This CVPR2015 paper is the Open Access version, provided by the Computer Vision Foundation.
The authoritative version of this paper is available in IEEE Xplore.

# Visual Vibrometry: Estimating Material Properties from Small Motions in Video

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

[1]Massachusetts Institute of Technology [2]Google Research
{abedavis,klbouman,ju21743,fredo,billf}@mit.edu mrub@google.com

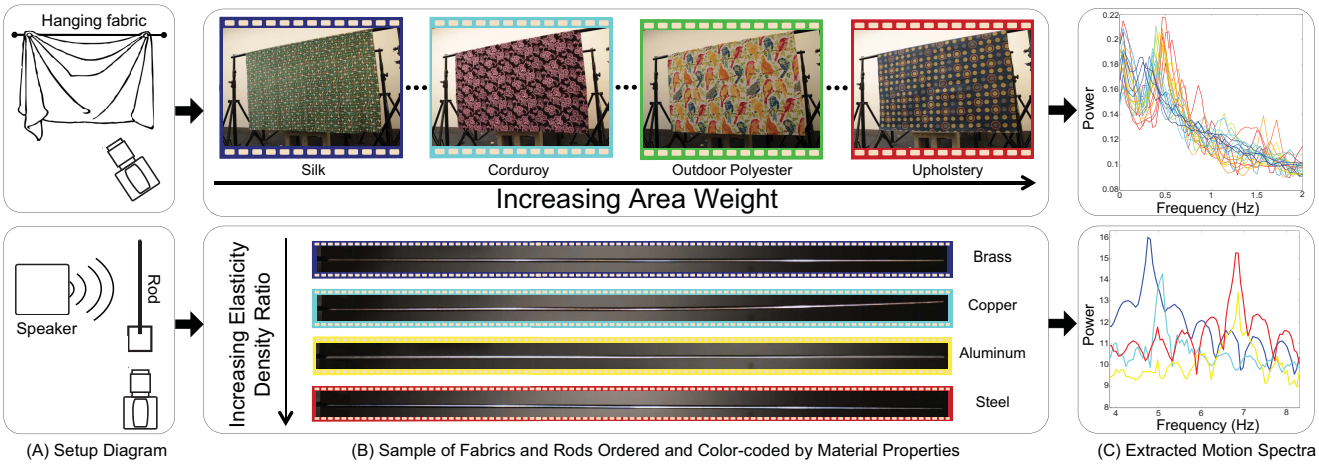


Figure 1. We present a method for estimating material properties of an object by examining small motions in video. (A) We record video of different fabrics and clamped rods exposed to small forces such as sound or natural air currents in a room. (B) We show fabrics (top) color-coded and ordered by area 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, resp (blue to red). By examining these spectra, we can make inferences about the material properties of objects.

## 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 shapes and frequencies of these modes depend on the structure and material properties of an object. Focusing on the case where geometry is known or fixed, we show how information about an object's modes of vibration can be extracted from video and used to make inferences about that object's material properties. We demonstrate our approach by estimating material properties for a variety of rods and fabrics by passively observing their motion in high-speed and regular framerate video.*

* Joint first author
† Part of this work was done while the author was at Microsoft Research

## 1. Introduction

Understanding a scene involves more than just recognizing object categories or 3D shape. The physical properties of objects, such as the way they move and bend, can be critical for applications that involve assessing or interacting with the world. In the field of non-destructive testing (NDT), an object's physical properties are often studied through the measurement of its vibrations using contact sensors or expensive laser vibrometers. In both cases, measurements are often limited to a small set of discrete points. In contrast, we leverage the ubiquity and high spatial resolution of video cameras to extract physical properties from video. These physical properties are then used to make inferences about the object's underlying material properties. We are inspired by recent work in computer vision, but seek to bridge the gap with engineering techniques and focus on fundamentals of vibration analysis.

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 related to its geometry and material properties. In this paper, we show how this connection can be used to estimate the material properties of an object with fixed or known geometry from video.

In this paper we review established theory on modal vibrations, and connect this theory to features that can be extracted from video. We then show how these features can be used to estimate the material properties of objects with fixed or known geometry. We demonstrate this on two sets of objects: clamped rods and hanging fabrics. With each set of objects we explore a different method to resolve the ambiguous contribution of structure (geometry) and material properties to an object's vibrations. Our rod experiments accomplish this with careful measurements in a setting that resembles typical engineering applications. Our fabric experiments instead explore the potential of a learning approach with more direct applications in computer vision.

## 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.

**Traditional Vibration Analysis:** The use of vibrations to estimate material properties is an established tool used in a variety of engineering disciplines. Especially related to this paper is work in the field of NDT, where techniques based on ultrasound are common. However, these techniques often require direct contact with the object being measured [25]. 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 [11]. Laser vibrometers have been used to non-destructively examine valuable paintings [5, 9], detect land mines [15, 1], test fruit [22], find defects in composite materials [4, 6, 12], and even test vibration modes of small structures [27]. 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 [27, 1], 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.

**Material Property Estimation from Video:** Previous work in computer vision has focused on estimating material properties from static images [24, 19, 16, 14]. 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 [2, 17]. Bouman et al. [3] adopted a learning approach that allowed them to estimate material properties from a video of fabric moving under wind forces. As with our experiments in Section 7, they estimate material properties directly from video statistics using a regression strategy. Their 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.

**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 [31, 28, 29, 21]. As with many of these works, our method uses phase variations of the complex steerable pyramid [26, 20] to represent small local motions in video. In recent work, Chen et al. [7, 8] 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. [10], but where they focus on using vibrations in video to recover sound, we use them to learn about the physical properties of visible objects.

## 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 much 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 [23].

### 3.1. Eigenmode Analysis

In modal analysis a solid object with homogeneous material properties is modeled as a system of point masses con-

nected by springs [23]. Intuitively, rigid objects are approximated with stiff springs and dense objects are approximated with heavy masses. Consider the mass matrix $\mathbf{M}$ of inertias between points, and the matrix $\mathbf{K}$ of spring stiffnesses. The differential equation of motion for this system is given by:

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

where $x$ and $\ddot{x}$ are vectors describing the displacement and acceleration of our points, respectively. Looking for solutions to this equation of the form:

$$x = A\sin(\omega t + \phi), \tag{2}$$

we obtain a standard eigenvalue problem for our original equation of motion:

$$[\mathbf{K} - \omega_i^2\mathbf{M}]A_i = 0. \tag{3}$$

The eigenvalues $\omega_i^2$ are the square of resonant natural frequencies of our object, and the eigenvectors $A_i$ describe the normal modes of vibration that take place at these frequencies. General deformations of our object can then be expressed as linear combinations of these mode shapes.

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 eigenvectors and eigenvalues for our 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 our system but leaving the eigenvectors 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 to estimate the material properties of objects with common geometry. Our task for the rest of the paper is to learn the material properties of objects by observing their resonant frequencies $\omega_i$. For this we will leverage the fact that $\omega_i$ are global properties of an object - meaning they do not vary across the object's surface; only the amplitudes and local phases of modal vibration vary spatially, according to the mode shapes $A_i$.

## 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 $A_i$ and frequencies $\omega_i$.

**Local Motion Signals:** Local motion signals are derived from phase variations of the complex steerable pyramid [26, 20, 13]. We weigh local motions by the square of amplitudes in the pyramid as in [10], 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, but alternatives like optical flow could be equally valid.

**Motion Spectra:** Recall that the frequencies $\omega_i$ do not vary across an object's surface. This means that the power spectra of local motions across an object should have spikes at the same resonant frequencies. Therefore, we compute the global motion power spectrum for a video by averaging the power spectra of local motions extracted at every pixel. This leaves us with a single temporal power spectrum describing the frequencies of motion that exist in a video.

**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 test data sets are taken from different viewpoints and cameras.

**Mode Shapes:** The theoretical mode shapes $A_i$ 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. We adopt the visualization used in [10], where the image of local Fourier coefficients at a given frequency is displayed by mapping phase to hue and magnitude to brightness.

## 5. Method

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

**Excitation:** An object must move in order for us to observe its vibration modes. Some very deformable objects, like hanging fabric, may move enough with natural air currents for no additional forces to be necessary. For more rigid objects like 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 has been used for this purpose previously in NDT [6, 9, 4, 12, 15].

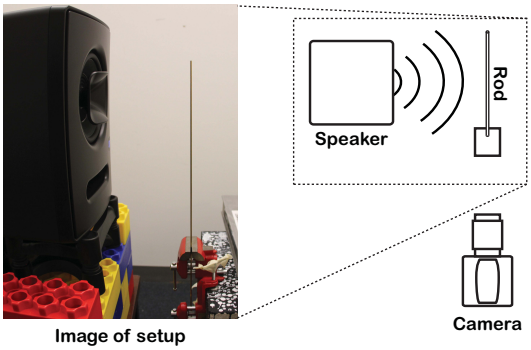


Figure 2. Rods were clamped to a concrete block next to a loudspeaker (shown left) and filmed with a high-speed camera. By analyzing small motions in the recorded video, we are able to find resonant frequencies of the rods and use them to estimate material properties.

**Video Capture:** To estimate an object's resonant frequencies we need to record at a high enough framerate to place these frequencies under the Nyquist limit. We should also ensure that videos capture enough periods at each mode frequency to sufficiently localize corresponding spikes in the Fourier domain. For rigid objects with high resonant frequencies this can be accomplished with short clips of high speed video. Deformable objects with low resonant frequencies can be captured with longer, lower-framerate video.

**Inference:** We explore two different strategies for inferring material properties from motion spectra. The first uses a voting scheme to extract resonant frequencies from motion spectra. This approach is very precise, but relies on detailed knowledge of the structure being observed (in our case rods). The second approach alleviates the need for detailed geometry and analysis by learning the relationship between recovered motion spectra and material properties from training data; we use this approach to estimate the properties of hanging fabrics.

# 6. Estimating Properties of Materials with Known Geometry: Rods

In our first experiments we estimate the material properties 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 show how this elastic modulus can instead be measured with a speaker and high-speed camera. .

**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 6.2 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 supplemental material.

**Excitation:** The excitation signal should be broad spectrum to ensure that multiple rod modes are activated. In [7, 8] 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 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 high speed camera. 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. Property Estimation

The vibrations of clamped rods are well studied [23]. A rod's fundamental frequency $\omega_1$ (corresponding to its first mode) is related to material properties by the equation:

$$\omega_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}} \tag{4}$$

where $d$ is the diameter of the rod, $L$ is its length, $\rho$ is its density and $E$ is its Young's modulus (measuring elasticity). Given the length and width of a rod, the task of estimating $\sqrt{\frac{E}{\rho}}$ 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:

$$\omega_i = \eta_i \omega_1,$$
$$\eta_1 = 1, \quad \eta_2 = 6.27, \quad \eta_3 = 17.55, \quad \eta_4 = 34.39 \tag{5}$$

where again $\omega_i$ is the resonant frequency for the $i$th mode. To distinguish modal frequencies from other spikes, we

[1] By solving the continuous analog to Equation 3 [23]

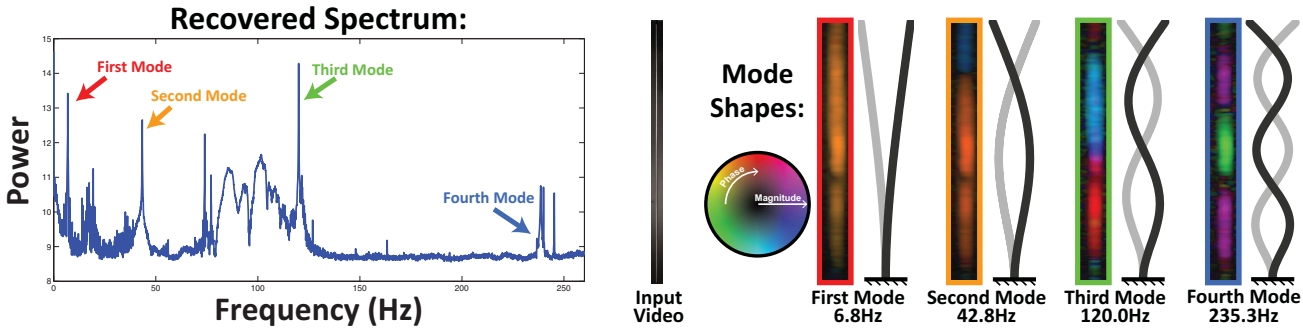


Figure 3. Finding vibration modes of a clamped brass rod: (Left) We recover a motion spectrum from 2.5 kHz video of a 22 inch clamped aluminum rod. Resonant frequencies are labeled. To distinguish resonant frequencies from other spikes in the spectrum, we look for energy at frequencies with ratios derived 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.

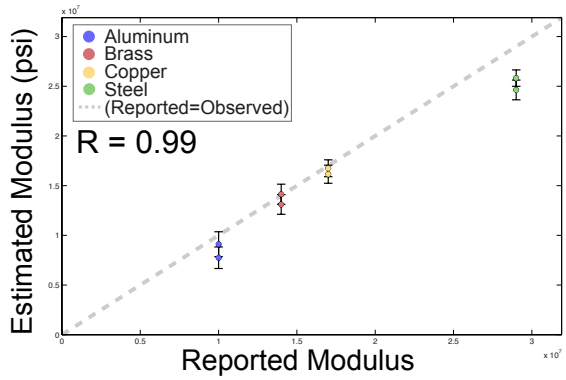


Figure 4. Estimating the elastic modulus of clamped rods: Young's moduli reported by the manufacturer are plotted against values estimated using our technique for aluminum, brass, copper, and steel. Estimated values are close to those reported by the manufacturer, with the largest discrepancies happening in 15 inch rods made of aluminum and steel.

look for energy in the recovered spectra that occurs in the ratios given by Equation 5. 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:

$$P\left(\omega = \omega_1 \middle| S\right) \propto \prod_{i=1}^{4} S(\omega\eta_i). \tag{6}$$

Using Equation 6, 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 5 to vote for the fourth resonant frequency.

### 6.2. Results

Under fixed but unknown geometry, the recovered fundamental frequencies provide a value proportional to $\sqrt{\frac{E}{\rho}}$. From this we can use Equation 4 with lengths and densities measured by a scale and measuring tape to compute the modulus of each rod. Figure 4 shows a plot of Young's moduli reported by the manufacturer against the values estimated using our technique. Error bars are calculated for each moduli by propagating error bounds for length, diameter, and density (see supplemental material for details). Percent errors are given in Table 1.

| % Error | Aluminum | Brass | Copper | Steel |
|---|---|---|---|---|
| 22 inches | -8.94 | 0.95 | -1.49 | -10.97 |
| 15 inches | -22.59 | -6.39 | -5.01 | -15.09 |

Table 1. Percent errors in estimating the Young's modulus for each rod.

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 the frequency of lights powered by AC current. However, it is unlikely that a majority of resonant frequencies will be masked in any single 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 supplemental material.

Our estimated moduli are close to, but consistently under, the reported values. One possible explanation for this is an incorrect estimate of where the clamp grabbed each rod in our setup. This suggests both a strength and weakness of the approach taken here - high precision that is very sensitive to accurate modeling of the structure being tested. Our next experiments address this issue by 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, $\eta_i$. 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 ge-

ometries 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 the current state of the art, even when trained using data captured from different viewpoints or using different excitation forces.

A number of metrics exist to describe the material properties of fabrics. These properties can be measured using setups such as the Kawabata system [18, 30]. In the work of Bouman, et al. [3], 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 [3] that relied on much larger motions produced by fans.

**Setup:** Each fabric specimen from [3] (width 43.5 to 44.5 inches across) was loosely draped over a bar and hung a length of 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.

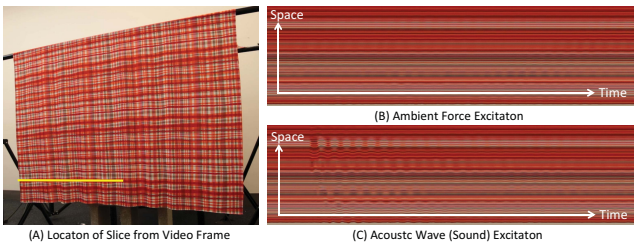


Figure 6. Videos of fabric excited by two different types of force were recorded. Here we see space $\times$ time slices from minute long videos of a fabric responding ambient forces (b) and sound (c). The motion is especially subtle in (b), but still encodes predictive information about the fabric's material properties.

**Excitation:** *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 6a 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" to the hanging fabric. We excited each fabric with a one second, logarithmic frequency ramp from 15 to 100 Hz. Figure 6b 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$\times$1080 pixel resolution) at 30 fps and a grayscale Point Grey camera (800$\times$600 pixel resolution) at 60 fps. The cameras recorded different viewpoints (see Figure 5), 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 (see Figure 1). Increased damping broadens the bandwidth around resonant frequencies, resulting in wide, overlapping resonant bands. Nonetheless, the distribution of energy in the motion spectrum is still very 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.

As features 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 [3]. Models are fit to the log of measurements in order to directly compare with results presented in [3]. 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

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

Figure 5. 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.

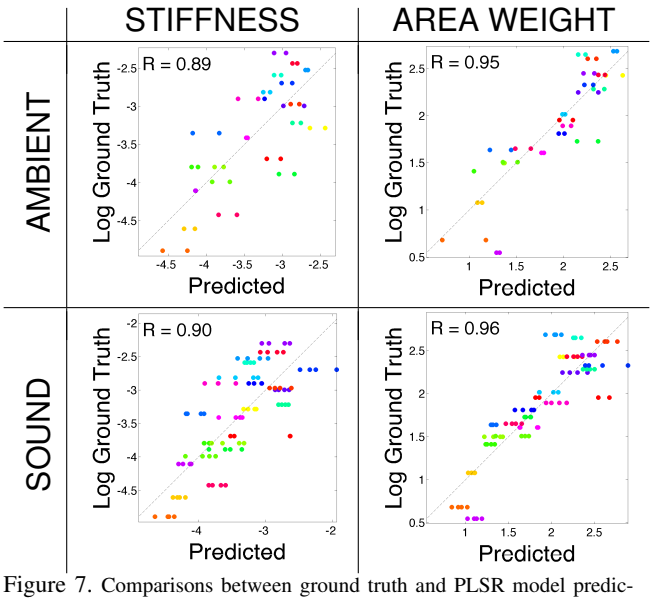


Figure 7. Comparisons between ground truth and PLSR model predictions on material properties estimated from videos of fabric excited by ambient forces and acoustic waves. Each circle in the plots represents the estimated properties from a single video. Identical colors correspond to the same fabric. The Pearson product-moment correlation coefficient (R-value) averaged across video samples containing the same fabric is displayed.

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.

**Testing Invariance:** We test the invariance of our features by training and testing on videos captured under different conditions. In total we have four conditions for fabrics: ambient and acoustic excitations, each captured from two different viewpoints. We used the same leave-one-out validation strategy when training and testing data were taken from different conditions.

| | [3] | Ambient | Sound |
|---|---|---|---|
| Stiffness | $R = 0.71$<br>$\% = 17.2$ | $R = 0.89$<br>$\% = 12.3$<br>$\tau = 0.70$ | $R = 0.90$<br>$\% = 12.5$<br>$\tau = 0.74$ |
| Area Weight | $R = 0.86$<br>$\% = 13.8$ | $R = 0.95$<br>$\% = 15.7$<br>$\tau = 0.86$ | $R = 0.96$<br>$\% = 13.3$<br>$\tau = 0.85$ |

Table 2. The Pearson correlation value ($R$), Percentage Error (%), and Kendall Tau ($\tau$) measures of performance for our PLSR model compared to the current state of the art algorithm [3]. The model was trained and tested separately on videos of fabric excited by acoustic waves (Sound) and ambient forces (Ambient).

### 7.2. Results

Our estimates of material properties are well correlated with the log of ground truth measurements (refer to Table 2). In all cases, even when testing under conditions with different viewpoints and excitation forces from the training data, our estimates outperform the current state of the art algorithm [3] in predicting both stiffness and area weight. The Pearson correlation R values obtained using every combination of viewpoint and excitation conditions for training and testing data can be found in the supplemental material.

Figure 7 contains correlation plots corresponding to the conditions presented in Table 2. These plots compare our algorithm's 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. Figure 8 shows that our estimates are still highly 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 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 pre-

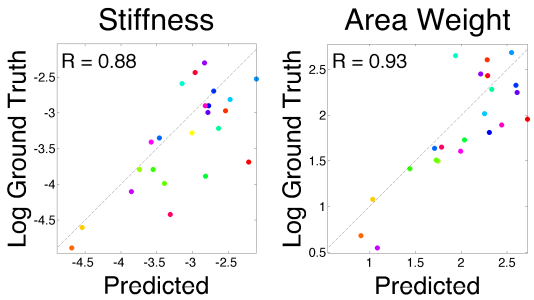


Figure 8. The features we use to estimate material properties are somewhat invariant to changes in excitation force and viewpoint. Here we show a comparison between ground truth material properties and PLSR model predictions when using models trained on Point Grey (left viewpoint) videos of fabric exposed to acoustic waves, but tested on SLR videos (right viewpoint) of fabric exposed to ambient forces. Although the training and testing conditions are different, we still perform well.

dictive of material properties in our fabrics. From the estimated regression coefficients ($\beta_m$) and dimensionality reducing basis vectors ($E_m$), the sensitivity ($S$) is computed as:

$$S = \sqrt{\left(\sum_{m=1}^{M} \beta_m E_m\right)^2} \tag{7}$$

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.

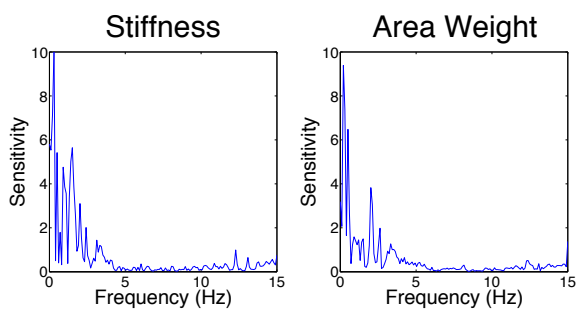


Figure 9. The sensitivity of each acoustically trained model to frequency regions in the motion spectrum. These sensitivity plots suggest that energy in the 0-5 Hz range is most predictive of a fabric's area weight and stiffness.

Figure 9 shows that frequencies in the 0Hz-5Hz range were most predictive of material properties in our fabrics. By visualizing recovered mode shapes at these frequencies, we see that they tend to contain the dominant vibration modes of our fabrics (see Figure 10), suggesting that our models use the same relationship between resonant frequencies and material properties predicted by modal analysis.

## 8. 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

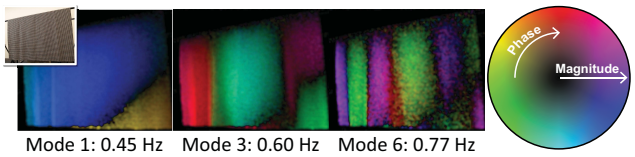


Figure 10. A sample of mode shapes extracted from predictive frequencies identified by the regression models. We see that these frequencies tend to contain dominant modes of the fabric.

active manner by recording video of an object responding to sound, or, in some cases, even passively by observing an object move naturally with 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 (aircraft, automobiles, etc). 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.

Our results suggest that the motion spectra we extract from video can be powerful features for scene understanding. The theory in Section 3 suggests that even when geometry is ambiguous, these 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 video motion spectra can be a powerful tool for reasoning about the physical properties of objects in the wild.

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. We believe that the spatial resolution and more passive nature of cameras will extend the applicability of techniques used for structural analysis to domains of interest in computer vision, such as object detection, classification and segmentation.

**Acknowledgements** This work was supported by NSF Robust Intelligence 1212849 Reconstructive Recognition, Shell Research, and Qatar Computing Research Institute. Abe and Katie were partially supported by NSF GRFP fellowships. We would also like to thank Neal Wadhwa, Gautham J. Mysore, and Danny M. Kaufman.

## 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

[2] 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. 2

[3] 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. 2, 6, 7

[4] 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. 2, 3

[5] 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. 2

[6] 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. 2, 3

[7] 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. 2, 4

[8] 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. 2, 4

[9] 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. 2, 3

[10] 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. 2, 3

[11] F. Durst, A. Melling, and J. H. Whitelaw. Principles and practice of laser-doppler anemometry. *NASA STI/Recon Technical Report A*, 76:47019, 1976. 2

[12] T. Emge and O. Buyukozturk. Remote nondestructive testing of composite-steel interface by acoustic laser vibrometry. *Materials evaluation*, 70(12):1401–1410, 2012. 2, 3

[13] 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. 3

[14] R. W. Fleming, R. O. Dror, and E. H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 2003. 2

[15] R. W. Haupt and K. D. Rolt. Standoff acoustic laser technique to locate buried land mines. *Lincoln Laboratory Journal*, 15(1):3–22, 2005. 2, 3

[16] Y.-x. Ho, M. S. Landy, and L. T. Maloney. How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, 2006. 2

[17] 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. 2

[18] S. Kawabata and M. Niwa. Fabric performance in clothing and clothing manufacture. *Journal of the Textile Institute*, 1989. 6

[19] C. Liu, L. Sharan, E. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. 2010. 2

[20] 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. 2, 3

[21] M. Rubinstein. *Analysis and Visualization of Temporal Variations in Video*. PhD thesis, Massachusetts Institute of Technology, Feb 2014. 2

[22] C. Santulli and G. Jeronimidis. Development of a method for nondestructive testing of fruits using scanning laser vibrometry (SLV). *NDT. net*, 11(10), 2006. 2

[23] A. A. Shabana. *Theory of vibration*, volume 2. Springer, 1991. 2, 3, 4

[24] 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. 2

[25] P. Shull. *Nondestructive evaluation: theory, techniques, and applications*, volume 142. CRC, 2002. 2

[26] 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. 2, 3

[27] A. Stanbridge and D. Ewins. Modal testing using a scanning laser doppler vibrometer. *Mechanical Systems and Signal Processing*, 13(2):255–270, 1999. 2

[28] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):80, 2013. 2

[29] 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. 2

[30] H. Wang, J. F. O'Brien, and R. Ramamoorthi. Data-driven elastic models for cloth: modeling and measurement. SIGGRAPH, 2011. 6

[31] 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. 2