Visual Vibrometry:
Using Cameras as Displacement Sensors

Miki Rubinstein
Google Research, Cambridge MA

(Work done while I was at MIT, Microsoft, Google)
Google Cambridge MA

MIT Stata Center (CSAIL)

Google Cambridge (~1000 people)
Google Cambridge Vision Group

- Started 2015, working at the intersection of computer vision and computer graphics.

- Group members:
  
  Bill Freeman (MIT), Ce Liu (MIT), Miki Rubinstein (MIT), Dilip Krishnan (NYU), Forrester Cole (Princeton), Inbar Mosseri, Aaron Sarna

- Hiring student summer interns
What is the area weight and elasticity of this dress?
Tiny Motions in Videos
The “Motion Microscope”
Not Just for Visualization

Recovering sound near objects

Estimating material properties (weight, stiffness)

Extracting modal frequencies, mode shapes
Vibrometry, Non-Destructive Testing
Common Vibrometry Tools

Contact Sensors, accelerometers

Laser Vibrometers
Vibrometry with Cameras

- **Small motions in videos:** hard to see, easier to analyze

- **Constraints:**
  - Textured surface,
  - Sufficient lighting,
  - Need high-speed camera for high-frequency motion

- **Advantages:**
  - Passive,
  - Non-contact,
  - Often more accessible / cheaper,
  - **Spatial resolution!**
Phase-based Motion Processing (SIGGRAPH 2013)

Complex steerable pyramid
[Simoncelli and Freeman 1995]
Complex Steerable Pyramid [Simoncelli and Freeman 1995]

- Basis functions are wavelets with even (cosine) and odd (sine) components which give local amplitude and phase.
Local Phase

- In a single subband, image is coefficients times translated copies of basis functions.
Phase-based Motion Processing (SIGGRAPH 2013)

Spatial decomposition (representation) \rightarrow Temporal filtering (per pixel)
Eulerian vs. Lagrangian processing (Fluid Dynamics)

Joseph Louis Lagrange
Track particles

Leonhard Euler
Measure changes within fixed voxels in space
Motion Magnification

Spatial decomposition (representation) → Temporal filtering (per pixel) → Amplify variation (per pixel)
Motion Magnification: Physiological Signals

Source

Motion-magnified (0.4-3 Hz, x10)

Source

Motion magnified x150

Ulnar artery
Motion Magnification: Machinery

Source (400 fps)

Motion-magnified x30
Motion Magnification: Vibration Modes

Source (20000 FPS)  480Hz (x200)  1200Hz (x400)  2400Hz (x1200)

With Justin Chen, MIT Civil Engineering

Theoretically-Derived Modal Shapes
[Wachel et al. 1990]
Sound and **Visual** Motion

Source video (2000 fps)

Motions magnified x100

90-110Hz
(male voice pitch)
Sound and Visual Motion

Riesz Pyramids for Fast Phase-Based Video Magnification (ICCP’14)
With Neal Wadhwa, Fredo Durand, William T. Freeman

CVPR 2014 Best Demo Award

Unprocessed Strobed Video:

Source video

Real-time Motion Magnified Video:

Motions magnified x250

Amplification: 0
Band: 1.15 - 1.25 Hz
Frame Rate: 20 Hz
Tone: 439 Hz

Loudspeaker
Sound-related Vibrations
Can We Recover Sound from Video?
“The Visual Microphone”
SIGGRAPH 2014
Sound Recovered from Video

Source sound in the room

Waveform

Spectrogram

Recovered sound

2200Hz (silent) video
Motion-magnified Chip Bag

Source video (2200 fps)

Sound spectrogram

Motion-magnified videos

C4 (261 Hz) x50

D4 (293 Hz) x150

E4 (330 Hz) x150

G4 (392 Hz) x120
Sound Recovered from Video

Source sound in the room

Waveform

Spectrogram

Recovered sound

2200Hz video
Remote Sound Acquisition

• Active techniques
  – Laser Microphone
Remote Sound Acquisition

- **Active** techniques
  - Laser Microphone
  - Video + speckle pattern [Zalevsky et al. 2009]
The Visual Microphone

- Active **Passive** technique to recover sound

High-speed camera
The Visual Microphone

Physical (unrelated to our processing)

Object response (A) → Camera (Projection) → Processing (B)

Unit-less (but correlated with the input pressure waves)

Assumption: Camera and object are static (any motion is due to sound)
Physical Analysis
Analysis

300 Hz at 90 dB $\rightarrow$ 0.1 micrometers motion (0.0001 millimeters!) $\rightarrow$ <0.01 pixel displacement
Video to Audio

Decomposition
- High-pass residual
- Orientation 1 (Quadrature pair)
- Orientation 2
- Scale 2
- Orientation 1
- Orientation 2
- Low-pass residual

Temporal filtering
- Amplitude
- Phase

Integration

Average local motions

“Global motion” (sound)
Combining Local Motions

- Telephone frequencies
  - 300Hz to 3.4kHz

- Speed of sound in air
  - ~343 m/s

3.4 kHz Wavelength: >10cm

300 Hz Wavelength: 1.15m
Speech Recovery

Camera

Bag of chips
Itsy Bitsy Spider

Candy wrapper (6420 fps)
Testing Visual Microphones
Rolling Shutter

https://www.flickr.com/photos/sorenragsdale/3904937619/
http://www.flickr.com/photos/boo66/5730668979/
Rolling Shutter

Image of vibrating object projected on sensor

Image read from rolling shutter sensor

**Motion and artifacts exaggerated here for illustration**
Rolling Shutter

- Slow 2D camera → Fast 1D camera!

(a) Rolling shutter in a video

(b) Converted to audio signal

Gaps (no signal)
Rolling Shutter

Input video (60 fps)

Regular SLR

Input

Recovered Sound
Rolling Shutter

Input video (60 fps)

Recovered Sound

400Hz (>6 times the frame rate!)
Range of Operation

4kHz, 400 x 480 video
Range of Operation
Wait Wait.. Don’t Tell Me!  (Aug 30 2014)
Vibration Depends on Object Properties
Estimating Material Properties with Sound and Cameras

Estimating Material Properties from Small Motions in Video, CVPR 2015

Power

Material properties (e.g. stiffness)

Frequency
Problem: Hard to Disambiguate Geometry and Material

- Motion spectra reflects a combination of material properties AND shape
Experiments

Estimating Material Properties from Small Motions in Video
Davis, Bouman, et al., CVPR 2015

Clamped Rods

Known Geometry

Hanging Fabric

Unknown but Similar Geometry
Material Properties in Clamped Rods

\[ \omega_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}} \]

- Diameter
- Elasticity
- Density
- Length
- Fundamental freq.
Material Estimation Pipeline

- Vibrating Rod
  - Exaggerated Motion
- 2500fps Video
  - Imperceptible Motion
- Brass, Copper, Aluminum, Steel
- Less than 1 Pixel

Image of setup

Speaker

Rod

Camera

Material Properties

- Elasticity
Processing

Decomposition
- High-pass residual
- Orientation 1
  (Quadrature pair)
- Orientation 2
- Orientation 1
- Orientation 2
- Low-pass residual

Temporal filtering
- Amplitude
- Phase

Integration

Average local motion spectra

Object Motion spectrum

Input

Power

Frequency
Recovered Resonant Frequencies

1\textsuperscript{st} Frequency : 6.8 Hz
2\textsuperscript{nd} Frequency : 43.4 Hz
3\textsuperscript{rd} Frequency : 121.5 Hz
4\textsuperscript{th} Frequency : 238.2 Hz
Verifying Recovered Modes: Phase Visualization

1\textsuperscript{st} Mode
6.8 Hz

2\textsuperscript{nd} Mode
43.4 Hz

3\textsuperscript{rd} Mode
121.5 Hz

4\textsuperscript{th} Mode
238.2 Hz
Material Properties from Recovered Frequencies

\[ \omega_1 = 0.1399 \frac{d}{L^2} \sqrt{\frac{E}{\rho}} \]

Diameter \downarrow \text{Elasticity} \uparrow \text{Density} \downarrow \text{Length}

2 points for each rod, for two lengths we tired:
15 and 22 inches

\[ R = 0.99 \]

Ground truth elasticity (psi)

Estimated elasticity

Compute diameter, length, density (mass/volume) with ruler and scale, solve for elasticity
Good for Objects with Known (Simple) Geometry
Motion Signals in Fabrics

Dataset of 23 fabrics of different materials, with lab measurements of area weight and stiffness
Identifying Trends in the Power Spectra

- Shifting Peak
- Increasing Area Weight
Identifying Trends in the Power Spectra

- Shifting Peak
- Increasing Area Weight

![Power Spectra Graph]

- Power Axis
- Frequency (Hz) Axis
- Color Scale: Increasing Area Weight
Learning Material Properties from Spectra

- Shifting Peak
- Regression Model
- Material Properties
  - Stiffness
  - Area Weight
Estimated Fabric Properties

Area weight

Stiffness

R = 0.96

R = 0.90

Legend:
- Lycra
- Faux Fur
- Silk
- Cotton
- Wool
- Linen
- Corduroy
- Cotton
- Fleece
- Upholstery
- Minky
- Damask Upholstery
- Flannel Backed Vinyl
- Upholstery
- Outdoor Polyester
- Wool
- Canvas
- Nylon Rip Stop
- Terry Knit
- Lycra
- Lycra
- Upholstery
Ongoing Work

Abe Davis, Justin Chen, Fredo Durand
Image-Space Modal Bases for Plausible Manipulation of Objects in Video
SIGGRAPH Asia 2015

Setup

Input Video (from SLR)

Extracted mode shapes

Interactive Modal Image Viewer

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>![6.7 Hz]</td>
<td>![16.7 Hz]</td>
<td>![28.8 Hz]</td>
</tr>
</tbody>
</table>

SIGGRAPH Asia 2015
Ongoing Work

Justen Chen, MIT Civil Engineering, with UNH and NHDOT
http://memorialbridgeproject.com/

Memorial Bridge, Portsmouth, NH

Captured video

1184 x 700 pixels at 30 fps
~80 meters from bridge

Frequency Response of Bridge Lower Deck
Ongoing Work

By Justin Chen with UNH and NHDOT
http://memorialbridgeproject.com/

Captured video (playing)
Ongoing Work

By Justin Chen with UNH and NHDOT
http://memorialbridgeproject.com/

Second vibrational mode, 2.4 Hz – 2.6 Hz, x400
Summary: Visual Vibrometry

One basic pipeline (phase-based processing)

- Magnification (visualization)
- Sound recovery
- Material property estimation
- Mode shape analysis

Graphs and images illustrating the process and results.
Miki Rubinstein, mrub@google.com
Google Research, Cambridge MA
(Work done while I was at MIT, Microsoft, Google)

Joint work with: Bill Freeman, Fredo Durand, Neal Wadhwa, Abe Davis, Katie Bouman, Justin Chen, Gautham Mysore

Project pages, code, demos:
Video magnification portal: http://people.csail.mit.edu/mrub/vidmag/
Phase-based processing: http://people.csail.mit.edu/nwadhwa/phase-video/
Material properties: http://visualvibrometry.com

TED talks by Abe Davis, Miki Rubinstein