IEEE TRANSACTIONS ON
PATTERN ANALYSIS AND
MACHINE INTELLIGENCE
A publication of the IEEE Computer Society

Visual Vibrometry: Estimating Material Properties from Small Motions in Video

Abe Davis *[1], Katherine L. Bouman *[1], Justin G. Chen [1], Michael Rubinstein [2†], Oral Büyüköztürk [1], Frédo Durand [1], and William T. Freeman [1]

Abstract—The estimation of material properties is important for scene understanding, with many applications in vision, robotics, and structural engineering. This paper connects fundamentals of vibration mechanics with computer vision techniques in order to infer material properties from small, often imperceptible motion in video. Objects tend to vibrate in a set of preferred modes. The frequencies of these modes depend on the structure and material properties of an object. We show that by extracting these frequencies from video of a vibrating object, we can often make inferences about that object's material properties. We demonstrate our approach by estimating material properties for a variety of objects by observing their motion in high-speed and regular frame rate video.

Index Terms—Material Properties, Vibration, Small Motion, Computational Photography, Computational Imaging

1 Introduction

Understanding a scene involves more than just recognizing object categories or 3D shape. Material properties like density, stiffness, and damping can play an important role in applications that involve assessing or interacting with the world. In the field of non-destructive testing (NDT), these properties are often recovered by analyzing the vibrations of an object. Typically, these vibrations are measured with contact sensors or expensive laser vibrometers, which limit sampling to only a small number of discrete points on an object's surface. We propose an alternative approach to vibration analysis that instead uses cameras to measure vibrations and make inferences about the object's underlying physical properties.

Objects tend to vibrate in a set of preferred modes. These vibrations occur in most materials, but often happen at scales and frequencies outside the range of human visual perception. Bells, for instance, vibrate at distinct audible frequencies when struck. We cannot usually see these vibrations because their amplitudes are too small and their frequencies are too high - but we hear them. Intuitively we know that large bells tend to sound deeper than small ones, or that a bell made of wood will sound muted compared to one made of silver. This is because an object's modes of vibration are closely and predictably related to its material properties. We show how this connection can be used to learn about the material properties of an object by analyzing its vibrations in video.

In this paper we review established theory on modal vibrations, and connect this theory to proposed features that can be extracted from video. These motion spectra features provide an ambiguous combination of structural and material information that can be used directly to make relative measurements, or in combination with structural information to make absolute measurements. We present three experiments showing how these features can be used to estimate structural or material properties given some prior information about an object. The first experiment, using a set of clamped rods, is designed to resemble typical engineering applications, and shows how our features can be used to resolve material properties in situations where geometry can be precisely measured. The second experiment, using a set of hanging fabrics, explores the idea of learning the relationship between our features and material properties when objects naturally occur with similar geometry, demonstrating the potential for data-driven approaches to material estimation. These first two experiments are summarized in Figure 1. The third experiment uses a set of wine glasses to demonstrate how our technique can be used to estimate *relative* properties, even without a prior on geometry, by comparing the resonance of objects within a group, or the resonance of a single object over time (refer to Figure 14).

Our previous work [13] introduced the idea of using vibrations extracted from video to estimate material properties of objects. This paper expands on that approach with additional results and analysis. In our rod experiments we perform new analysis to evaluate damping, and show how geometry can be inferred from vibrations when material properties are known. In our study of fabrics, we include additional training/testing scenarios to further demonstrate the invariance of our features to changes in viewpoint and excitation force. We also present new experiments: one using wine glasses to explore how our proposed features can be used in contexts where information about geometry is not available, and an experiment comparing our method to sensors traditionally used for vibration analysis (a laser Doppler vibrometer, and an accelerometer). Additional information and data can be found on our project website (visualvibrometry.com).

2 Related Work

This paper connects related works in computer vision, graphics, and civil engineering through common theory and uses these connections to extend existing methods.

2.1 Traditional Vibration Analysis

Vibration analysis is an established tool used in a variety of engineering disciplines. Especially related to this paper is work in the

* *Joint first author*
[1] *Massachusetts Institute of Technology,*
{abedavis,klbouman,ju21743,obuyuk,fredo,billf}@mit.edu
[2] *Google Research, mrub@google.com*
† *Part of this work was done while the author was at Microsoft Research*

Digital Object Identifier no. 10.1109/TPAMI.2016.2622271

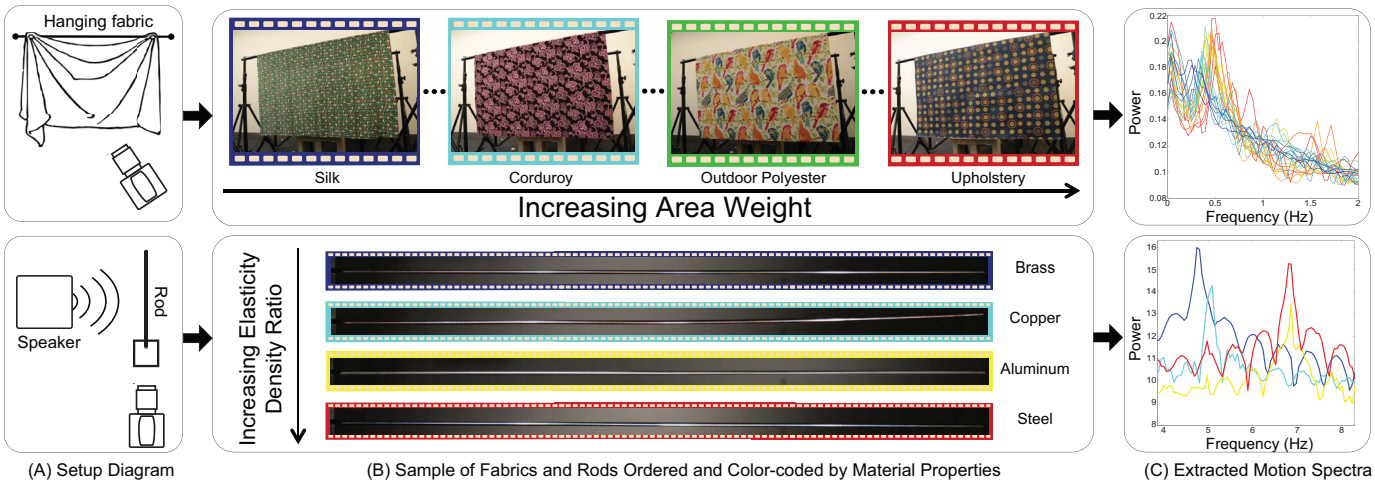


Fig. 1. We present a method for estimating material properties of an object by examining small motions in video. (A) We record video of different fabrics and clamped rods exposed to small forces such as sound or natural air currents in a room. (B) We show fabrics (top) color-coded and ordered by area weight, and rods (bottom) similarly ordered by their ratio of elastic modulus to density. (C) Local motion signals are extracted from captured videos and used to compute a temporal power spectrum for each object. These motion spectra contain information that is predictive of each object's material properties. For instance, observe the trends in the spectra for fabrics and rods as they increase in area weight and elasticity/density, respectively (blue to red). By examining these spectra, we can make inferences about the material properties of objects.

field of NDT, where techniques based on ultrasound are common. However, these techniques often require direct contact with the object being measured [31]. Non-contact vibration measurement is usually accomplished with a laser Doppler vibrometer, which computes the velocity of a surface by measuring the Doppler shift of a reflected laser beam [16]. Laser vibrometers have been used to non-destructively examine valuable paintings [7], [11], detect land mines [1], [20], test fruit [28], find defects in composite materials [6], [8], [17], and even test vibration modes of small structures [33]. However, laser vibrometers are active in nature and generally only measure the vibration of a single surface point. While scanning or multi-beam laser vibrometers exist [1], [33], they are still active and can be prohibitively expensive - costing several times more than even the most expensive high-speed camera used in this work.

2.2 Material Property Estimation from Video

Previous work in computer vision has focused on estimating material properties from static images [19], [21], [25], [30]. In contrast, our goal is to use video in order to estimate material properties that characterize the motion of an object.

A number of works in vision and graphics have been used to estimate properties of fabric, which we also do in this paper. Early approaches worked by fitting the parameters of cloth-specific models to video and depth information [4], [22]. Bouman et al. [5] adopted a learning approach to estimate material properties from a video of fabric moving under wind forces. As with our experiments in Section 7, [5] estimated material properties directly from video statistics using a regression strategy. That work found the local autocorrelation of optical flow to be especially predictive of a fabric's area weight and stiffness, suggesting a possible connection between material properties and the spectrum of an object's motion in video. Our work uses established vibration theory to explain this connection and improve on the features used in their paper.

2.3 Small motions

Our approach to material property estimation is based on linear approximations of object deformation that hold when displacement is small. We build on several recent works in vision and graphics that address small motions in video [27], [34], [36], [38]. As with many of these works, our method uses spatial phase variations of the complex steerable pyramid [26], [32] to represent small local motions in video. In recent work, Chen et al. [9], [10] use these phase variations to quantify the vibration modes of pipes and cantilever beams. Our features and analysis also bear some resemblance to the work of Davis et al. [15], but where that work focuses on using vibrations in video to recover sound, we use them to learn about the physical properties of visible objects.

Our work is also closely related to [14], which uses projected vibration modes recovered from video to create plausible image space simulations of objects. However, where that work focuses on plausibility over accuracy (e.g. for entertainment, film effects), ours focuses on accurate estimation of properties for scientific and engineering applications.

3 Theory of Vibration

The object motion we consider in this paper is small by computer vision standards. While this sometimes makes the motion difficult to extract, it makes it simpler to analyze. General deformations of an object may be governed by complex nonlinear relationships, but small deformations from a rest state are often well-approximated by linear systems. The theory of such linear systems is well established, and used in work spanning a variety of disciplines. We review eigenmode analysis, a subset of this theory that is especially relevant to our work. In Section 4 we connect this analysis to the features we extract from video, and use it to motivate our approach for material property estimation. The goal of this section is to provide intuition; for detailed derivations we

recommend [29]. A closely related analysis with greater focus on vibration mode shapes can also be found in [14].

3.1 Eigenmode Analysis

In modal analysis, a solid object is modeled as a system of point masses connected by springs and dampers [29]. Intuitively, rigid objects are approximated with stiff springs, highly damped objects are approximated with stiff dampers, and dense objects are approximated with heavy masses. Consider the mass matrix $\mathbf{M}$ of inertias between points, the matrix $\mathbf{C}$ of viscous damping values between points, and the matrix $\mathbf{K}$ of spring stiffnesses between points. The differential equation of motion for this system is given by:

$$\mathbf{M}\ddot{x} + \mathbf{C}\dot{x} + \mathbf{K}x = 0, \tag{1}$$

where x, $\dot{x}$, and $\ddot{x}$ are vectors describing the displacement, velocity, and acceleration of the points, respectively. Under the common assumption of Rayleigh damping, the matrix $\mathbf{C}$ is a linear combination of $\mathbf{M}$ and $\mathbf{K}$ given by $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$. In this case, the eigenmodes of the system are the solutions to the generalized eigenvalue problem given by $\mathbf{K}\phi_i = \omega_i^2\mathbf{M}\phi_i$. The eigenmodes $\phi_1 ... \phi_N$ define a modal matrix $\mathbf{\Phi}$ that diagonalizes the mass and stiffness matrices into modal masses $\mathbf{m}_i$, and stiffnesses $\mathbf{k}_i$:

$$\mathbf{\Phi} = \begin{bmatrix}\phi_1 & \phi_2 & \dots & \phi_N\end{bmatrix} \tag{2}$$

$$\mathbf{\Phi}^T\mathbf{M}\mathbf{\Phi} = diag(\mathbf{m}_i) \tag{3}$$

$$\mathbf{\Phi}^T\mathbf{K}\mathbf{\Phi} = diag(\mathbf{k}_i) \tag{4}$$

The matrix $\mathbf{\Phi}$ defines modal coordinates $\mathbf{q}(t)$, where $\mathbf{x}(t) = \mathbf{\Phi}\mathbf{q}(t)$, which decouple the system into single degree of freedom systems defined by modal masses $\mathbf{m}_i$, stiffnesses $\mathbf{k}_i$, and dampings $\mathbf{c}_i = \alpha\mathbf{m}_i + \beta\mathbf{k}_i$. Defining the undamped natural frequency of a mode as $\omega_i = \sqrt{\frac{\mathbf{k}_i}{\mathbf{m}_i}}$, we get the decoupled equation of motion for each mode:

$$\ddot{\mathbf{q}}(t) + 2\xi_i\omega_i\dot{\mathbf{q}}(t) + \omega_i^2\mathbf{q}(t) = 0 \tag{5}$$

where ξ_i is a modal damping ratio, defined as:

$$\xi_i = \frac{\mathbf{c}_i}{2\mathbf{m}_i\omega_i} = \frac{1}{2}\Big(\frac{\alpha}{\omega_i} + \beta\omega_i\Big). \tag{6}$$

3.2 Transfer Functions of Modal Systems

The impulse response of the system can then be decoupled into a superposition of the impulse responses for individual modes. We obtain the unit impulse response for the i^{th} mode by solving Equation 5

$$h_i(t) = \Big(\frac{e^{-\xi_i\omega_i t}}{\mathbf{m}_i\omega_{di}}\Big)\sin(\omega_{di}t) \tag{7}$$

where the damped natural frequency is $\omega_{di} = \omega_i\sqrt{1-\xi_i^2}$. The Fourier transform of the unit impulse response, $h_i(t)$ in Equation 7, results in the convolution

$$H_i(\omega) = \Big(\frac{1}{\mathbf{m}_i\omega_{di}}\frac{\xi_i\omega_i}{\xi_i^2\omega_i^2 + \omega^2}\Big) * \Big(\frac{\delta(\omega-\omega_{di}) - \delta(\omega+\omega_{di})}{i}\Big). \tag{8}$$

The transfer function of a single mode is thus the convolution of a spike at its resonant frequency and a Lorentzian distribution (the Fourier transform of the decaying exponential), which has a width that depends on modal frequency and damping.

3.3 Resonance and Material Properties

Both mode shapes and frequencies will depend on an object's geometry. If a piece of an object is removed, for instance, it will change the sparsity of both $\mathbf{M}$ and $\mathbf{K}$, potentially changing both eigenmodes and eigenvalues of the system. But if geometry is held constant and only material properties are changed – say by making an object uniformly heavier or stiffer – this can only scale $\mathbf{M}$ and $\mathbf{K}$, scaling the eigenvalues of the system but leaving the eigenmodes unchanged. This implies that *different objects with the same geometry have the same set of mode shapes, but their resonant frequencies scale in proportion to material properties*. In our experiments we use this property to estimate material properties among objects with common geometry by observing their modal vibrations. We also leverage the fact that resonant frequencies ω_i are global properties of an object - meaning they do not vary across the object's surface; only the amplitudes and local phases of vibration vary spatially, according to the mode shapes, ϕ_i. This suggests that we can learn about global properties of an object by observing vibrations at any part of the object, even if these vibrations are affected by hidden or occluded structure.

4 Extracting Motion Features

We use small local motions in video to reason about the modes of recorded objects. For each spatial point in a video, we compute the local motion around that point over time. Our analysis relates the spectra of these motion signals to mode shapes, ϕ_i, and frequencies, ω_i.

4.1 Local Motion Signals

Local motion signals are derived from phase variations of the complex steerable pyramid (CSP) [18], [26], [32]. A CSP is computed for each input frame of video, breaking it into complex-valued sub-bands for different scales and orientations. Previous work has shown that temporal variations in the spatial phases of these sub-bands are an effective measure of small motions in video [15], [34]. For our local motion signals, we weigh the phase variation of each sub-band by the squared amplitude of that sub-band, as in [15], and spatially filter these weighted signals with a Gaussian kernel to account for noise in texture-less regions of video. We chose this representation of motion for its simplicity and robustness, but alternatives such as optical flow could be equally valid.

4.2 Motion Spectra

Recall that the frequencies ω_i do not vary across an object's surface. This implies that the power spectra of local motions across an object should have spikes at the same resonant frequencies. Thus, we compute the global motion power spectrum for a video by averaging the power spectra of local motions extracted at every pixel location, scale, and orientation. This allows us to reduce noise, and leaves us with a single temporal power spectrum describing the frequencies of motion that exist in a video.

4.3 Viewpoint Invariance

An advantage of using temporal spectra as features is that they offer invariance to changes in scale and viewpoint. This invariance agrees with what we know from theory: the resonant frequencies of an object are global to that object and should not differ

according to how it is viewed. In Section 7 we use this to estimate the material properties of fabrics in experiments where training and testing data sets are taken from different viewpoints and cameras.

4.4 Damping

Under a broad-spectrum excitation force, the recovered temporal motion spectra should take the shape of an object's transfer function. In Section 3 we showed that damping determines the width of resonant spikes in this transfer function. Therefore, by observing the width of resonant spikes in recovered motion spectra we can reason about damping in an object. The effect of damping can be learned by observing many motion spectra, or it can be estimated explicitly by fitting Lorentzian distributions to spikes in these spectra (see Section 6.2).

4.5 Mode Shapes

The theoretical mode shapes ϕ_i describe spatially varying amplitudes of a vibration mode across the surface of an object. Positive and negative amplitudes vibrate with opposite phase, and zeros indicate static nodal points in a vibration mode. Therefore, by visualizing the phases and amplitudes of our local motion spectra at a resonant frequency, we can picture the shape of the corresponding mode. While we do not use these shapes to estimate material properties, visualizing them helps to verify the presence of a vibration mode at a specific frequency, since spatial coherancy is not likely to happen at random. We adopt the visualization used in [15], where the image of local Fourier coefficients at a given frequency is displayed by mapping phase to hue and magnitude to brightness (see Figure 13).

5 Method

Our task is to estimate the material properties of objects using the motion spectra described in Section 4. Our method has three components that vary depending on the object being observed.

5.1 Excitation

An object must move in order for us to observe its vibration modes and resonant frequencies. Some very deformable objects, such as hanging fabric, may move enough with natural air currents for no additional forces to be necessary. For more rigid objects, like wine glasses or metal rods, we use sound to induce motion. The excitation should be strong enough to create a recoverable motion signal, and should contain energy at each of the objects resonant frequencies. Sound excitation has been used for this purpose previously in NDT [6], [8], [11], [17], [20].

5.2 Video Capture

To estimate an object's resonant frequencies we need to record at a high enough frame rate to place these frequencies under the Nyquist limit. We should also ensure that videos capture enough periods at each modal frequency to sufficiently localize corresponding spikes in the Fourier domain. For objects with high resonant frequencies this can be accomplished with short clips of high-speed video. Objects with low resonant frequencies (like hanging fabric) can be captured with longer, lower-frame rate video.

5.3 Inference

The motion spectrum of an object provides us with an ambiguous combination of structural and material information. In some cases, this combination is directly useful (e.g. tuning an instrument or identifying a source of unwanted noise). In others, it provides constraints from which we can infer more specific properties. This inference depends on the type of information available and the properties being inferred. We explore three approaches to estimating physical properties in this paper – each addressing a different scenario, with different information available about the object being filmed. The first method uses measured or known geometry to directly estimate material properties. This method can be very precise, but requires additional measurement (usually through some means other than video). The second method alleviates the need for careful measurement by learning the relationship between recovered motion spectra and material properties from training data. This approach is convenient, but depends on the availability and accuracy of a learned prior. Finally, the third method is to sidestep the need for any prior on geometry by simply comparing spectra in order to detect changes over time, or variations within a group of objects. This approach is simple, and promising for applications in structural health monitoring, where any significant change in resonance may indicate a problem, and reference spectra are often available.

6 Estimating Properties of Materials with Known Geometry: Rods

In our first set of experiments we estimate the material properties or geometry of various rods by extracting their resonant frequencies from video. The simple geometry of a clamped rod makes it easy to solve for vibration modes analytically as a function of length, diameter, density, and an elastic modulus. While length, diameter, and density can all be measured with a simple ruler and scale, the elastic modulus is usually measured with a tensile test, which requires expensive equipment and usually damages the object being tested. In these experiments we first show how this elastic modulus can instead be measured with a speaker and high-speed camera. Just as our recovered spectra can be used to resolve unknown material properties (i.e. elasticity) given known geometry, we also show that they can be used to resolve unknown geometry given known material properties. This second case could be used to resolve an ambiguity of scale when a filmed object is made of a known material.

Setup: We filmed rods made from four different metals - steel, aluminum, copper, and brass. Rods were clamped to a block of concrete next to a loudspeaker (see Figure 2), and each rod was tested twice: once clamped to a length of 15 inches and once clamped to a length of 22 inches. In Section 6.3 we compare material properties derived from our observations to estimates provided by the manufacturer. Recovered frequencies and mode shapes for all of the rods, as well as birch and fiberglass rods with unreported material properties, can be found in the provided supplemental material.

Excitation: The excitation signal should be broad spectrum to ensure that multiple rod modes are activated. In [9], [10] this is accomplished by striking the beam with a hammer. To avoid damage to the rod, we instead use sound - specifically, a linear ramp of frequencies from 15 Hz to 2250 Hz played through the

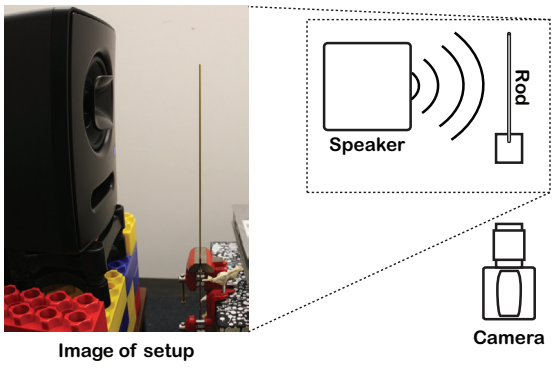


Fig. 2. Rods were clamped to a concrete block next to a loudspeaker (shown left) and filmed with a high-speed camera. By analyzing small motions in the recorded video, we are able to extract the locations of the rod's resonant frequencies and use these values to estimate the rod's material properties.

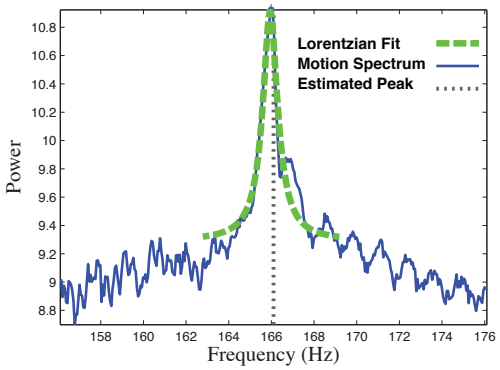


Fig. 4. Our damping selection interface, inspired by the standard procedure defined in [2], presents users with a view of the recovered motion spectra around a predicted rod resonant frequency and asks them to click and drag over the spike region. A Lorentzian is then fit to the selected region and presented to the user for evaluation of accuracy.

loudspeaker at each rod. We found that modes at frequencies below 15 Hz were still activated by this signal, possibly due to the presence of some signal components below 15 Hz and the relatively high sensitivity of lower modes.

Video Capture: Rods were filmed with a Phantom v10 high-speed camera (80×2016 pixel resolution). Given the lengths and thicknesses of our rods, a conservative estimate of material properties put the fourth mode of each rod well below 1250 Hz. We filmed at 2500 fps to ensure a sampling rate high enough to recover this mode for each rod.

6.1 Finding Resonant Frequencies

The vibrations of clamped rods are well studied [29]. A rod's fundamental frequency f_1 (corresponding to its first mode) is related to material properties by the equation:

$$f_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}} \tag{9}$$

where d is the diameter of the rod, L is its length, ρ is its density, E is its Young's modulus (measuring elasticity), and $\omega = 2\pi f$. Given the length and width of a rod, the task of estimating $\sqrt{E/\rho}$ can then be reduced to finding its fundamental frequency. Under ideal conditions this would amount to finding the largest spike in the rod's motion spectrum. However, real spectra tend to also contain spikes at non-modal frequencies (see Figure 3). To distinguish these from the rod's resonant frequencies we recall from Section 3 that changes in material properties only scale the modal frequencies - leaving their ratios constant. In clamped rods, ratios for the first four resonant frequencies can be found analytically[1], and are given by:

$$f_i = \eta_i f_1,$$
$$\eta_1 = 1, \quad \eta_2 = 6.27, \quad \eta_3 = 17.55, \quad \eta_4 = 34.39 \tag{10}$$

where again f_i is the resonant frequency for the ith mode. To distinguish modal frequencies from other spikes, we look for energy in the recovered spectra that occurs in the ratios given by Equation 10. We assume that the probability of a rod mode at a given frequency is proportional to the power at that frequency. Given the recovered spectrum S, we then have:

$$P\,(f = f_1 |\, S) \propto \prod_{i=1}^{4} S(\eta_i f). \tag{11}$$

Using Equation 11, we can find the most likely fundamental frequency using a simple voting scheme. In practice, since we operate in the discrete Fourier domain, we achieve higher precision at the fundamental by using the relations of Equation 10 to vote for the fourth resonant frequency.

6.2 Estimating Damping

As discussed in Section 3, the damping of a mode appears in an object's transfer function as convolution with a Lorentzian distribution that depends on the damping ratio ξ. To find ξ, we fit a Lorentzian distribution around the modes identified by our voting scheme. Automatically fitting these distributions using a fixed range of frequencies around identified modes produces poor fits, as different damping values affect different ranges of frequencies. We address this using a manual selection strategy, inspired by the procedures set by the ASTM for measuring the material damping or loss factor in materials [2]. Our selection interface is similar to the one used for mode selection in [14], where users are presented with the motion spectrum of a video and asked to click on peaks. However, our selection process uses the frequencies predicted with our voting scheme as an initial estimate, zooming in on each predicted frequency one at a time. Users are then asked to select the range of frequencies between the resonant peak and noise floor using their mouse. A Lorentzian is immediately fit to the selected region using non-linear least squares, and presented for the user to evaluate (Figure 4). If the fit looks good, the user proceeds to the next mode. If the fit does not look good, they can press a button to indicate that the damping on the corresponding mode cannot be accurately estimated, a result often caused by mode masking. The full width at half maximum, Δf, of the Lorentzian can then be used to calculate the modal damping ratio as $\xi_i = \frac{\Delta f}{2 f_{di}}$ [2]. Note that we measure the damped frequency, f_{di}, directly from our power spectrum. In Section 3.2 $\omega_{di} = 2\pi f_{di}$ is discussed in more detail.

1. By solving the continuous analog to Equation 1 [29]

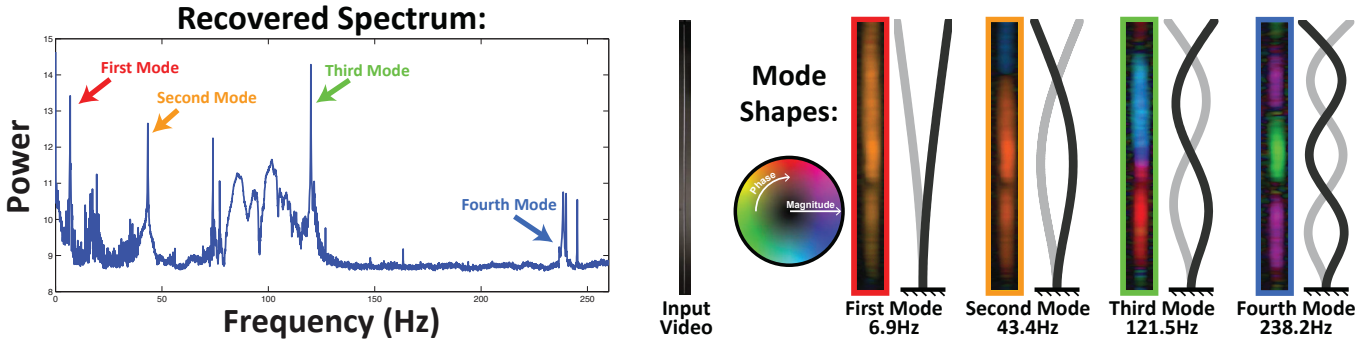
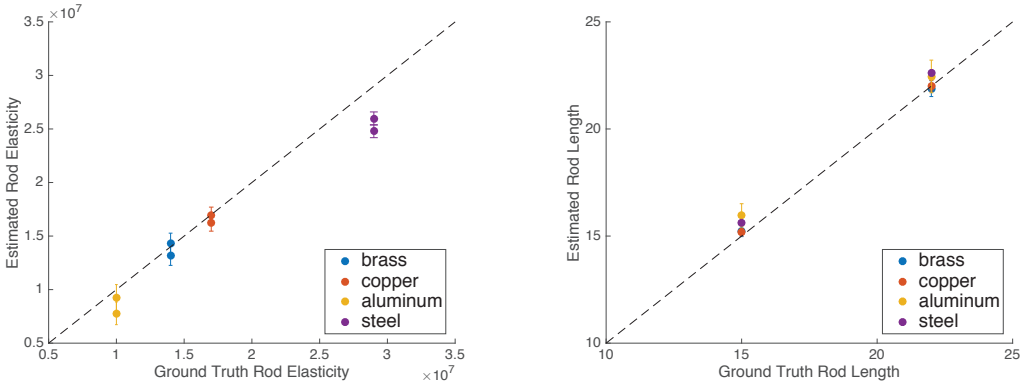


Fig. 3. Finding vibration modes of a clamped brass rod: (Left) We recover a motion spectrum from 2500 Hz video of a 22 inch clamped aluminum rod. Resonant frequencies are labeled. To distinguish resonant frequencies from other spikes in the spectrum, we look for energy at frequencies with ratios derived from the known geometry of the rod. (Middle) A sample frame from the 80×2016 pixel input video. (Right) Visualizations of the first four recovered mode shapes are shown next to the corresponding shapes predicted by theory.



(a) Elastisity Estimation (force per squared inch)

(b) Length Estimation (inches)

Fig. 5. Estimating the elastic modulus and length of clamped rods: (a) Young's moduli (force per squared inch) reported by the manufacturer plotted against values estimated using our technique. Estimated values are close to those reported by the manufacturer, with the largest discrepancies happening in 15 inch rods made of aluminum and steel. (b) The length (inches) of each rod measured to the base of the clamp plotted against values estimated using our technique.

6.3 Results

Young's Modulus / Elasticity: Under fixed but unknown geometry, the recovered locations of fundamental frequencies provide a value proportional to $\sqrt{E/\rho}$. From this we can use Equation 9 with lengths and densities measured by a scale and measuring tape to compute the modulus of each rod. Figure 5a shows a plot of Young's moduli (in force per squared inch) reported by the manufacturer against the values estimated using our technique. Percent errors are given in Table 1.

Length: By rearranging Equation 9, we see that the length of a rod can be estimated as a function of the fundamental frequency, rod diameter, elasticity, and density:

$$L = \sqrt{0.1399 \frac{d\eta_i}{f_i} \sqrt{\frac{E}{\rho}}}. \quad (12)$$

We can use this equation to estimate a rod's length given the measured resonant frequencies and the Young's modulus reported by the manufacturer. Figure 5b shows a plot of the measured length (in inches) of each rod verses the value estimated in this manor. Percent errors are given in Table 2.

Error: Error bars in Figure 5 are calculated for each Young's modulus and length estimate by propagating error bounds for each measured variable. Error propagation was done assuming independent variables [24]. Given a function $F(a, b, c, ...)$, the equation for the error σ_F depending on the errors $\sigma_a, \sigma_b, \sigma_c ...$ is given as:

$$\sigma_F = \sqrt{\left(\frac{\partial F}{\partial a}\right)^2 \sigma_a^2 + \left(\frac{\partial F}{\partial b}\right)^2 \sigma_b^2 + \left(\frac{\partial F}{\partial c}\right)^2 \sigma_c^2 + ...} \quad (13)$$

Young's modulus estimates were calculated by propagating error from length, diameter, and density. Length estimates were calculated by propagating error from only diameter and density. The calculated errors for length estimation are smaller than expected due to the lack of reported tolerances on Young's modulus values. Refer to the Appendix for further information on error approximation.

Mode Shapes: For each rod, we can further verify recovered modes by visualizing the recovered shapes corresponding to estimated resonant frequencies (see Figure 3). Mode shapes are sometimes masked by vibrations from other parts of the experimental setup - for instance, vibrations of the camera or

% Error	Brass	Copper	Aluminum	Steel
22 inches	2.13	-0.40	-7.82	-10.40
15 inches	-5.98	-4.69	-22.13	-14.53

TABLE 1
Percent errors in estimating the Young's modulus (force per squared inch) for each rod.

% Error	Brass	Copper	Aluminum	Steel
22 inches	-0.52	0.10	2.06	2.78
15 inches	1.55	1.21	6.45	4.00

TABLE 2
Percent errors in estimating the length (inches) for each rod.

the frequency of lights powered by AC current. However, it is unlikely that a majority of resonant frequencies will be masked in any single rod. In practice we see the predicted shapes of multiple modes in the data recovered for each rod. All 48 mode shapes recovered in our experiments can be found in the provided supplemental material.

Damping: Material damping properties are not as well characterized as other mechanical properties, such as Young's modulus for stiffness. This is, in part, because it is very difficult to control for external sources of damping. Additionally, damping can vary across the different modes of a given system. As a result, manufacturers do not typically report damping ratios. However, some general trends are accepted for different materials. For example, metals tend to have very low material damping compared to rubber. In addition to our metal rods, for which the manufacturer reported Young's moduli, we also tested a rod made of wood (birch). While material property values for wood are highly variable (likely the reason no Young's modulus was provided), wood is generally accepted to have higher damping than most metals, and quantitative studies of different vibrating systems (e.g. [12]) have supported this claim. Figure 6 shows our damping estimates of different rod modes as a function of frequency (damping was evaluated at each unmasked rod mode). As expected, we see that the wooden rod has the highest damping ratio at every mode.

Discussion: Our estimated moduli are close to, but consistently under, the reported values (Figure 5a and Table 1). One possible explanation for this is an incorrect estimate of where the clamp grabbed each rod in our setup. Similarly, Figure 5b and Table 2 show that our length estimates are close to, and correlated with, but consistently longer than our measured values - which could be explained by the same source of measurement error.

Our damping results show that our wooden rod has consistently higher damping than the metal rods, which is expected given their material differences. However, the relative damping ratios of our metal rods are less consistent across different modes. These results suggest that we are able to distinguish between materials with significantly different levels of damping (such as metal and wood), though additional experiments would be needed to better understand how well we distinguish damping between more similar materials (e.g. among the different metals).

Our Young's modulus and length results suggest both a strength and weakness of an approach that pairs recovered motion spectra with careful measurement for inference - high precision that is very sensitive to accurate modeling of the structure being tested. Our next experiments address this issue by instead attempting to learn the relationship between material properties and resonant frequencies.

7 Learning Properties of Materials with Unknown Geometry: Fabrics

The inference described in Section 6.1 relies on knowing the ratios between resonant frequencies, η_i. These ratios are simple to derive in clamped rods, but can be prohibitively difficult to compute in more general structures. As a result, many applications of vibrometry are limited to simple geometries that can be precisely measured (as is the case with rods) or man-made structures (airplanes, buildings, cars, etc) with resonant frequencies that can be derived from detailed CAD models through FEM analysis. The ubiquity and passive nature of video offers the potential to address this limitation by providing sufficient data to learn relationships between motion spectra and the material properties of objects. In this section, we explore that potential by using a learning approach to estimate the material properties of hanging fabrics from video. We show that our technique outperforms a previous video-based fabric property estimation method, even when trained using data captured from different viewpoints or different excitation forces.

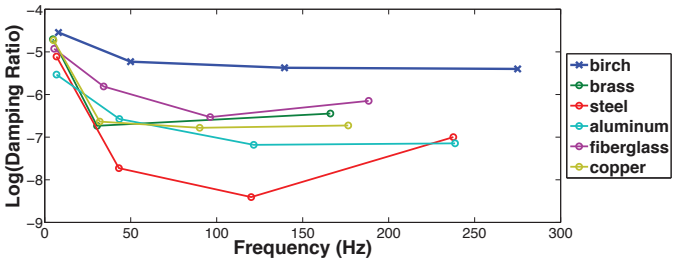


Fig. 6. The damping ratio estimated from the recovered motion spectra for each automatically identified resonant frequency. While reported damping ratios for different materials vary greatly, general trends are recognized. Our recovered rod damping ratios show recognized trends of higher damping in wood than in metals [12], and higher damping in lower fundamental modes due to their high amplitude [3].

A number of metrics exist to describe the material properties of fabrics. These properties can be measured using setups such as the Kawabata system [23], [37]. In the work of Bouman, et al. [5], a dataset of 30 fabrics along with ground truth measurements of stiffness and area weight were collected. We extend this dataset to predict the material properties from videos exhibiting small motions that are often invisible to the naked eye, in contrast to [5] that relied on much larger motions produced by fans.

Setup: Each fabric specimen from [5] (width of approximately 43.5 to 44.5 inches across) was loosely draped over a bar and hung a length of approximately 29.25 to 32.25 inches from the top of the bar. Notice that although the geometry was kept relatively constant, these measurements vary a great deal compared to those used in Section 6.

Excitation: We explore two different types of excitation forces in estimating the material properties of fabric.

Ambient Forces: Even without an explicit excitation force applied, hanging fabric is almost always moving. Ambient forces, such as air currents in the room or small vibrations in the building induce small motions in fabric. Figure 8a shows a space-time slice of a fabric moving due to ambient forces in the room.

Sound: As an alternative, we also tested sound as a source of excitation. Sound was used to provide a small, controlled "kick"

(a) Setup Diagram (b) Setup Image (c) Point Grey Camera (grayscale) (d) Canon 6D SLR Camera

Fig. 7. Videos were recorded of the fabric moving from (c) a grayscale Point Grey camera (800×600 pixel resolution) at 60 fps and (d) an RGB SLR Camera (Canon 6D, 1920×1080 pixel resolution) at 30 fps. The experimental layout (a,b) consisted of the two cameras observing the fabric from different points of view.

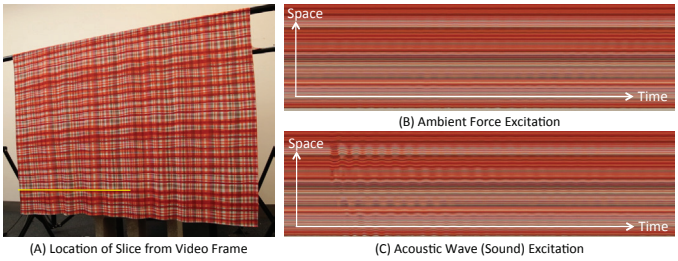


Fig. 8. Videos of fabric excited by two different types of force were recorded. Here we see space × time slices from minute long videos of a fabric responding ambient forces (b) and sound (c). The motion is especially subtle in (b), but still encodes predictive information about the fabric's material properties.

to the hanging fabric. We excited each fabric with a one second, logarithmic frequency ramp from 15 to 100 Hz. Figure 8b shows a space-time slice of a fabric moving due to this "kick."

Video Capture: Each combination of fabric and excitation force was captured simultaneously by two cameras: an RGB SLR camera (Canon 6D, 1920×1080 pixel resolution) at 30 fps, and a grayscale Point Grey camera (800×600 pixel resolution) at 60 fps. The cameras recorded different viewpoints (see Figure 7), which we use to test the invariance of our trained models to changes in perspective. Each video is approximately one-minute long and can be found, along with the corresponding fabric measurements (width and height), on our project website.

7.1 Property Estimation

Feature Extraction: Due to their comparatively high damping, fabric motion spectra do not contain the same clean, narrow peaks seen in rods. Damping causes the bandwidth around resonant frequencies to overlap, making it difficult to identify individual modes (see Figure 1). As a result, the inference strategies we used for rods will not work. However, the distribution of energy in the motion spectrum is still predictive of the fabric's material properties. For example, note how in Figure 1 the location of a fabric's resonant band shifts to the right with increasing area weight. Our approach is to use the motion spectra directly as features, and learn a regression model that maps these features to material properties.

As feature vectors we chose $N = 150$ uniform samples of the normalized motion spectra from 0 to 15 Hz. To reduce the effect of noise, we smooth the recovered motion spectra using a Gaussian with standard deviation $\frac{15}{2(N-1)}$ Hz.

Inference: We learn regression models that map the motion spectra to the log of ground truth stiffness or area weight measurements provided in [5]. Models are fit to the log of measurements in order to directly compare with results presented in [5]. Fitting a regression model directly to the processed motion spectra results in overfitting. Instead, we have explored two standard regression methods that reduce the dimensionality of the data: Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR). Both methods perform comparably, suggesting that the power of our algorithm is in the features, the recovered motion spectra, rather than the regression model. In this paper, we show results of the trained PLSR model. Additional results from PCR can be found in the supplemental material.

Cross Validation: Due to the small number of fabrics in the dataset, we use a leave-one-out method for training and testing. Precisely, all data corresponding to a fabric are removed from training of the regression parameters when predicting the material properties of that fabric. Using this method, we estimate the performance of our model on predicting the material properties of a previously unseen fabric. Performance was evaluated using a varying number of PLSR components. From this evaluation we chose a reduced number of PLSR dimensions, M, that is both robust and results in high accuracy for both material properties. For results presented in this paper, we used $M = 2$ and $M = 5$ for the ambient force model and acoustic model respectively. Refer to Figure 9.

Testing Invariance: We saw in Section 4 that our features should be invariant to changes in viewpoint. Here we test this invariance by training and testing on videos captured under different conditions. In total we have four conditions for fabrics: ambient (A) and acoustic (S) excitations, each captured from two different viewpoints (the left point grey (L) and right SLR (R) cameras). We used the same leave-one-out validation strategy when training and testing data were taken from different conditions.

7.2 Results

Our estimates of material properties are well correlated with the log of ground truth measurements (refer to Table 4). In all cases, even when testing under conditions with different viewpoints and excitation forces from the training data, our estimates outperform previous video-based fabric measurements [5] in predicting both stiffness and area weight.

Figure 10 contains correlation plots corresponding to the conditions presented in Table 4. These plots compare our algorithm's

Stiffness - Testing

Training	A/LR	A/L	A/R	S/LR	S/L	S/R
A/LR	0.89	-	-	0.80	-	-
A/L	-	0.89	0.89	-	0.73	0.80
A/R	-	0.87	0.89	-	0.74	0.77
S/LR	0.88	-	-	0.90	-	-
S/L	-	0.87	0.88	-	0.87	0.89
S/R	-	0.86	0.87	-	0.88	0.88

Area Weight - Testing

Training	A/LR	A/L	A/R	S/LR	S/L	S/R
A/LR	0.95	-	-	0.90	-	-
A/L	-	0.94	0.95	-	0.87	0.90
A/R	-	0.94	0.95	-	0.87	0.87
S/LR	0.93	-	-	0.96	-	-
S/L	-	0.92	0.93	-	0.96	0.96
S/R	-	0.91	0.92	-	0.96	0.95

TABLE 3
The Pearson correlation R value obtained when training and testing a PLSR model on videos captured under different excitation and viewpoint conditions. The testing and training shorthand notation specifies excitation/viewpoint using abbreviations for the four possible conditions: ambient excitation (A), acoustic excitation (S), left camera viewpoint (L) and right camera viewpoint (R). Results are comparable to training and testing on the same viewpoint, suggesting that our features are somewhat invariant to the direction in which the material is observed. Note that all combinations of excitation and viewpoint perform better than results reported in [5].

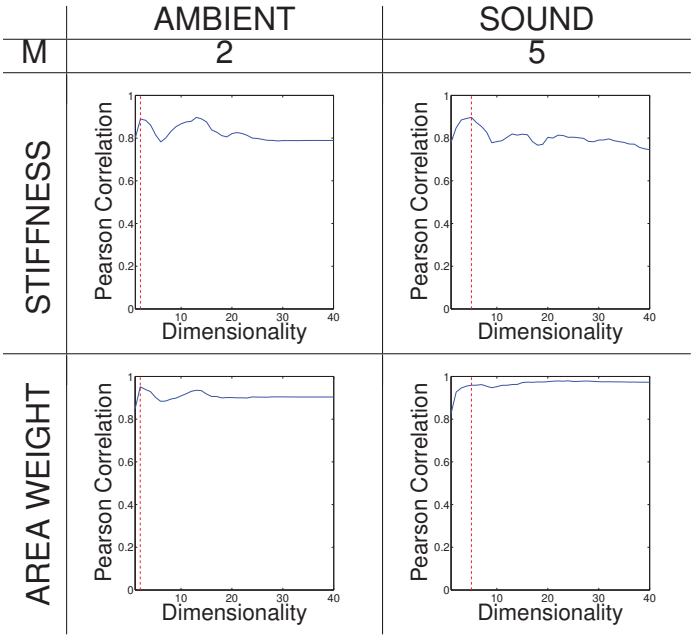


Fig. 9. The Pearson product correlation value between predicted results and the ground truth measured properties when fitting a model with a varying number of components (dimensionality). The number of components, M, was selected for each model by choosing a value that resulted in good accuracy for both material properties (stiffness and area weight). These selected M values are specified above and are indicated on the plots as a vertical red line.

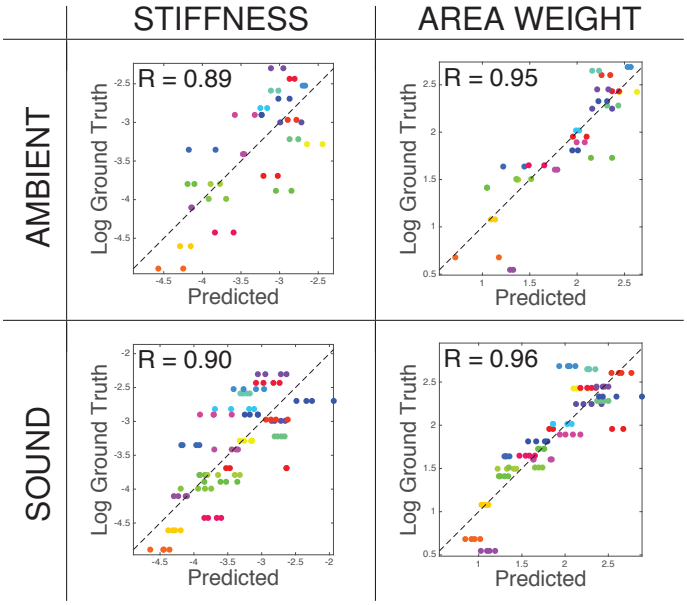


Fig. 10. Comparisons between ground truth and PLSR model predictions on material properties estimated from videos of fabric excited by ambient forces and acoustic waves. Each circle in the plots represents the estimated properties from a single video. Identical colors correspond to the same fabric. The Pearson product-moment correlation coefficient (R-value) averaged across video samples containing the same fabric is displayed.

	[5]	Ambient	Sound
Stiffness	$R = 0.71$ $\% = 17.2$	$R = 0.89$ $\% = 12.3$ $\tau = 0.70$	$R = 0.90$ $\% = 12.5$ $\tau = 0.74$
Area Weight	$R = 0.86$ $\% = 13.8$	$R = 0.95$ $\% = 15.7$ $\tau = 0.86$	$R = 0.96$ $\% = 13.3$ $\tau = 0.85$

TABLE 4
The Pearson correlation value (R), Percentage Error ($\%$), and Kendall Tau (τ) measures of performance for our PLSR model compared to the performance of a previous video-based fabric property estimation method [5]. The model was trained and tested separately on videos of fabric excited by acoustic waves (Sound) and ambient forces (Ambient).

predicted measurements of stiffness and area weight to the log of ground truth measurements when models were trained and tested on videos of fabrics excited by ambient forces and acoustic waves separately.

We test the invariance of an object's extracted motion spectra to excitation and viewpoint change by training the regression model on the extracted features from one excitation/viewpoint combination and testing on the extracted features from another combination. Table 3 shows that correlation results across all combinations of training and testing data are comparable to training and testing on the same viewpoint and excitation. Figure 11 visually shows our estimates are still well correlated with ground truth measurements when the training and testing is performed using different cameras, viewpoints, and excitation forces.

Frequency Sensitivity and Modes The theory in Section 3 describes a predictable relationship between resonant frequencies and material properties. However, our regression model has no explicit notion of resonant frequencies; it simply looks for predictive patterns in the spectra of training data. By analyzing the sensitivity of our recovered regression models we can see which frequencies are most predictive of material properties in our fabrics. From the estimated regression coefficients (β_m) and

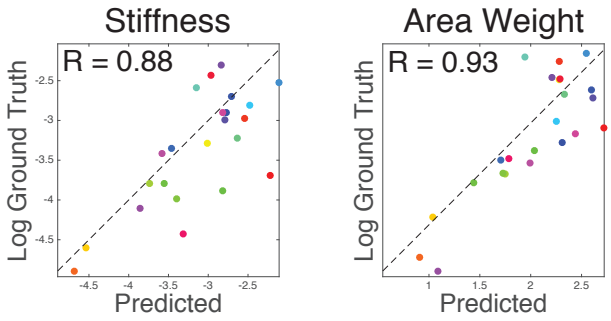


Fig. 11. The features we use to estimate material properties are somewhat invariant to changes in excitation force and viewpoint. Here we show a comparison between ground truth material properties and PLSR model predictions when using models trained on Point Grey (left viewpoint) videos of fabric exposed to acoustic waves, but tested on SLR videos (right viewpoint) of fabric exposed to ambient forces. Although the training and testing conditions are different, there is still a strong correlation between our results and the ground truth properties.

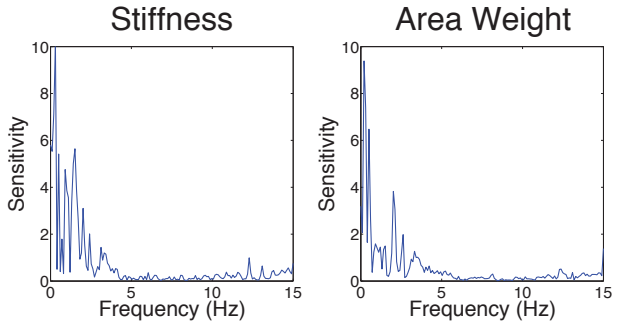


Fig. 12. The sensitivity of each acoustically trained model to frequency regions in the motion spectrum. These sensitivity plots suggest that energy in the 0-5 Hz range is most predictive of a fabric's area weight and stiffness.

dimensionality reducing basis vectors (E_m), the sensitivity (S) is computed as:

$$S = \sqrt{\left(\sum_{m=1}^{M} \beta_m E_m\right)^2} \tag{14}$$

Since the regression model for each of our fabrics is recovered using leave-one-out cross validation, we average the computed sensitivities across models to obtain a single measure of sensitivity for each material property.

Figure 12 shows that frequencies in the 0-5 Hz range were most predictive of material properties in our fabrics. By visualizing the pattern of relative pixel motion recovered for a specific frequency, we see that the fabrics' dominant vibration modes often appear in the frequency range of 0-5 Hz (see Figure 13). This suggests that our models use the same relationship between resonant frequencies and material properties predicted by modal analysis.

8 Detecting Changes in Resonance: Glasses of Water

There are many cases where *changes* in an object's resonant frequencies may be useful even when the contributions of material and geometry are left ambiguous. For example, the resonant frequencies of a leaking container will change over time as the container empties. In such a case, the changing resonance indicates a leak, regardless of specific structural or material properties. Similarly, a change in the resonance of a load-bearing structure may call for close attention, regardless of whether the change is caused by material weakening or an unseen change in geometry. One advantage of using resonance in such a scenario is that the source of the problem, or change, does not have to be visible - shifting frequencies at visible parts of the object may reveal hidden or occluded changes. In this section we show an experiment, analogous to the example of a leaking container, to demonstrate how our recovered motion spectra could be used to detect hidden changes to an object.

The following experiment demonstrates that we can infer when a wine glass is empty or full by observing the vibrations of its rim. For this to be the case, the changes in resonant frequencies that result from adding liquid to a glass must be significant compared to natural variations over time, or natural variations in resonant frequencies among the glasses. We compare motion spectra extracted from two videos and show that the addition of water results in a shift of the spectra's peaks. In the first video, all three glasses were left empty. In the second, the middle glass was filled with water.

Setup: Three wine glasses were placed on a table (Figure 14 left) next to a loudspeaker and partially occluded so that their contents were hidden from view (Figure 14 middle, top). The tops of these wine glasses were filmed to recover vibrations caused by a loudspeaker - once with all three glasses empty and once with only the center glass filled approximately two-thirds with water. Our goal was to see whether the hidden addition of water to the center glass could be easily detected in our recovered motion spectra.

Excitation: We played a 15 second linear chirp of frequencies ranging from 200 Hz to 800 Hz through the loudspeaker.

Video Capture: The tops of the glasses were filmed with a Phantom v10 high-speed camera at 2500 fps for approximately 17.3 seconds. The video was captured at a resolution of 1248x153 pixels (an example frame is given in Figure 14 middle, bottom). To evaluate the motion spectrum for each glass separately, a mask that segmented a single glass from the video frame was applied to the local, pixel motion spectra before averaging down to a single spectrum.

8.1 Results

Figure 14 (right) shows the motion spectra recovered from each glass in each of the two videos. In the spectra recovered from the first video, we see that the empty glasses have resonant peaks within 30 Hz of one another. In the spectra recovered from the second video, we don't see noticeable changes in the resonant frequencies of the empty glasses, but the water has shifted the resonant frequencies of the middle glass by approximately 76 Hz.

9 Comparison With Traditional Vibrometry

The motion spectra we recover from video are analogous to spectra derived from laser vibrometers and accelerometers for traditional vibration analysis. To compare these different types of sensors we conducted an experiment where a steel cantilever beam was measured simultaneously with a high-speed camera, a laser vibrometer, and a piezoelectric accelerometer (refer to

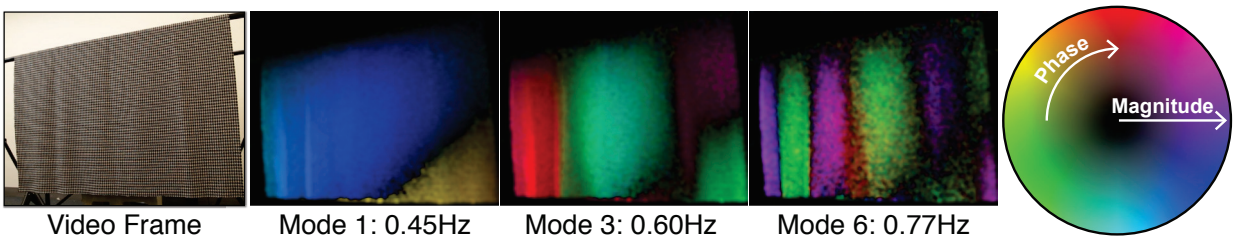


Fig. 13. A sample of the recovered motion patterns for predictive frequencies identified by the regression models. These recovered motion patterns often resemble a fabric's mode shapes. Phase specifies the relative direction of the motion signal. Pixels moving in opposite directions are colored with hue from opposite sides of the color wheel.

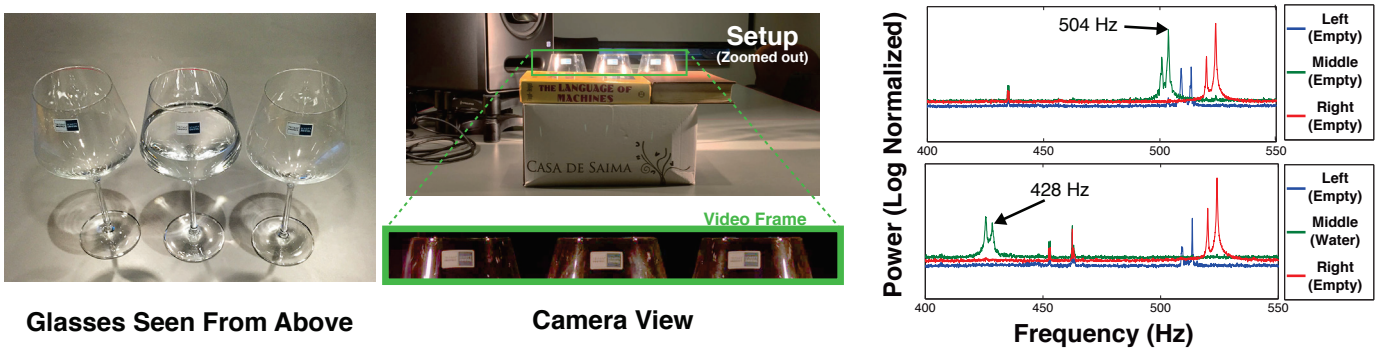


Fig. 14. (Left) Three wine glasses are set on a table. They are filmed twice - once with all three glasses empty and once with the middle glass partially filled with water (shown left). (Middle, top) The glasses are partially occluded to that their contents are not visible, and a nearby loudspeaker plays a 15 second linear chirp of frequencies ranging from 200 Hz to 800 Hz. (Middle, bottom) The rims of the glasses are filmed at 2500 Hz. (Right) Masks are used to extract the motion spectra of each glass from each video separately. (Right, top) When all glasses are empty, they show resonant peaks within the range of 500-530 Hz. (Right, bottom) When only the middle glass is filled with water, resonant frequencies of the empty glasses remain unchanged, while the resonant peak of the glass containing water shifts by 76 Hz, to 428 Hz.

Figure 15). A shaker was mounted to the top of the beam, and driven with a sum of sinusoids at resonant modes of the beam. The accelerometer was mounted directly to the beam, the laser vibrometer measured the motion of the accelerometer, and a high-speed camera recorded a video of the accelerometer and beam motion. All three measurement methods were used concurrently in time, measuring the same vibrations of the beam at the same location. The laser vibrometer and accelerometers sampled at 9000 Hz, while the video captured 2000 fps. each sensor recorded for approximately 15 seconds.

The work of [35] also compared accelerations measured with a laser vibromater with video measurements, focusing on a time domain analysis. Here, we study differences in the spectra of recovered motions. It is natural for each sensor to produce slightly different spectra, as each tests a different derivative of position (the accelerometer measures acceleration, the vibrometer measures velocity, and our method measures position). However, we focus specifically on comparing the resonant frequencies and damping estimated in each case.

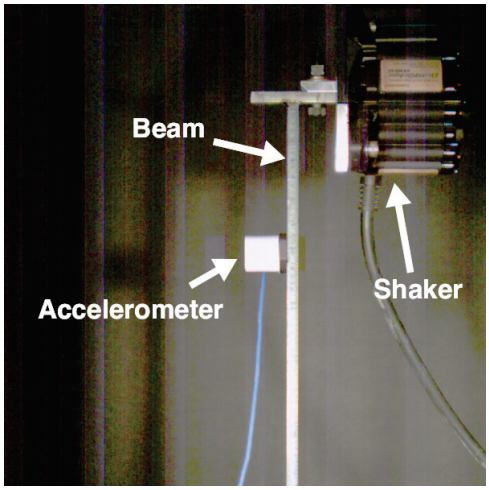


Fig. 15. Example frame from our video of a forced beam, captured simultaneously with a video, laser vibrometer, and accelerometer.

9.1 Frequency and Damping Estimates

Spectra recovered using each of the three techniques can be seen in Figure 16. Mode frequencies for each of these spectra were detected as the local maximum around each resonant peak, and are shown in Table 5. As all three sensors were recording the same object, we used the same range of frequencies to fit damping around each peak (± 3 Hz). Recovered damping values can be found in Table 6.

Figure 16 shows that the overall shape of spectra recovered using each of the three methods is very similar, though some harmonic artifacts are present in the spectra recovered using our technique. Table 5 shows that all three methods agree on the locations of resonant frequencies to within quantization errors. Table 6 shows that our method disagrees with the accelerometer and vibrometer on two out of three of the modes, with our strongest disagreement in the fundamental, where our estimate is approximately 39% higher. This amount of error is large relative

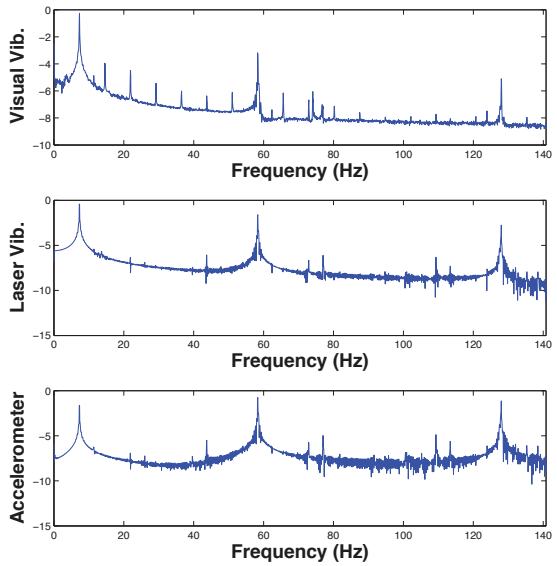


Fig. 16. Recovered motion spectra from our beam experiment using our visual vibrometry technique (top), a laser Doppler vibrometer (middle), and an accelerometer (bottom).

Estimated Frequency	Mode 1	Mode 2	Mode 3
Visual Vibrometry	7.3Hz	58.3Hz	128Hz
Laser Vibrometer	7.3Hz	58.3Hz	128Hz
Accelerometer	7.3Hz	58.3Hz	128Hz

TABLE 5
Recovered beam mode frequencies using our technique, a laser Doppler vibrometer, and an accelerometer. All mode frequencies agree to within the quantization error of our sampling.

to the differences in damping ratios for similar metals, but small compared to the differences between metals and materials like wood or rubber.

10 Discussion

We have shown that it is possible to learn about the material properties of visible objects by analyzing subtle, often imperceptible, vibrations in video. This can be done in an active manner by recording video of an object responding to sound, or, in some cases, even passively by observing an object move naturally within its environment.

The rod experiments in Section 6 demonstrate how our technique can be used as a low cost alternative to laser vibrometers in settings that are typical for testing manufactured parts (e.g. for airplanes and automobiles). Our technique also offers an affordable way to apply established methods from structural engineering to applications that require more than single point measurements.

The fabric experiments in Section 7 address a relatively unexplored area of potential for vibration analysis. While traditional applications of vibrometry are often limited by the need for detailed measurements and analysis of geometry, the ubiquity and passive nature of video offers unique potential as a way to enable data-driven alternative approaches. Our results on fabrics demonstrate that the relationship between motion spectra and material properties can be learned, and suggests that traditional vibration analysis may be extended to applications where geometry is unknown and only loosely controlled.

Damping Ratio	Mode 1	Mode 2	Mode 3
Visual Vibrometry	6.1×10^{-3}	6.5×10^{-4}	3.9×10^{-4}
Laser Vibrometer	4.4×10^{-3}	6.5×10^{-4}	2.9×10^{-4}
Accelerometer	4.4×10^{-3}	6.5×10^{-4}	2.9×10^{-4}

TABLE 6
Damping ratios computed using spectra derived from the three different sensors. Each damping ratio was computed by fitting a Lorentzian to a 6 Hz region around each identified mode frequency.

The simple wine glass experiment in Section 8 highlights a use case that could be applicable to structural health monitoring and quality control in manufacturing. In these scenarios, precise geometry and material properties are not necessary; by directly comparing the motion spectra of similar objects, or of one object over time, it may be possible to detect failures or defects.

Our work offers cameras as a promising alternative to the specialized, laser-based equipment that is traditionally used in many applications in civil engineering and manufacturing. The presented results suggest that the motion spectra we extract from video can be a powerful feature for scene understanding. Vibration theory in Section 3 suggests that even when geometry is ambiguous, the spectra constrain the physical properties of visible objects. These constraints could be useful for many tasks in computer vision - just as color is often useful despite being an ambiguous product of reflectance and illumination. We believe that a video's extracted motion spectra can be a powerful tool for reasoning about the physical properties of objects, in controlled settings as well as in the wild.

Acknowledgments

Dr. Dirk Smit of Shell Research proposed to us the analysis of small displacements for structural health monitoring. We would also like to thank Neal Wadhwa, Gautham J. Mysore, and Danny M. Kaufman. This work was supported by NSF Robust Intelligence 1212849 Reconstructive Recognition, NSF CGV-1111415, Shell Research, and Qatar Computing Research Institute. A. Davis and K. Bouman were partially supported by NSF GRFP fellowships.

References

[1] V. Aranchuk, A. Lal, J. M. Sabatier, and C. Hess. Multi-beam laser doppler vibrometer for landmine detection. *Optical Engineering*, 45(10):104302–104302, 2006.

[2] ASTM Standard E756 - 05. Standard Test Method for Measuring Vibration-Damping Properties of Materials. ASTM International, West Conshohocken, PA, 2010.

[3] W. E. Baker, W. E. Woolam, and D. Young. Air and internal damping of thin cantilever beams. *International Journal of Mechanical Sciences*, 9(11):743–766, 1967.

[4] K. S. Bhat, C. D. Twigg, J. K. Hodgins, P. K. Khosla, Z. Popović, and S. M. Seitz. Estimating cloth simulation parameters from video. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '03, pages 37–51, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[5] K. L. Bouman, B. Xiao, P. Battaglia, and W. T. Freeman. Estimating the material properties of fabric from video. *Computer Vision, IEEE International Conference on*, 0:1984–1991, 2013.

[6] O. Buyukozturk, R. Haupt, C. Tuakta, and J. Chen. Remote detection of debonding in frp-strengthened concrete structures using acoustic-laser technique. In *Nondestructive Testing of Materials and Structures*, pages 19–24. Springer, 2013.

[7] P. Castellini, N. Paone, and E. P. Tomasini. The laser doppler vibrometer as an instrument for nonintrusive diagnostic of works of art: application to fresco paintings. *Optics and Lasers in Engineering*, 25(4):227–246, 1996.

[8] J. G. Chen, R. W. Haupt, and O. Buyukozturk. Acoustic-laser vibrometry technique for the noncontact detection of discontinuities in fiber reinforced polymer-retrofitted concrete. *Materials evaluation*, 72(10):1305–1313, 2014.
[9] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk. Structural modal identification through high speed camera video: Motion magnification. In *Topics in Modal Analysis I, Volume 7*, pages 191–197. Springer, 2014.
[10] J. G. Chen, N. Wadhwa, Y.-J. Cha, F. Durand, W. T. Freeman, and O. Buyukozturk. Modal identification of simple structures with high-speed video using motion magnification. *Journal of Sound and Vibration*, 345:58–71, 2015.
[11] L. Collini, R. Garziera, and F. Mangiavacca. Development, experimental validation and tuning of a contact-less technique for the health monitoring of antique frescoes. *NDT & E International*, 44(2):152–157, 2011.
[12] L. Cremer and M. Heckl. *Structure-borne sound: structural vibrations and sound radiation at audio frequencies*. Springer Science & Business Media, 2013.
[13] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman. Visual vibrometry: Estimating material properties from small motion in video. June 2015.
[14] A. Davis, J. G. Chen, and F. Durand. Image-space modal bases for plausible manipulation of objects in video. *ACM Trans. Graph.*, 34(6):239:1–239:7, Oct. 2015.
[15] A. Davis, M. Rubinstein, N. Wadhwa, G. J. Mysore, F. Durand, and W. T. Freeman. The visual microphone: Passive recovery of sound from video. *ACM Trans. Graph.*, 33(4):79:1–79:10, July 2014.
[16] F. Durst, A. Melling, and J. H. Whitelaw. Principles and practice of laser-doppler anemometry. *NASA STI/Recon Technical Report A*, 76:47019, 1976.
[17] T. Emge and O. Buyukozturk. Remote nondestructive testing of composite-steel interface by acoustic laser vibrometry. *Materials evaluation*, 70(12):1401–1410, 2012.
[18] D. J. Fleet and A. D. Jepson. Computation of component image velocity from local phase information. *International Journal of Computer Vision*, 5(1):77–104, 1990.
[19] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 2003.
[20] R. W. Haupt and K. D. Rolt. Standoff acoustic laser technique to locate buried land mines. *Lincoln Laboratory Journal*, 15(1):3–22, 2005.
[21] Y.-x. Ho, M. S. Landy, and L. T. Maloney. How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, 2006.
[22] N. Jojic and T. S. Huang. Estimating cloth draping parameters from range data. In *In International Workshop on Synthetic-Natural Hybrid Coding and 3-D Imaging*, pages 73–76, 1997.
[23] S. Kawabata and M. Niwa. Fabric performance in clothing and clothing manufacture. *Journal of the Textile Institute*, 1989.
[24] H. Ku. Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards*, 70(4), 1966.
[25] C. Liu, L. Sharan, E. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. 2010.
[26] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vision*, 40(1):49–70, Oct. 2000.
[27] M. Rubinstein. *Analysis and Visualization of Temporal Variations in Video*. PhD thesis, Massachusetts Institute of Technology, Feb 2014.
[28] C. Santulli and G. Jeronimidis. Development of a method for nondestructive testing of fruits using scanning laser vibrometry (SLV). *NDT.net*, 11(10), 2006.
[29] A. A. Shabana. *Theory of vibration*, volume 2. Springer, 1991.
[30] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson. Image statistics for surface reflectance perception. *Journal of the Optical Society of America. A, Optics, image science, and vision*, Apr. 2008.
[31] P. Shull. *Nondestructive evaluation: theory, techniques, and applications*, volume 142. CRC, 2002.
[32] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger. Shiftable multi-scale transforms. *IEEE Trans. Info. Theory*, 2(38):587–607, 1992.
[33] A. Stanbridge and D. Ewins. Modal testing using a scanning laser doppler vibrometer. *Mechanical Systems and Signal Processing*, 13(2):255–270, 1999.
[34] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):80, 2013.
[35] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. *ACM Trans. Graph. (Proceedings SIGGRAPH 2013)*, 32(4), 2013.
[36] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Riesz pyramid for fast phase-based video magnification. In *Computational Photography (ICCP), 2014 IEEE International Conference on*. IEEE, 2014.
[37] H. Wang, J. F. O'Brien, and R. Ramamoorthi. Data-driven elastic models for cloth: modeling and measurement. SIGGRAPH, 2011.
[38] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4):65, 2012.

Abe Davis is a PhD candidate in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), where he is advised by Frédo Durand. He received his B.S. with honors in computer science from Stanford University in 2010, and his M.S. in computer science and electrical engineering from MIT in 2013. His research spans computer vision, graphics, and computational photography, as well as nondestructive testing and structural health monitoring. Abe's PhD dissertation focuses on the analysis of vibrations in video, with a broad range of applications. Abe is a recipient of the NSF Graduate Research Fellowship, and the MathWorks Fellowship.

Katherine L. Bouman is a Ph.D. candidate in the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT), under the supervision of William T. Freeman. She previously received a B.S.E. in Electrical Engineering from the University of Michigan, Ann Arbor, MI in 2011 and an S.M. degree in Electrical Engineering and Computer Science from MIT, Cambridge, MA in 2013. Her masters thesis on estimating material properties of fabric from video won the Ernst A. Guillemin Thesis Prize for outstanding S.M. thesis in electrical engineering at MIT. Katherine is the recipient of the NSF Graduate Fellowship, the Irwin Mark Jacobs and Joan Klein Jacobs Presidential Fellowship, and is a Goldwater Scholar. Her research interests include computer vision, computational photography, and computational imaging.

Justin G. Chen received his Ph.D in Structures and Materials from the Massachusetts Institute of Technology in 2016, as well as an S.M. in Civil and Environmental Engineering also from MIT in 2013. He received his B.S. in Physics from the California Institute of Technology in 2009. He is currently working as a post-doctoral associate at MIT and his main research interests are in the application of novel sensor systems to structural health monitoring and non-destructive testing. Justin was a recipient of an American Society for Nondestructive Testing (ASNT) Fellowship Award in 2011.

Michael Rubinstein is a Research Scientist at Google, Cambridge MA. His research is at the intersection of computer vision and graphics, and focuses on areas in image and video processing, and computational photography. Michael received his PhD in Computer Science from MIT in 2013. He was awarded the Microsoft Research PhD Fellowship in 2012, and the NVIDIA Graduate Fellowship in 2011. His doctoral dissertation on analysis and visualization of temporal variations in video won the George M. Sprowls Award for outstanding doctoral thesis in Computer Science at MIT. Prior to joining Google he was a Postdoctoral Researcher at Microsoft Research.

Oral Büyüköztürk received his Ph.D. degree in Structural Engineering from Cornell University in 1970. He joined the Massachusetts Institute of Technology (MIT), Cambridge MA, in 1976, where he is Professor of Civil and Environmental Engineering, and Director of the Laboratory for Infrastructure Science and Sustainability. His early work prior to joining MIT involved design and safety analysis of nuclear power structures, and at Brown University, development of nonlinear finite element models and computational engineering capabilities. His early research at MIT involved design and analysis of major energy facilities such as nuclear energy and offshore oil production structures, and thermo-mechanical analysis of coal gasification vessels. His recent and current research focuses on infrastructure sustainability, design for durable and energy efficient materials through multiscale analysis using molecular dynamics (MD), intelligent structures and materials, structural health monitoring (SHM), and nondestructive testing (NDT). His work also includes design and assessment of concrete structures, nuclear containment systems, durability of materials, earthquake engineering, interface fracture mechanics, and fiber-reinforced polymer (FRP) composites in structural rehabilitation. He has extensively published through refereed journals and edited books, made more than 200 invited and keynote presentations around the world, and served in different capacities in over 20 technical committees. His awards include Golden Mirko Ro Medal of the Swiss Federal Research Laboratory for Materials Science and Technology; Fellow (non-resident) Royal Society of Edinburgh, Scotlands National Academy of Science and Letters; 2008 and 2011 ASNT National Faculty Fellowship Awards; Fellow, American Concrete Institute (ACI), and various Best Paper Awards jointly with his students.

Frédo Durand received the PhD degree from Grenoble University, France, in 1999, supervised by Claude Puech and George Drettakis. He is an associate professor in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology (MIT) and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL). From 1999 till 2002, he was a postdoctoral research in the MIT Computer Graphics Group with Julie Dorsey. He works both on synthetic image generation and computational photography, where new algorithms afford powerful image enhancement and the design of imaging system that can record richer information about a scene. His research interests span most aspects of picture generation and creation, with emphasis on mathematical analysis, signal processing, and inspiration from perceptual sciences. He coorganized the first Symposium on Computational Photography and Video in 2005, the first International Conference on Computational Photography in 2009, and was on the advisory board of the Image and Meaning 2 conference. He received an inaugural Eurographics Young Researcher Award in 2004, an US National Science Foundation (NSF) CAREER award in 2005, an inaugural Microsoft Research New Faculty Fellowship in 2005, a Sloan fellowship in 2006, a Spira award for distinguished teaching in 2007, and the ACM SIGGRAPH Computer Graphics Achievement Award in 2016.

William T. Freeman is the Thomas and Gerd Perkins Professor of Electrical Engineering and Computer Science at MIT, and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL) there. He was the Associate Department Head from 2011 - 2014. His current research interests include machine learning applied to computer vision, Bayesian models of visual perception,and computational photography. He received outstanding paper awards at computer vision or machine learning conferences in 1997, 2006, 2009 and 2012, and test-of-time awards for papers from 1990 and 1995. Previous research topics include steerable filters and pyramids, orientation histograms, the generic viewpoint assumption, color constancy, computer vision for computer games, and belief propagation in networks with loops. He is active in the program or organizing committees of computer vision, graphics, and machine learning conferences. He was the program co-chair for ICCV 2005, and for CVPR 2013.