. We model the object as a mesh and sample the visible mesh vertices in the image-space modes. From the image-space modes sampled at these points, we are able to recover a voxelized volume of the Young's modulus and density throughout the object.

An object's interior material properties, while invisible to the human eye, determine motion observed on its surface. We propose an approach that estimates heterogeneous material properties of an object directly from a monocular video of its surface vibrations. Specifically, we estimate Young's modulus and density throughout a 3D object with known geometry. Knowledge of how these values change across the object is useful for characterizing defects and simulating how the object will interact with different environments. Traditional non-destructive testing approaches, which generally estimate homogenized material properties or the presence of defects, are expensive and use specialized instruments. We propose an approach that leverages monocular video to (1) measure an object's sub-pixel motion and decompose this motion into image-space modes, and (2) directly infer spatially-varying Young's modulus and density values from the observed image-space modes. On both simulated and real videos, we demonstrate that our approach is able to image material properties simply by analyzing surface motion. In particular, our method allows us to identify unseen defects on a 2D drum head from real, high-speed video.

1 INTRODUCTION

The subtle motions of objects around us are clues to their physical properties. Among such properties are stiffness and density,

which dictate how an object will interact with its environment. As humans, we can vaguely characterize how stiff or heavy a material is, such as when we infer that a rubber basketball will bounce higher than a ceramic bowling ball by tapping on its surface. Most engineering applications, however, require a greater level of detail, such as when an aeronautical engineer must faithfully simulate how an airplane wing will react to wind turbulence. In graphics applications, a full characterization of an object's material properties allows one to faithfully simulate its behavior. These scenarios require non-destructive testing to obtain physical properties of the object without altering it.

We propose *visual vibration tomography*, an approach that estimates the material properties of an object directly from vibration signals extracted from monocular video. Much of non-destructive testing (NDT) and structural health monitoring (SHM) has focused on measuring vibrations to identify the presence and location of defects in structures with a known geometry. However, the tools developed for NDT/SHM are not generally used to determine the precise spatial distribution of different physical properties in objects with a heterogeneous interior structure.

Here, we show that we can measure structural vibrations as subpixel motion in 2D video, and then use this motion to constrain the downstream task of material-property estimation. Videos have several advantages over existing NDT techniques: while contact

Authors' addresses: Berthy Feng, California Institute of Technology, Pasadena, USA, bfeng@caltech.edu; Alexander C. Ogren, California Institute of Technology, Pasadena, USA, aogren@caltech.edu; Chiara Daraio, California Institute of Technology, Pasadena, USA, daraio@caltech.edu; Katherine L. Bouman, California Institute of Technology, Pasadena, USA, klbouman@caltech.edu.

sensors and laser vibrometers take point measurements, videos offer spatially dense measurements of surface vibrations. While laser vibrometers are expensive and specialized, cameras are ubiquitous and general-purpose. While digital image correlation (DIC) requires stereo cameras for 3D motion tracking, our method shows that in many cases, a monocular view is all you need.

When objects vibrate with small displacements, their motion can be decomposed into independent *modes* at *resonant frequencies*. As a simple example, when a guitar string is plucked, it vibrates in a standing wave at its fundamental frequency and integer multiples of the fundamental, but imperfections in the string cause more complex movement at resonant frequencies. Under known geometry, a set of vibrational modes determines an object's material properties up to a scaling factor and vice versa; this is a key fact that motivates our approach to solving for material properties from motion features. Surface motion can be extracted from video and decomposed into periodic motion at different frequencies [Davis et al. 2014]. These decomposed motion fields are 3D modes projected onto 2D image-space, which we refer to as *image-space modes*. The key challenge of our task is to deal with incomplete and 2D (as opposed to full-field) modes. Despite these challenges, we show that we are able to recover material properties from image-space motion and recover full-field modes in the process.

In this paper, we first review related work and the theoretical relationship between modes and material properties. We then show how to extract image-space modes from video and recover material properties and full-field modes from these measurements (Fig. 1 shows an overview of the method pipeline). Finally, we demonstrate our approach on simulated data generated from a variety of 3D and 2D geometries. As a proof-of-concept experiment on real-world data, we introduce material inhomogeneities to the bottom of a real drum head and show that we are able to image these defects, even though they are unseen in the captured video.

2 RELATED WORK

In this section, we review traditional NDT, as well as recent works in computer vision and graphics that inspire our visual approach to vibration tomography.

2.1 Traditional Material Characterization

A non-destructive way to characterize materials is to study their movement in response to non-damaging, external forces. To that end, laser vibrometry and digital image correlation (DIC) are popular *non-contact* means to measure surface displacements.

A laser Doppler vibrometer sends laser beams towards the moving surface of interest and then, applying the Doppler effect, recovers displacement and velocity directly from the frequency and phase changes of the backscattered light [Durst et al. 1981]. Laser vibrometry has been used for a variety of applications, from monitoring bridges [Nassif et al. 2005] and buildings [Roozen et al. 2015], to testing composite steel [Emge and Buyukozturk 2012] and concrete [Chen et al. 2014a], and even to assessing the health of fruits [Landahl and Terry 2020; Santulli and Jeronimidis 2006] and trees [Zorović and Čokl 2015]. A single laser vibrometer is limited in its measurement capacity: it can only measure 1D motion along

a single surface point at a time. Three independent laser beams can be used to measure 3D vibration [Khalil et al. 2016], and a scanning laser vibrometer can be used to scan a beam across the surface, but these upgrades quickly become prohibitively expensive.

DIC, similarly to our method, computes displacements from digital images. Sub-regions of the image are tracked across time, and local 3D displacements and strains are resolved [Chu et al. 1985]. DIC has been used to identify defects in materials including glass [Speranzini and Agnetti 2014], concrete [Feiteira et al. 2017; Helm 2008; Wu et al. 2011], and masonry [Tung et al. 2008]. Such defects are identified as regions of abnormal strain – not through solving an inverse problem. Measuring 3D strains also requires at least two cameras, whereas our framework uses a monocular view.

Modal analysis is a common downstream task of laser vibrometry [MacPherson et al. 2007] and DIC [Ha et al. 2015; Helfrick et al. 2011; Trebuňa and Hagara 2014], among other NDT methods. While usually regarded as a verification tool, rather than a means to directly infer material properties, modal frequencies [Gentile and Saisi 2007; Girardi et al. 2020] and shapes [Beardslee et al. 2021] have been used to solve for *homogenized* material properties. Our work presents a new method to (1) identify modes in image-space and (2) directly infer spatially heterogeneous material properties from image-space modes.

In the graphics community, systems have been specially built to probe material properties of real-world objects for downstream simulation of their behavior. However, they are often restricted to homogenized material properties and particular object classes (e.g., fabrics) [Bickel et al. 2010; Miguel et al. 2012; Wang et al. 2011]. Previous works have inferred material properties from 3D point clouds [Kim et al. 2017; Wang et al. 2015] and known external forces [Xu et al. 2015].

2.2 Material Characterization from Video

In computer vision, scene understanding is an important goal that includes, among many other tasks, classification and characterization of materials. Static images have been used to classify [Liu et al. 2010] and describe [Ho et al. 2006; Sharan et al. 2008] materials. Using video to identify the *dynamic* properties of materials has been recently explored with applications for fabrics [Bhat et al. 2003; Bouman et al. 2013] and trees [Wang et al. 2017]. Davis & Bouman et al. introduced the idea of "visual vibrometry" [Davis et al. 2017, 2015a], using frequencies extracted from a video's motion spectrum to estimate stiffness and damping of fabrics and rods with known geometry. Videos have also been used for modal analysis [Chen et al. 2014b, 2015; Harmanci et al. 2019; Yang et al. 2020] and structural health monitoring [Sarrafi et al. 2018]. However, none of these past works estimate material properties. Our work leverages information from both frequencies and modes to quantify the *heterogeneous* material properties of *unseen* parts of objects.

2.2.1 Small motions. Since our approach relies on small, often imperceptible, motions, we need a way to extract sub-pixel motions from video. Several works in vision and graphics that have been published in SIGGRAPH [Liu et al. 2005; Wadhwa et al. 2013; Wu et al. 2012] have developed ways to magnify small motions [Wadhwa et al. 2014]. As proposed by Wadhwa et al. [Wadhwa et al. 2013], local
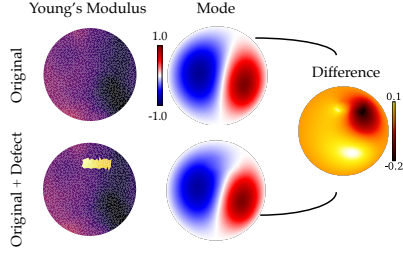
Fig. 2. Small changes in material properties affect motion. Here a small region of a circular membrane becomes stiffer from "Original" to "Original + Defect." This change appears as a slight change in the mode shown. In this paper, we propose a technique to use small changes in observed modal motion to recover the locations and magnitudes of defects in an object.

phase shifts in a complex steerable pyramid are computed [Portilla and Simoncelli 2000; Simoncelli and Freeman 1995; Simoncelli et al. 1992]. This phase-based method has the advantage over other tracking methods (e.g., optical flow) of being robust to extremely small motions, down to one thousandth of a pixel [Davis et al. 2015a]. The phase shifts can be converted to pixel displacements using the approach originally discussed by Fleet and Jepson [Fleet and Jepson 1990; Wadhwa et al. 2017].

The phase-based approach for motion extraction was used by Davis et al. [Davis et al. 2014] to convert small motions of everyday objects into ambient sound signals. It was also used to extract image-space modal bases from video for simulating plausible motion of objects [Davis et al. 2015b]. In our work, we apply the same process to extract image-space modes and then directly infer underlying physical properties from these measurements.

3 MODELING OBJECT MOTION

3.1 Modal Analysis

Every object has resonant, or natural, frequencies. At each resonant frequency, the object vibrates periodically in a particular shape, called a mode. When you hit a drum, for example, you send an impulse force to the drum head, which excites a range of resonant frequencies in the membrane. The membrane responds by moving up and down in a complicated way. When the displacements are small, this motion can be decomposed into modes occurring at independent frequencies.

The finite element method (FEM) models an object as a mesh, composed of elements that each take on material-property values. The material properties that determine an object's vibration are Young's modulus (E), Poisson's ratio, and density (ρ). Young's modulus and Poisson's ratio define the stiffness of connections between vertices, while density defines the mass distribution over all vertices. One can model a mesh as an interconnected system of springs and masses, where the stiffness matrix K and mass matrix M describe those connections. K and M can be computed using an FEM solver.

The stiffness matrix and mass matrix fully determine the modes and resonant frequencies of an object with known geometry. A mode u and resonant frequency ω are an eigenvector-eigenvalue pair solving the generalized eigenvalue problem:

$$Ku = \omega^2 Mu. \tag{1}$$

The mode u describes the motion of the mesh at a resonant frequency ω. 3D objects have three directions of motion, meaning that there are three degrees of freedom (DOFs) per mesh vertex. The vector u therefore contains $3 \times (\# \text{ vertices})$ entries, giving the direction and relative magnitude of motion for every DOF on the exterior and interior of the object. When the object vibrates, its motion is a linear combination of its modes.

As Fig. 2 illustrates, a small change in material properties within a fixed geometry appears as a small change in modal motion. As most solid materials have Poisson's ratio ≈ 0.3 [Poplavko 2019], the principal material properties affecting motion are Young's modulus, which determines K, and density, which determines M. Our method is based on the insight that mode shapes on the surface of an object may reveal internal spatial inhomogeneities in these properties.

3.2 Inferring Material Properties from Motion

In solving the inverse problem, we would like to infer the material-property values that cause observed motion. Assuming we perfectly measure all modes u and frequencies ω, then by Eq. 1, we can set up the following minimization problem:

$$K^*, M^* = \arg\min_{K,M} \|KU - MU\Lambda\|_2^2, \tag{2}$$

where U is the matrix whose columns are modes u, and Λ is the diagonal matrix containing eigenvalues ω^2. For a known geometry, this is a convex problem with respect to K and M (with the constraints on K and M specified in Eq. 4). However, we will be working with experimentally-observed, image-space modes, which incur the following challenges:

(1) *Unseen degrees of freedom (DOFs).* When we cannot observe the interior or all surfaces of an object, we do not have modal information for all of its vertices. For example, when observing a 3D cube with a single monocular camera, one can see at most three of its sides, projected onto two directions of motion. Consider an 8x8x8 cubic mesh, which has $(8+1)^3 = 729$ vertices. Since it has three directions of motion, it has $3 \times 729 = 2187$ total DOFs. But a single monocular view of three sides of the cube can only observe 217 vertices, moving in two directions of motion, amounting to $2 \times 217 = 434$ image-space DOFs. This alone limits us to observing *fewer than 20%* of the full-field DOFs for an 8x8x8 cube.

(2) *Unseen modes.* Theoretically, for discrete meshes, there are as many modes as there are DOFs. However, we can only capture modes at resonant frequencies below the Nyquist sampling rate of the camera. For a video with frames-per-second FPS, the Nyquist sampling rate is FPS/2.

(3) *Noise.* Aside from camera noise, there is noise from motion extraction, particularly in non-textured regions of the image.

As a consequence of having limited numerical data, the problem of solving for K and M from the observed data (Eq. 2) is ill-posed.[1] That is, many solutions for K and M may match the projected modes. More concretely, let n be the number of total DOFs and $n' \ll n$ be the number of observed DOFs. If we organize the theoretical

---

[1] For a known geometry and complete mode and eigenvalue information, K and M are fully determined up to a scaling factor.

modes as columns of a matrix $U \in \mathbb{R}^{n \times n}$, and we observe k modes $u_1, \ldots, u_k \in \mathbb{R}^{n'}$, then – ignoring projection onto image-space for now – we are only able to measure an $n' \times k$ submatrix of the $n \times n$ matrix U. Similarly, we are only given $k << n$ eigenvalues $\omega_1^2, \ldots, \omega_k^2$. Fig. 3 visualizes the matrices involved. In the following section (Sec. 4), we explain how to deal with these limitations.
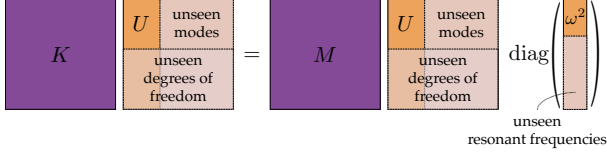

Fig. 3. The generalized eigenvalue equation (Eq. 1) defines the relationship between K, M and U, ω^2. The matrix U has columns corresponding to modes and rows corresponding to DOFs (i.e., a direction of motion for each mesh vertex). The vector ω^2 contains associated eigenvalues. We would like to solve for K and M given partial information about U and ω^2.

4 APPROACH

Our aim is to extract and analyze motion features from a video to estimate material properties. This involves two stages: (1) motion extraction and image-space mode identification, and (2) solving for material properties that best match the observed image-space modes. The input is a video of a vibrating object, of which a mesh is known, and we assume that it is vibrating under linear elasticity (i.e., small motions). The output is a 3D volume showing voxelized Young's modulus and density values throughout the object.

4.1 Extracting Image-Space Modes from Video

4.1.1 Motion extraction. From an input video (which can be synthetic or real), we compute local motion across the image over time. Using complex steerable pyramids, we decompose each frame into different spatial scales and orientations. The response of the original image to each filter is represented as a complex image showing local amplitude and phase. In each sub-band, we compute the local phase change that occurred since the first frame. Using the method detailed by Wadhwa et al. and Fleet and Jepson [Fleet and Jepson 1990; Wadhwa et al. 2017], we convert these phase changes to pixel-wise displacements. To increase SNR, we first filter out outlier pixels (i.e., top 1% of displacement magnitudes) and then apply an amplitude-weighted Gaussian blur to reduce noise. The result of this motion-extraction step is a motion field for each frame, which quantifies the horizontal and vertical displacement of each pixel relative to the first frame.

4.1.2 Identifying image-space modes. Modes are simply periodic motions occurring at particular frequencies, so we would expect them to appear as peaks in the power spectrum of extracted motion amplitudes. As is done in previous works that extract image-space modes [Davis et al. 2017, 2015b, 2014], we perform a discrete Fourier transform of the motion fields to extract the mode shapes. Our motion-field data is stored as a three-dimensional array of size (T, h, w), where T is the number of frames, and h and w are the image height and width. We take the FFT along the time-axis of

this array, resulting in a complex-valued image showing the amplitude and phase for each frequency bin. To make this more concrete, let $\Delta x_t(x, y)$ and $\Delta y_t(x, y)$ be the horizontal and vertical displacements from the first frame, respectively, of pixel (x, y) at frame t. We take the one-dimensional FFT of $\Delta x_t(x, y)$ and $\Delta y_t(x, y)$, resulting in two (complex-valued) images, $\widehat{\Delta x}_\ell(x, y)$ and $\widehat{\Delta y}_\ell(x, y)$, for each frequency bin $\ell = 1, \ldots, T$. From here, we compute the power spectrum by plotting magnitudes, $\left\| \left[\widehat{\Delta x}_\ell, \widehat{\Delta y}_\ell \right] \right\|_2^2$, versus frequency, (FPS · ℓ/T) Hz. (The (x, y) notation has been dropped for clarity.) A peak ℓ^* in the spectrum ideally corresponds to a resonant frequency (FPS · ℓ^*/T) Hz and image-space mode $\left[\mathrm{Re} \left(\widehat{\Delta x}_{\ell^*} \right), \mathrm{Re} \left(\widehat{\Delta y}_{\ell^*} \right) \right]$.

4.1.3 Sampling image-space modes at mesh vertices. Since we model the object of interest as a mesh, we need a way to pinpoint where mesh vertices appear in the image. The 3D vertices are mapped by a 3D-to-2D projection matrix to their image-space locations. To approximate this projection matrix, a user manually identifies the pixel locations of several "reference" mesh vertices and computes the projection matrix P that best maps the corresponding 3D coordinates from the mesh to the image.

Using P, we map all of the mesh vertices from their 3D coordinates to 2D image coordinates. We then sample each image-space mode at the pixel locations of *visible* mesh vertices. For mode j, we construct a vector γ_j that contains the horizontal and vertical displacement of each mesh vertex at the corresponding resonant frequency. Supposing we observe q' out of q mesh vertices, the vector γ_j has the form

$$\gamma_j = \left[\Delta x_1, \Delta y_1, \ldots, \Delta x_{q'}, \Delta y_{q'}, 0, \ldots, 0 \right]^\top \in \mathbb{R}^{2q}, \quad (3)$$

where Δx_i is the horizontal (pixel) displacement and Δy_i the vertical (pixel) displacement of vertex i. Unseen vertices are assigned displacements of 0, and for notational clarity, we choose to position them at the end of the vector.
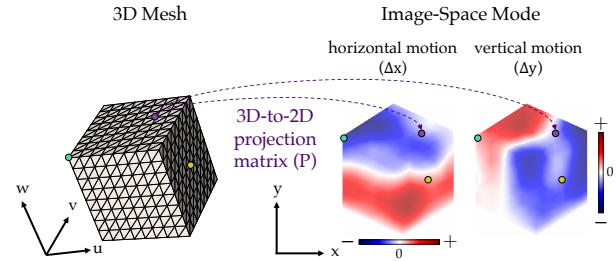

Fig. 4. P is a 3D-to-2D projection matrix mapping 3D mesh vertices to 2D image coordinates. Once we have mapped surface mesh vertices to locations in the image, we can sample those points in the image-space modes.

4.2 Estimating Material Properties

4.2.1 Matrices K and M as functions of material properties. As discussed in Sec. 3.1, the matrices K and M are functions of Young's modulus and density (as well as geometry, which we assume is already known). Computing them by hand is non-trivial, so an FEM solver such as FEniCS [Alnæs et al. 2015] can be used to assemble these matrices. K and M are typically expressed as *global* matrices,

which are sparse and contain information for all vertices and edges in the mesh. However, they can be decomposed into *local* matrices, which contain edge information for a local collection of mesh elements. The global K and M can both be written as a sum of local matrices, each of which can also be assembled with FEniCS.

One useful property is that these local matrices scale linearly with local material properties. We can voxelize the volume containing the mesh such that each voxel contains a collection of mesh elements. Suppose that a voxel has a local Young's modulus $E = 1$ and associated local stiffness matrix K_e. If the local Young's modulus is changed to $E = 100$, then the new local stiffness matrix becomes $100 \cdot K_e$. The same is true for mass matrices. Given Young's modulus values w_e and density values v_e for each voxel, we can express the global stiffness matrix and global mass matrix as

$$K = \sum_{e=1}^{m} w_e K_e \quad \text{and} \quad M = \sum_{e=1}^{m} v_e M_e. \quad (4)$$

This allows us to represent K and M as functions of vectors $w, v \in \mathbb{R}^m$, which contain voxel-wise Young's moduli and densities. The complexity of our variables has now been reduced from $O(n \times n)$ to $O(m)$, where n is the total number of DOFs and m the number of voxels covering the mesh ($m << n$).

4.2.2 Optimization formulation. Supposing we have modes $u_1, \ldots, u_k$ and their frequencies $\omega_1, \ldots, \omega_k$, we would like to determine the voxel-wise Young's modulus values w and density values v such that, for each $i = 1, \ldots, k$ (refer to Eq. 1),

$$\left(\sum_{e=1}^{m} w_e K_e \right) u_i \approx \omega_i^2 \left(\sum_{e=1}^{m} v_e M_e \right) u_i. \quad (5)$$

Before we can formulate Eq. 5 as an optimization problem, we have to address that our observations are modes that have been *projected* onto image-space. Therefore, we need to recover the 3D modes u_i from the 2D projected modes γ_i (Eq. 3). For a given observation γ_i, it should be the case that

$$GPu_i \approx \gamma_i, \quad (6)$$

where P is the 3D-to-2D projection matrix discussed in Sec. 4.1.3, and G is a binary matrix that sets unseen DOFs to 0. Remember that γ_i also contains some noise from the camera and motion-extraction step, so we want to avoid overfitting to γ_i.

Given Eqs. 5 and 6, we have the following optimization problem:

$$w^*, v^* = \underset{\substack{w, v \in \mathbb{R}^m \\ u_i \in \mathbb{R}^n, i=1,\ldots,k}}{\operatorname{argmin}} \sum_{i=1}^{k} \left\| \left(\sum_{e=1}^{m} w_e K_e \right) u_i - \omega_i^2 \left(\sum_{e=1}^{m} v_e M_e \right) u_i \right\|_2^2$$

$$+ \sum_{i=1}^{k} \| GPu_i - \gamma_i \|_2^2. \quad (7)$$

Note that ω_i and γ_i are the i-th observed frequency and image-space mode and will contain noise.

As discussed in Sec. 3.2, Eq. 7 is under-determined, so we must introduce regularization to make the solution well-defined. We choose to minimize total squared variation (TSV) of w and v, which ensures that the material properties vary minimally between adjacent voxels. Moreover, since we are estimating both stiffness and mass, the objective function can become arbitrarily low if we do not constrain one of these material properties to a certain range of values. To see this, consider the eigenvalue equation in Eq. 1. Scaling M and K in this equation by a factor of s still satisfies the equation: $(sK)u = \omega^2(sM)u$. To resolve this ambiguity, we choose to minimize the deviation of the Young's modulus vector w from a mean value $\bar{w}$. Regardless of $\bar{w}$, the relative differences in w^*, v^* will not change. For defect detection, we generally only care about relative changes in material properties. With regularization, the full optimization is

$$w^*, v^* = \underset{\substack{w, v \in \mathbb{R}^m \\ u_i \in \mathbb{R}^n, i=1,\ldots,k}}{\operatorname{argmin}} \frac{1}{2k} \sum_{i=1}^{k} \left\| \left(\sum_{e=1}^{m} w_e K_e \right) u_i - \omega_i^2 \left(\sum_{e=1}^{m} v_e M_e \right) u_i \right\|_2^2$$

$$+ \frac{\alpha_u}{2k} \sum_{i=1}^{k} \| GPu_i - \gamma_i \|_2^2 \quad (8)$$

$$+ \frac{\alpha_w}{2m} \| \nabla w \|_2^2 + \frac{\alpha_v}{2m} \| \nabla v \|_2^2 + \left(\frac{1}{m} \left(\sum_{e=1}^{m} w_e \right) - \bar{w} \right)^2,$$

where α_u, α_w, and α_v are hyperparameters that determine the relative weights of the observation-matching term, regularization on w, and regularization on v, respectively.

4.2.3 Optimization procedure. Eq. 8 is a nonconvex problem, but it is quadratic with respect to w, v when u_i is fixed, and it is quadratic with respect to u_i when w, v are fixed. Our optimization procedure is to iteratively compute the closed-form solution for $U = [u_1 \ldots u_k]$ and then $z = [w^\mathsf{T}, v^\mathsf{T}]^\mathsf{T}$, thereby minimizing the objective function at each step. For an 8x8x8 cube, one iteration takes about 10 seconds (tested on 8-core Intel Core i9, 32 GB RAM). Once the value of the objective function has converged — which typically takes fewer than 100 iterations — we output the minimizing solution $z^* = [w^{*\mathsf{T}}, v^{*\mathsf{T}}]^\mathsf{T}$. The vectors w^* and v^* define the voxel-wise reconstructed Young's modulus and density values.

5 SYNTHETIC EXPERIMENTS

With FEM modeling, it is possible to generate physics-based simulations of any object defined by a mesh. In our synthetic experiments, we test our approach on videos of simulated vibration of various objects, ranging from 2D surfaces to more complex 3D structures.

5.1 Creating Synthetic Data

By voxelizing the volume containing a mesh, we can simulate the modal vibration of any mesh geometry. In this paper, we demonstrate our proposed method on synthetic 2D drum heads, 3D cubes, and the Stanford bunny.

5.1.1 Mesh Geometries.
Drums: The circular drum head is modeled as a triangular membrane mesh, where the mesh vertices each have one DOF. We assign material properties on a pixel grid, where each mesh element is given the properties of its nearest pixel. In these experiments, we use a mesh with 1530 (linear) triangular elements, laid on a 20x20 pixel grid. Each pixel within the circle covers about 5 elements.
Cubes: The cube is modeled as a tetrahedral mesh, and we assign material-property values on a voxel grid. For an 8x8x8 grid, for example, we assign a Young's modulus and density to each of the $8^3 = 512$ voxels. Each voxel covers six (linear) tetrahedral elements,
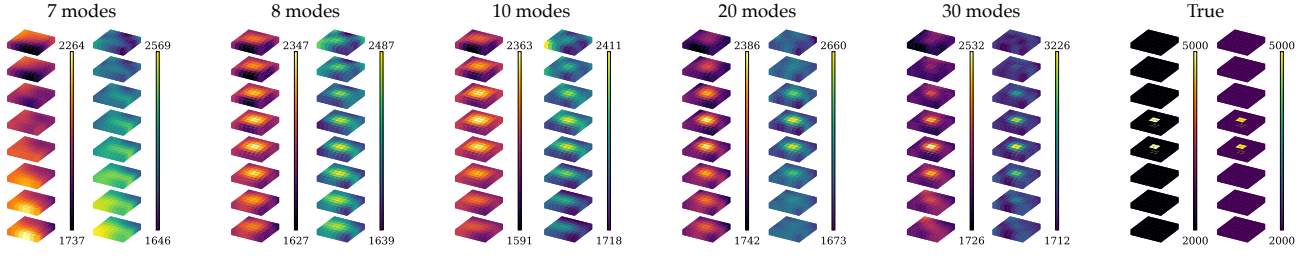
Fig. 5. Changing the number of observed image-space modes. From a synthetic video of an impulse 8x8x8 cube, more than 30 image-space modes were extracted. Keeping hyperparameters fixed, we estimated material properties given the 5, then 10, then 20, and then 30 extracted image-space modes at the lowest resonant frequencies. This "impulse" cube has a jump in material properties in its core, which is not detected until more than 7 modes are used. For each example, Young's modulus is shown on the left, and density is shown on the right.
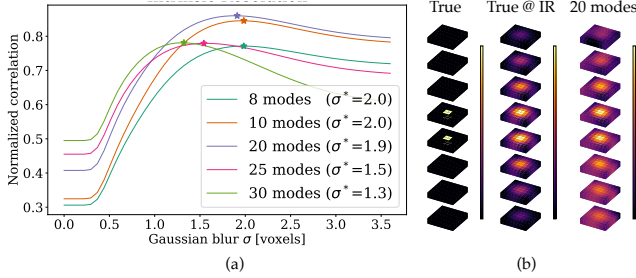
Fig. 6. Intrinsic resolution of reconstructed volumes. Based on normalized correlation with the ground-truth material properties smoothed at different scales, one can approximate the intrinsic resolution of the reconstructed material properties. In (a), we plot normalized correlation versus Gaussian blur standard deviation σ, for the reconstruction of Young's modulus using different numbers of image-space modes (keeping all other hyperparameters fixed). As the number of observed modes increases, the reconstructed resolution also increases (i.e., smaller blur σ). Subfigure (b) shows the true Young's modulus volume blurred at the intrinsic resolution (IR) of the reconstruction given 20 image-space modes ($\sigma^* = 1.9$ voxels).

so the mesh contains 3072 elements.

Bunny: The Stanford Bunny is modeled as a tetrahedral mesh.[2] We voxelize the volume containing the mesh into an 8x8x8 grid, resulting in about 21 mesh elements per voxel.

5.1.2 Defining material properties: We consider two cases for the way material properties are spread across the mesh: smooth and discontinuous. In the smooth case, material properties vary smoothly between elements. To generate examples, we draw random samples from a multivariate Gaussian distribution with a Matern covariance matrix, which allows for correlation between elements at different spatial scales. For discontinuous examples, we assign a primary Young's modulus and density across the mesh, but with an "impulse" jump in material properties in the middle of the mesh. For 3D objects, these impulse defects are not visible on the surface.

5.1.3 Simulating movies. For the objects we consider, we create physics-based simulations of their vibration. After meshing and

[2]The surface mesh is from https://www.thingiverse.com/thing:151081, and tetrahedralization was done using TetWild [Hu et al. 2018].

defining material properties, we assemble the global stiffness matrix

K and global mass matrix M using FEniCS, and then use scipy's built-in eigen-solver to solve for the r lowest modes u (e.g., $r = 50$). The motion is synthesized as a linear combination of the modes $u_1, \ldots, u_r$. Letting p_0 be the vector containing the 3D positions of all the mesh vertices at time 0, the position p_t at time t is

$$p_t = p_0 + \sum_{i=1}^{r} A_i \sin(2\pi f_i t) \cdot u_i, \qquad (9)$$

where A_i and f_i are the amplitude and frequency, resp., of mode i. Note that in these simulations, we assume free vibration without damping and directly prescribe the amplitude of each mode. Using time-dependent analysis, more realistic simulations can be made of the object's reaction to an external force.

To create texture for motion extraction, we plot random scatter points on the surface of the cube (see Fig. 1 for an example frame). For the drum, simply plotting the mesh vertices and edges provides enough texture. Using matplotlib to plot surface points and OpenCV and imageio for image I/O, we input the resulting animations to our material-estimation pipeline (Sec. 4).

5.2 Reconstruction from Synthetic Data

In this subsection, we show a series of experiments run on synthetic data. Due to space limitations, we choose to highlight one geometry for each experiment. Unless otherwise specified, all results are using (noisy) motion-extracted image-space modes.

5.2.1 Effect of number of modes. As more image-space modes are included, the inverse problem becomes better constrained. Therefore, we expect that adding more image-space modes will improve the material-property reconstruction. Fig. 5 shows how the Young's modulus and density reconstructions of a cube, whose defect lies in its interior, improves as the number of image-space modes increases. Note that, for this cube, more than seven image-space modes are required before an interior defect can be detected.

Because of the need for regularization in the inverse problem, the estimated material properties are often smoother than reality. As Fig. 6 shows, in terms of normalized correlation, there is an optimal blur kernel at which the ground-truth closely correlates with the estimated material properties. Since normalized correlation with the ground-truth is robust to scaling, it is an appropriate metric for our technique, which only recovers relative changes in material
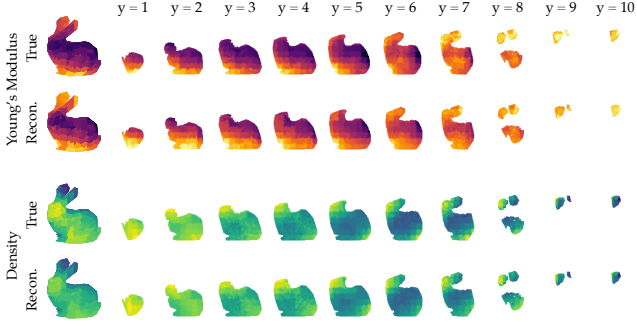
Fig. 8. Reconstruction for a bunny mesh. Given a mesh of the Stanford bunny, we voxelize the volume containing the mesh as an 8x8x8 grid. Using the same approach for the synthetic cubes, we are able to recover Young's modulus and density from (true) image-space modes. Slicing along the y-axis reveals interior material properties.

properties. Given 20 motion-extracted image-space modes (for an 8x8x8 voxel impulse cube), the intrinsic resolution has a blur kernel of $\sigma \approx 1.9$ voxels, and this intrinsic resolution becomes sharper (i.e., smaller σ) as the number of given image-space modes increases.

5.2.2 Regularization. The strength of spatial regularization on material properties affects the smoothness of the estimation. In Eq. 8's optimization formulation, the smoothness regularization terms are $\|\nabla w\|_2^2$ and $\|\nabla v\|_2^2$, which have scalar weights α_w and α_v, respectively. A lower regularization weight α_w or α_v will result in a less-smooth image of estimated material properties. In Fig. 7, we show how the image of estimated density becomes sharper as α_v decreases while keeping α_w fixed. This can result in a crisper picture of a defect, but can also make reconstruction more sensitive to noise.
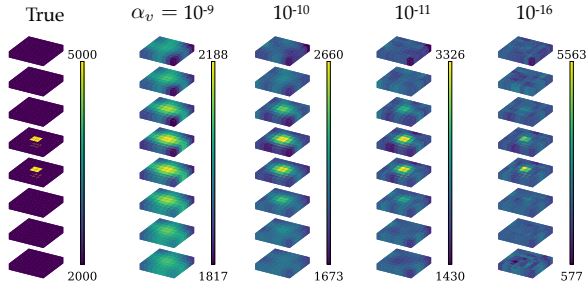

Fig. 7. Effect of regularization. The hyperparameters α_w and α_v control the weight of the smoothness regularization terms in Eq. 8. Here we show the estimated density for a cube with an impulse in the center. Keeping $\alpha_w = 10^{-10}$ fixed, as α_v becomes smaller, the image of the defect becomes crisper, but more sensitive to noise. These estimations were done from the same 20 motion-extracted, image-space modes.

5.2.3 Effect of spatially correlated noise. Our current optimization formulation only accounts for i.i.d. Gaussian noise in the image-space modes (Eq. 8). Since noise in the extracted motion fields is likely to be correlated in image-space, it is important to investigate the effect of spatially correlated noise. We study this by adding different types of noise to the true image-space modes, with noise falling under one of three categories:

(1) *i.i.d. Gaussian noise.* All DOFs in γ (Eq. 3) have i.i.d. Gaussian noise with mean 0.
(2) *Correlated Gaussian noise.* The random noise applied to each DOF in γ is correlated with the noise of nearby surface points in the image. To better model realistic noise, we synthesize a noise field by drawing a random sample from a multivariate Gaussian distribution with mean 0 and Matern covariance matrix. By tuning the correlation scale of the Matern covariance function, we can consider noise at different correlation lengths. A higher correlation length causes noise to be similar across a larger region of the image, while a correlation length of 0 corresponds to i.i.d. noise.
(3) *Noise from motion extraction.* In this case, we use the image-space modes obtained through motion extraction (Sec. 4.1).

The image-space modes with artificial noise all have an average PSNR of about 42 dB , and the motion-extracted modes for this example contain a PSNR of 36 dB. The correlation length of the noise appears to be a major factor in the quality of reconstruction. Fig. 9 shows the imaged material properties, given these different types of noise, for a drum with a discontinuous jump in material properties. As the correlation length increases, the estimated image of material properties becomes more prone to error; even so, the defects appear strongly in the recovered image.

5.2.4 Inference resolution. One of the hyperparameters of our method is the resolution of the grid used to infer material properties. Since we cannot expect to match the infinite resolution of the real world, we test our approach on situations where data is generated on a finer grid than the inference grid. Fig. 10 shows the reconstruction of a smooth cube as the inference resolution is decreased incrementally from 8x8x8 to 5x5x5, keeping the source of data fixed. As we would hope, the reconstruction degrades smoothly as the resolution becomes coarser. An additional demonstration of geometric model mismatch is shown in Fig. 11. As the inference geometry becomes less accurate, the reconstruction quality decreases. A mismatch of 10% still identifies large defects.

5.2.5 Predicted image-space modes. Another way to assess reconstructed properties is to verify that they produce the same image-space modes as the true properties. We can do so by running the FEM forward model on the estimated Young's moduli and densities and looking at the resulting predicted image-space modes. Also recall that 3D modes are a decision variable in our optimizaton scheme (Eq. 8). As Fig. 12 shows, it is informative to compare the true FEM modes, observed modes U, optimized modes, and predicted FEM modes, in image-space. The optimization process usually de-noises observed modes. For some modes, spatially correlated noise may make it difficult to recover the true mode. In such cases, it is possible for the predicted FEM modes to still be similar to the truth.

5.2.6 Complex geometries. Our method can be applied to any geometry, simply by voxelizing the volume containing the mesh. Fig. 8 shows a volume reconstruction of material properties for the Stanford bunny. Twenty true image-space modes of the monocular view of the bunny shown in the figure were given for this reconstruction.
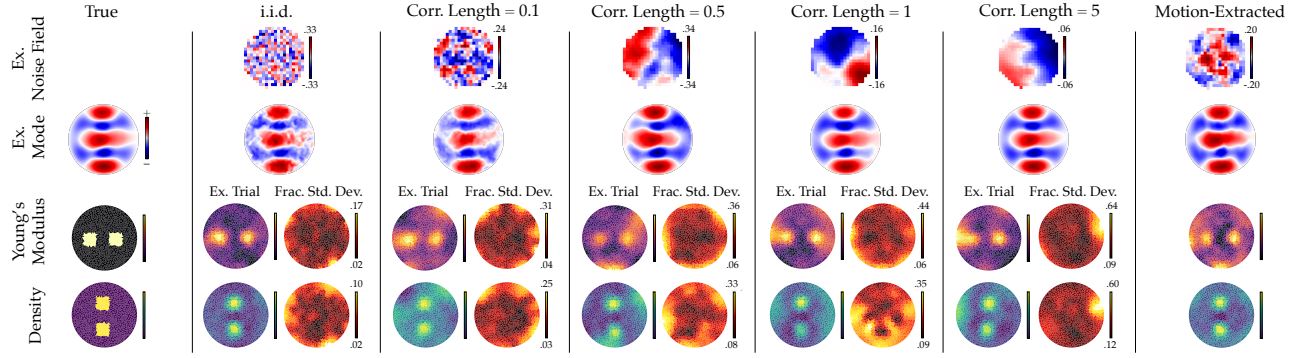
Fig. 9. Effect of i.i.d., correlated, and motion-extracted noise. We first solved for the true image-space modes given the true material properties. For each column (except for "True"), we added noise to the true image-space modes (reconstruction using 10 modes shown here). "Corr. Length" is the correlation scale of the Matern covariance function used to synthesize a random noise field. In "Motion-Extracted," these 10 image-space modes were extracted from a video synthesized as vibrations at the 50 lowest true modes. The top row ("Ex. Noise Field") shows the noise field that was applied to one of the true image-space modes. The second row ("Ex. Mode") shows the resulting image-space mode that was given as an observation to optimize material properties. For "Young's Modulus" and "Density," we show an example reconstruction and the fractional standard deviation (σ/μ) across 10 trials of random noise applied to the true image-space modes. In the case of "Motion-Extracted" image-space modes, we show the reconstruction from the synthesized video.
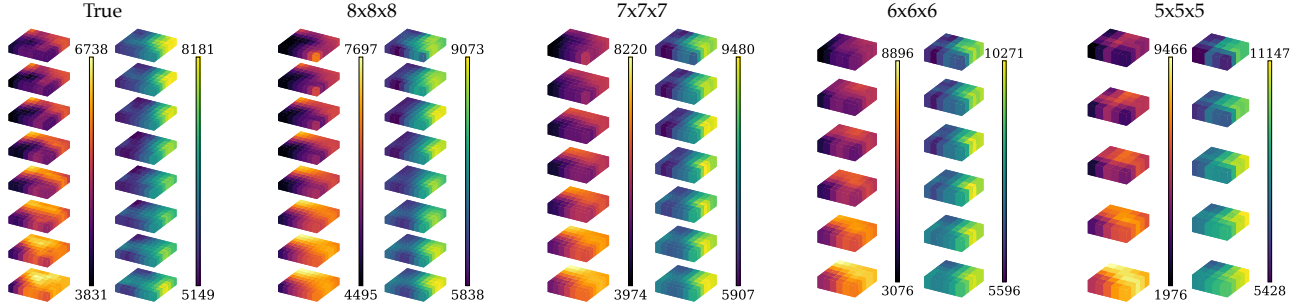

Fig. 10. Mesh resolution model mismatch. After synthesizing a video for a cube with smooth material properties defined over an 8x8x8 voxel grid, we infer the material properties on a mesh of various resolutions. This introduces model mismatch in our mesh. A coarse, yet still reasonable, inference resolution captures large-scale variations but loses fine detail. For each example, Young's modulus is shown on the left, and density is shown on the right.
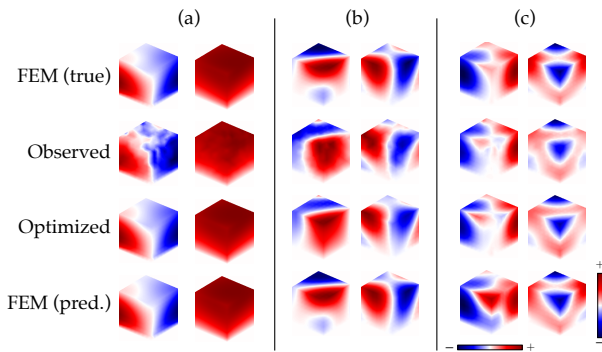

Fig. 12. Similarity of predicted image-space modes to true, observed, and optimized image-space modes. The "Observed" mode is often a noisy version of "FEM (true)." "Optimized" refers to the optimized solution U^* in Eq. 8. "FEM (pred.)" is the mode resulting from the estimated material properties.
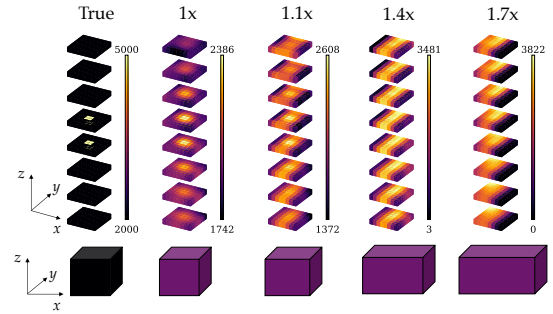

Fig. 11. Geometric model mismatch. After synthesizing a video of a square cube with impulse material properties, we infer on a mesh of various incorrect geometries. In particular, we extend the inferred geometry width by a multiple of the true width. This introduces model mismatch: fitting to a rectangular prism rather than a cube. Results on Young's modulus show that geometric mismatch of 10% still performs reasonably well.

6 REAL-WORLD EXPERIMENTS

As a proof-of-concept, we created real-world versions of the drum head and cube geometries used in synthetic experiments.

6.1 Real Drums

We tested our method on a dataset of real drum heads, each altered with an unseen defect on its bottom surface. Although we are able to see all DOFs of the drum surface in the video, the problem of solving for material properties is still ill-posed because we observe a limited number of projected modes.

6.1.1 Drum construction. The drums were constructed by fixing a thin rubber sheet over a 4"x4" PVC adaptor with a rubber band. We tested six defects with two different materials: nail hardening gel and acrylic plastic circles. The defects included circles of varying sizes and locations, and a rectangle. For the gel defects, we painted the shape onto the bottom side of the rubber sheet and set the gel under a UV lamp. The acrylic had been pre-cut into circles of radius 1.85 cm, which we glued to the bottom of the rubber using Gorilla Glue. For each defect, we recorded a video of the homogeneous drum before the defect was applied, for direct comparison to the drum after the defect was introduced. We drew a speckle pattern on the drum head for texture.

6.1.2 Vibration-capture setup. Fig. 13 shows a schematic of the experimental setup. We taped the drum onto an optical table, with the high-speed camera standing on the same optical table. The excitation source was a PreSonus Sceptre S8 loudspeaker, which sat on a platform separate from the optical table and was pointed at the drum. For each video, we recorded the drum head's vibration in response to a 3.5-second linear frequency sweep (50 Hz to 1000 Hz) being played through the speaker.
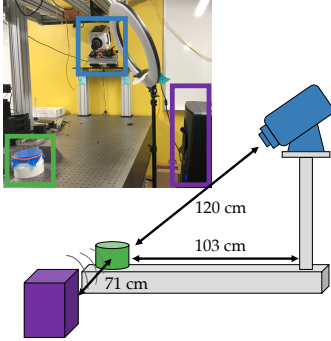

Fig. 13. Experimental setup for real drum heads. Vibrations of the drum are induced by a loudspeaker and recorded with a high-speed camera.

6.1.3 Video capture. To capture monocular video, we used a Phantom V1610 high-speed camera. Each video was captured at 6000 FPS at an image resolution of 288×384. To reduce camera noise, we averaged every two frames for a resulting temporal frequency of 3000 FPS. Note that in Fig. 14, the drums vibrate at frequencies below 120 Hz. While we chose to first demonstrate our approach using a high-speed camera, where compression and camera noise are less challenging, many modal frequencies can be captured on a consumer camera.

6.1.4 Extracting image-space modes. We followed the step-by-step process outlined in Sec. 4.1 by first extracting motion fields throughout the video and then identifying image-space modes. We found that in real videos, some level of manual selection was necessary to verify peaks in the motion amplitude spectrum as modal motion (without the aid of a more sophisticated peak-selection algorithm). For instance, spurious camera motion would often appear as spikes in the spectrum. Verification was done by visually inspecting the magnified motion in the video at the frequency in question (following the method proposed in [Wadhwa et al. 2013]). Fig. 14 shows some image-space modes for the drum with a gel bar as its defect, before and after the defect was introduced. The defect does not appear in the low modes, but in higher modes, one can see regions of slightly smaller motion than the rest of the drum head.

We modeled the drum as a triangular mesh with 1530 linear elements and inferred material properties on a 20x20 pixel grid. To sample image-space modes at the mesh vertices, we manually selected the top-most, bottom-most, left-most, and right-most vertices of the drum head in the video and recorded their image coordinates. Following the procedure described in Sec. 4.1.3, we were able to estimate the 3D-to-2D projection matrix from those four points, and then map the remaining mesh vertices to image coordinates.
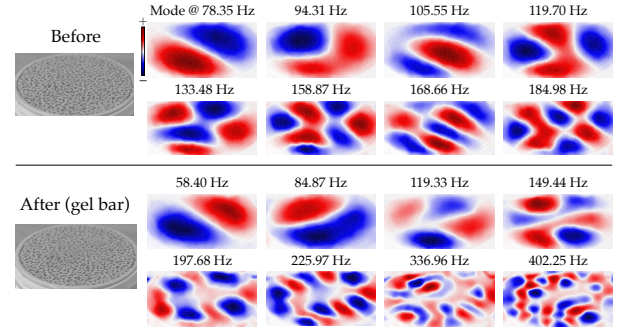

Fig. 14. Extracted image-space modes from real videos of a drum, before and after a defect had been introduced. The defect shown here is a gel rectangle, which was painted on the bottom of the drum head. Differences in the image-space modes pre- and post-defect already appear among the lowest resonant frequencies, but become more noticeable as regions of smaller motion at higher frequencies. Only vertical motion is shown.

6.1.5 Results and discussion. Fig. 15 shows estimated Young's modulus and density for various real drum heads, before and after defects were introduced. It is important to remember that the defects are not visible in the video. The defects tend to appear more clearly in the reconstructed Young's modulus than in density. For both materials (gel and acrylic), the defect appears as a bright region in Young's modulus. Interestingly, gel and acrylic appear in different ways in the density estimations. For gel defects, there is a bright, filled region in the density map that corresponds to a higher mass from the defect. For acrylic defects, this change appears on the edges of the defect. A possible explanation is that the acrylic plastic circles are much stiffer than gel, which bends along with the rubber membrane.
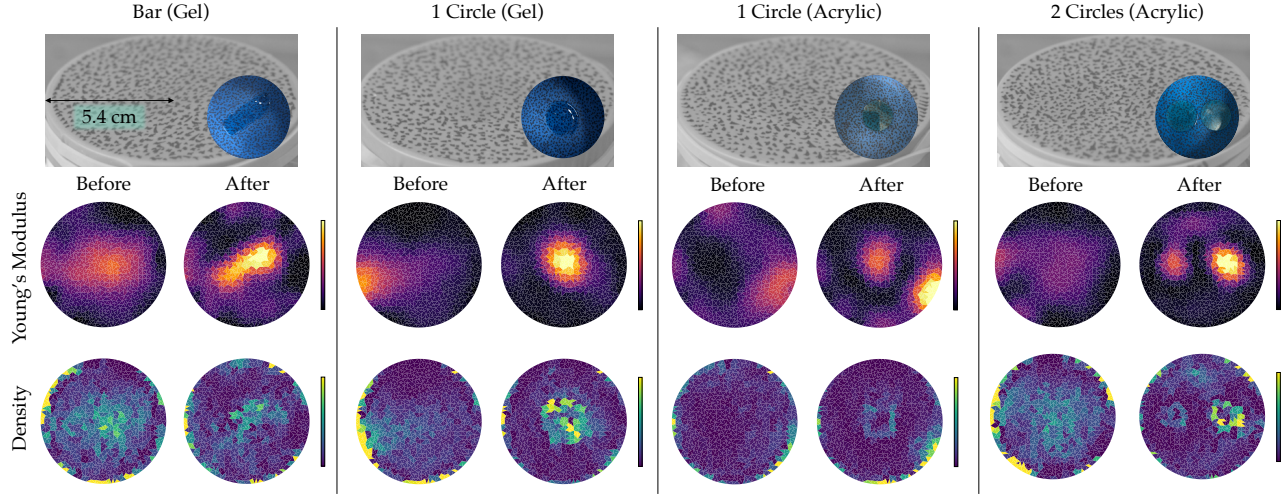
Fig. 15. Reconstruction from real videos of drums. The defects shown here are a gel bar, gel circle, acrylic circle, and two acrylic circles, applied to the underside of the drum head. For each defect, we recorded a video of the drum before and after the defect was applied. One cannot see the defect in a video frame, but after applying our method, we were able to image the defects as changes in Young's modulus and density. For each type of defect, the "Before" and "After" material properties are plotted with the same normalized colormap. Notice that all of the defects appear as sharp rises in Young's modulus.
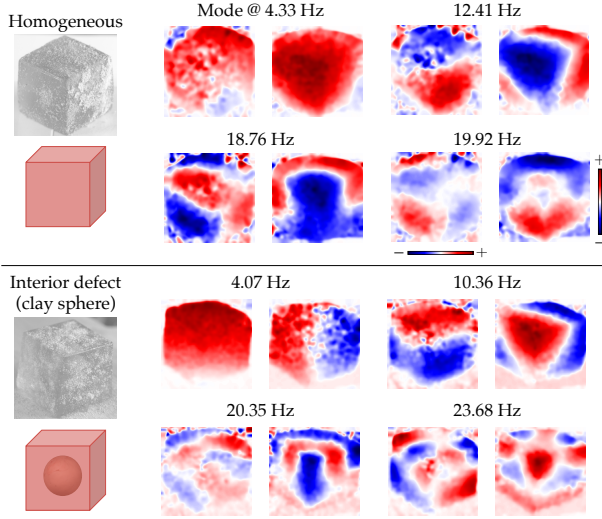

Fig. 16. Extracted image-space modes from real videos of a jello cube, which had a clay sphere placed in its center. The first four out of eight identified image-space modes are shown here. Note that the recovered image-space modes and corresponding resonant frequencies are slightly different between the two cubes, indicating a change of material properties.

6.2 Real Cubes

To further test our method on real data, we recorded videos of real gelatinous cubes. These included homogeneous cubes and cubes with a hidden defect in the center. Various excitation methods were tried, including a hammer hitting the table near the cube to mimic an impulse force. While this method excited the most resonant frequencies, in most videos, we could only identify at most eight

image-space modes. This presents a significant practical challenge. As Fig .5 shows, at least eight image-space modes are required to image an impulse change in material properties within a synthetic cube. Real-world videos contain more noise due to background and camera motion, which means that it is likely that more than eight modes are required to detect an interior defect. Future work can explore how to excite more ideal modal responses in a real 3D object. Fig. 16 shows some of the image-space modes that were extracted from a real video of a gelatinous cube with a clay sphere in its center.

7 DISCUSSION

We have shown that it is possible to recover spatially-varying material properties of 3D objects from monocular video, even in regions unseen in the image. This can be done by decomposing 2D surface motion into image-space modes, and then solving for the Young's modulus and density values that agree with the observed modes. We demonstrated our method on synthetic and real-world data of objects ranging from 2D drum heads to a 3D bunny. Future work to apply this approach in the real world might incorporate a model of damping, explore optimal ways to excite modes in a 3D object, and develop a mesh-fitting method for unknown geometries.

Our results highlight that monocular videos are a simple, yet powerful, source of data for understanding the physical properties of objects around us. Our proposed approach could be useful as an efficient form of non-destructive testing, as it only requires one monocular camera. Of particular interest to the graphics community, future technology based on this idea could facilitate automatic identification of a real-world object's physical properties, which could then be used for physics-based animations. We believe that videos are a promising domain for further research into non-destructive testing and characterizing objects for physics-based graphics, turning everyday visual sensors into tools for material characterization.

ACKNOWLEDGMENTS

REFERENCES

Martin Alnæs, Jan Blechta, Johan Hake, August Johansson, Benjamin Kehlet, Anders Logg, Chris Richardson, Johannes Ring, Marie E Rognes, and Garth N Wells. 2015. The FEniCS project version 1.5. *Archive of Numerical Software* 3, 100 (2015).

Luke B. Beardslee, Marcel C. Remillieux, and T.J. Ulrich. 2021. Determining material properties of components with complex shapes using Resonant Ultrasound Spectroscopy. *Applied Acoustics* 178 (2021), 108014. https://doi.org/10.1016/j.apacoust.2021.108014

Kiran S Bhat, Christopher D Twigg, Jessica K Hodgins, Pradeep Khosla, Zoran Popovic, and Steven M Seitz. 2003. Estimating cloth simulation parameters from video. (2003).

Bernd Bickel, Moritz Bächer, Miguel A Otaduy, Hyunho Richard Lee, Hanspeter Pfister, Markus Gross, and Wojciech Matusik. 2010. Design and fabrication of materials with desired deformation behavior. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 1–10.

Katherine L Bouman, Bei Xiao, Peter Battaglia, and William T Freeman. 2013. Estimating the material properties of fabric from video. In *Proceedings of the IEEE international conference on computer vision*. 1984–1991.

Justin G Chen, Robert W Haupt, and Oral Buyukozturk. 2014a. The acoustic-laser vibrometry technique for the noncontact detection of discontinuities in fiber reinforced polymer-retrofitted concrete. *Materials evaluation* 72, 10 (2014).

Justin G Chen, Neal Wadhwa, Young-Jin Cha, Frédo Durand, William T Freeman, and Oral Buyukozturk. 2014b. Structural modal identification through high speed camera video: Motion magnification. In *Topics in Modal Analysis I, Volume 7*. Springer, 191–197.

Justin G Chen, Neal Wadhwa, Young-Jin Cha, Frédo Durand, William T Freeman, and Oral Buyukozturk. 2015. Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration* 345 (2015), 58–71.

TC Chu, WF Ranson, and Michael A Sutton. 1985. Applications of digital-image-correlation techniques to experimental mechanics. *Experimental mechanics* 25, 3 (1985), 232–244.

Abe Davis, Katherine L. Bouman, Justin G. Chen, Michael Rubinstein, Oral Büyüköztürk, Frédo Durand, and William T. Freeman. 2017. Visual Vibrometry: Estimating Material Properties from Small Motions in Video. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 4 (2017), 732–745. https://doi.org/10.1109/TPAMI.2016.2622271

Abe Davis, Katherine L Bouman, Justin G Chen, Michael Rubinstein, Fredo Durand, and William T Freeman. 2015a. Visual vibrometry: Estimating material properties from small motion in video. In *Proceedings of the ieee conference on computer vision and pattern recognition*. 5335–5343.

Abe Davis, Justin G Chen, and Frédo Durand. 2015b. Image-space modal bases for plausible manipulation of objects in video. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–7.

Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham Mysore, Fredo Durand, and William T. Freeman. 2014. The Visual Microphone: Passive Recovery of Sound from Video. *ACM Transactions on Graphics (Proc. SIGGRAPH)* 33, 4 (2014), 79:1–79:10.

Franz Durst, Adrian Melling, and James H Whitelaw. 1981. *Principles and practice of laser-Doppler anemometry*. Number BOOK. Academic press.

Timothy Emge and Oral Buyukozturk. 2012. Remote nondestructive testing of composite-steel interface by acoustic laser vibrometry. *Materials evaluation* 70, 12 (2012).

João Feiteira, Eleni Tsangouri, Elke Gruyaert, Christine Lors, Ghislain Louis, and Nele De Belie. 2017. Monitoring crack movement in polymer-based self-healing concrete through digital image correlation, acoustic emission analysis and SEM in-situ loading. *Materials & Design* 115 (2017), 238–246.

David J Fleet and Allan D Jepson. 1990. Computation of component image velocity from local phase information. *International journal of computer vision* 5, 1 (1990), 77–104.

Carmelo Gentile and A Saisi. 2007. Ambient vibration testing of historic masonry towers for structural identification and damage assessment. *Construction and building materials* 21, 6 (2007), 1311–1321.

Maria Girardi, Cristina Padovani, Daniele Pellegrini, Margherita Porcelli, and Leonardo Robol. 2020. Finite element model updating for structural applications. *J. Comput. Appl. Math.* 370 (2020), 112675.

NS Ha, HM Vang, and NS Goo. 2015. Modal analysis using digital image correlation technique: an application to artificial wing mimicking beetle's hind wing. *Experimental Mechanics* 55, 5 (2015), 989–998.

Yunus Emre Harmanci, Utku Gülan, Markus Holzner, and Eleni Chatzi. 2019. A Novel Approach for 3D-Structural Identification through Video Recording: Magnified

Tracking. *Sensors* 19, 5 (2019). https://doi.org/10.3390/s19051229

Mark N Helfrick, Christopher Niezrecki, Peter Avitabile, and Timothy Schmidt. 2011. 3D digital image correlation methods for full-field vibration measurement. *Mechanical systems and signal processing* 25, 3 (2011), 917–927.

Jeffrey D Helm. 2008. Digital image correlation for specimens with multiple growing cracks. *Experimental mechanics* 48, 6 (2008), 753–762.

Yun-Xian Ho, Michael S Landy, and Laurence T Maloney. 2006. How direction of illumination affects visually perceived surface roughness. *Journal of vision* 6, 5 (2006), 8–8.

Yixin Hu, Qingnan Zhou, Xifeng Gao, Alec Jacobson, Denis Zorin, and Daniele Panozzo. 2018. Tetrahedral Meshing in the Wild. *ACM Trans. Graph.* 37, 4, Article 60 (July 2018), 14 pages. https://doi.org/10.1145/3197517.3201353

Hossam Khalil, Dongkyu Kim, Joonsik Nam, and Kyihwan Park. 2016. Accuracy and noise analyses of 3D vibration measurements using laser Doppler vibrometer. *Measurement* 94 (2016), 883–892.

Meekyoung Kim, Gerard Pons-Moll, Sergi Pujades, Seungbae Bang, Jinwook Kim, Michael J Black, and Sung-Hee Lee. 2017. Data-driven physics for human soft tissue animation. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–12.

Sandra Landahl and Leon A Terry. 2020. Non-destructive discrimination of avocado fruit ripeness using laser Doppler vibrometry. *Biosystems Engineering* 194 (2020), 251–260.

Ce Liu, Lavanya Sharan, Edward H Adelson, and Ruth Rosenholtz. 2010. Exploring features in a bayesian framework for material recognition. In *2010 ieee computer society conference on computer vision and pattern recognition*. IEEE, 239–246.

Ce Liu, Antonio Torralba, William T Freeman, Frédo Durand, and Edward H Adelson. 2005. Motion magnification. *ACM transactions on graphics (TOG)* 24, 3 (2005), 519–526.

William N MacPherson, Mark Reeves, David P Towers, Andrew J Moore, Julian DC Jones, Martin Dale, and Craig Edwards. 2007. Multipoint laser vibrometer for modal analysis. *Applied optics* 46, 16 (2007), 3126–3132.

Eder Miguel, Derek Bradley, Bernhard Thomaszewski, Bernd Bickel, Wojciech Matusik, Miguel A Otaduy, and Steve Marschner. 2012. Data-driven estimation of cloth simulation models. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 519–528.

Hani H Nassif, Mayrai Gindy, and Joe Davis. 2005. Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *Ndt & E International* 38, 3 (2005), 213–218.

Yuriy M. Poplavko. 2019. Chapter 2 - Mechanical properties of solids. In *Electronic Materials*, Yuriy M. Poplavko (Ed.). Elsevier, 71 – 93. https://doi.org/10.1016/B978-0-12-815780-0.00002-5

Javier Portilla and Eero P Simoncelli. 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision* 40, 1 (2000), 49–70.

Nicolaas Bernardus Roozen, Ludovic Labelle, Monika Rychtáriková, and Christ Glorieux. 2015. Determining radiated sound power of building structures by means of laser Doppler vibrometry. *Journal of Sound and Vibration* 346 (2015), 81–99.

C Santulli and G Jeronimidis. 2006. Development of a method for nondestructive testing of fruits using scanning laser vibrometry (SLV). *NDT. net* 11, 10 (2006), 1–12.

Aral Sarrafi, Zhu Mao, Christopher Niezrecki, and Peyman Poozesh. 2018. Vibration-based damage detection in wind turbine blades using Phase-based Motion Estimation and motion magnification. *Journal of Sound and Vibration* 421 (2018), 300 – 318. https://doi.org/10.1016/j.jsv.2018.01.050

Lavanya Sharan, Yuanzhen Li, Isamu Motoyoshi, Shin'ya Nishida, and Edward H Adelson. 2008. Image statistics for surface reflectance perception. *JOSA A* 25, 4 (2008), 846–865.

Eero P Simoncelli and William T Freeman. 1995. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Proceedings., International Conference on Image Processing*, Vol. 3. IEEE, 444–447.

Eero P Simoncelli, William T Freeman, Edward H Adelson, and David J Heeger. 1992. Shiftable multiscale transforms. *IEEE transactions on Information Theory* 38, 2 (1992), 587–607.

Emanuela Speranzini and Stefano Agnetti. 2014. The technique of digital image correlation to identify defects in glass structures. *Structural Control and Health Monitoring* 21, 6 (2014), 1015–1029.

F Trebuňa and M Hagara. 2014. Experimental modal analysis performed by high-speed digital image correlation system. *Measurement* 50 (2014), 78–85.

Shih-Heng Tung, Ming-Hsiang Shih, and Wen-Pei Sung. 2008. Development of digital image correlation method to analyse crack variations of masonry wall. *Sadhana* 33, 6 (2008), 767–779.

Neal Wadhwa, Justin G Chen, Jonathan B Sellon, Donglai Wei, Michael Rubinstein, Roozbeh Ghaffari, Dennis M Freeman, Oral Büyüköztürk, Pai Wang, Sijie Sun, et al. 2017. Motion microscopy for visualizing and quantifying small motions. *Proceedings of the National Academy of Sciences* 114, 44 (2017), 11639–11644.

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

Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. 2014. Riesz pyramids for fast phase-based video magnification. In *2014 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–10.

Bin Wang, Longhua Wu, KangKang Yin, Uri M Ascher, Libin Liu, and Hui Huang. 2015. Deformation capture and modeling of soft objects. *ACM Trans. Graph.* 34, 4 (2015), 94–1.

Bohan Wang, Yili Zhao, and Jernej Barbič. 2017. Botanical materials based on biomechanics. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–13.

Huamin Wang, Ravi Ramamoorthi, and James F. O'Brien. 2011. Data-Driven Elastic Models for Cloth: Modeling and Measurement. *ACM Transactions on Graphics* 30, 4 (July 2011), 71:1–11. http://graphics.berkeley.edu/papers/Wang-DDE-2011-08/ Proceedings of ACM SIGGRAPH 2011, Vancouver, BC Canada.

Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)* 31, 4 (2012), 1–8.

ZhiMin Wu, Hua Rong, JianJun Zheng, Feng Xu, and Wei Dong. 2011. An experimental investigation on the FPZ properties in concrete using digital image correlation technique. *Engineering Fracture Mechanics* 78, 17 (2011), 2978–2990.

Hongyi Xu, Yijing Li, Yong Chen, and Jernej Barbič. 2015. Interactive material design using model reduction. *ACM Transactions on Graphics (TOG)* 34, 2 (2015), 1–14.

Yongchao Yang, Charles Dorn, Charles Farrar, and David Mascareñas. 2020. Blind, simultaneous identification of full-field vibration modes and large rigid-body motion of output-only structures from digital video measurements. *Engineering Structures* 207 (2020), 110183. https://doi.org/10.1016/j.engstruct.2020.110183

M Zorović and A Čokl. 2015. Laser vibrometry as a diagnostic tool for detecting wood-boring beetle larvae. *Journal of pest science* 88, 1 (2015), 107–112.