Video Magnification

Magritte, “The Listening Room”
Imperceptible Motions and Changes
Magnified Imperceptible Motions and Changes
Approach 1: Point Tracking

Motion Magnification (SIGGRAPH 2005)

Ce Liu  Antonio Torralba  William T. Freeman  Frédo Durand  Edward H. Adelson

Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology

Following slides based on SG 2005 presentation:
http://people.csail.mit.edu/celiu/motionmag/motionmag.html
Naïve Approach

- Magnify the estimated optical flow field
- Rendering by warping
Tracking-based Motion Magnification

(a) Registered input frame
(b) Clustered trajectories of tracked features
(c) Layers of related motion and appearance

(d) Motion magnified, showing holes
(e) After texture inpainting to fill holes
(f) After user's modifications to segmentation map in (c)

Liu et al. Motion Magnification, 2005
Robust Video Registration

- Find feature points with Harris corner detector on the reference frame
- Track feature points
- Select a set of robust feature points with inlier and outlier estimation (most from the rigid background)
- Warp each frame to the reference frame with a global affine transform
Feature tracking trick 1: Adaptive Region of Support

- SSD patch matching search
- Learn adaptive region of support using expectation-maximization (EM) algorithm

Confused by occlusion!
Feature tracking trick 2: trajectory pruning

- Tracking with adaptive region of support

Nonsense at full occlusion!
Comparison

Without adaptive region of support and trajectory pruning
Cluster trajectories based on normalized complex correlation

• The similarity metric should be independent of phase and magnitude

• Normalized complex correlation

\[
\rho_{n,m} = \frac{\sum_{k} (v_{nk}^x + jv_{nk}^y)(v_{mk}^x + jv_{mk}^y)}{\sqrt{(\sum_{k}(v_{nk}^x)^2 + (v_{nk}^y)^2)(\sum_{k}(v_{mk}^x)^2 + (v_{mk}^y)^2)}}
\]
Spectral Clustering

Affinity matrix

Clustering

Reordering of affinity matrix
Clustering Results
From Sparse Feature Points to Dense Optical Flow Field

Interpolate dense optical flow field using locally weighted linear regression
Motion Layer Assignment

• Assign each pixel to a motion cluster layer, using four cues:
  – **Motion likelihood**—consistency of pixel’s intensity if it moves with the motion of a given layer (dense optical flow field)
  – **Color likelihood**—consistency of the color in a layer
  – **Spatial connectivity**—adjacent pixels favored to belong the same group
  – **Temporal coherence**—label assignment stays constant over time

• Energy minimization using graph cuts
Segmentation Results

Two additional layers: static background and outlier
Layered Motion Representation for Motion Processing

- Background
- Layer 1
- Layer 2

Layer mask

Occluding layers

Appearance for each layer before texture filling-in

Appearance for each layer after texture filling-in
Discussion of point tracking approach

- Good: applies to any motion
- Bad: requires accurate point tracking, clustering and texture synthesis, so likely to fail
Approach 2: pixelwise processing

Eulerian Video Magnification for Revealing Subtle Changes in the World
Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Fredo Durand, William T. Freeman
ACM Transactions on Graphics, Volume 31, Number 4 (Proc. SIGGRAPH) 2012

Phase-based Video Motion Processing
Neal Wadhwa, Michael Rubinstein, Fredo Durand, William T. Freeman
ACM Transactions on Graphics, Volume 32, Number 4 (Proc. SIGGRAPH) 2013

Following slides based on Siggraph presentations:
http://people.csail.mit.edu/mrub/vidmag/
http://people.csail.mit.edu/nwadhwa/phase-video/
Lagrangian and Eulerian Perspectives: Fluid Dynamics

Lagrangian

Eulerian
Eulerian Perspective: Videos

- Each pixel is processed independently
- Treat each pixel as a time series and apply signal processing to it
Method Overview

Spatial Decomposition

Temporal filtering

Reconstruction

Laplacian Pyramid

Bandpass filter intensity at each pixel over time

Amplify bandpassed signal and add back to original
Subtle Color Variations

- The face gets slightly redder when blood flows
- Unfortunately usually below the per pixel noise level
Subtle Color Variations

1. Average spatially to overcome sensor and quantization noise

Input frame

Spatially averaged luminance trace

Luminance trace (zero mean) pulses

Spatially averaged luminance trace
Amplifying Subtle Color Variations

2. Filter temporally to extract the signal of interest

Spatially averaged luminance trace

Temporal filter

=

Temporally bandpassed trace
Color Amplification Results

Source

Color-amplified (x100)
0.83-1 Hz (50-60 bpm)
Heart Rate Extraction

Peak detection

Temporally bandpassed trace (one pixel)

Pulse locations
Heart Rate Extraction

Thanks to Dr. Donna Brezinski and the Winchester Hospital staff 2.33-2.67 Hz (140-160 bpm)
Why It Amplifies Motion
Relating Temporal and Spatial Changes

Courtesy of Lili Sun
When Does It Break?

- Clipped

![Graph showing intensity versus space with clipped regions indicated.](Image)
When does it break?
Motion Magnification Artifacts

Source

Motion-magnified (3.6-6.2 Hz, x60)
When does it break?
Method Overview

Spatial Decomposition

Temporal filtering

Reconstruction

Laplacian Pyramid

Bandpass filter intensity at each pixel over time

Amplify bandpassed signal and add back to original

\[ \alpha_1, \alpha_2, \ldots, \alpha_n \]

\[ \sum \]

40
Scale-varying Amplification

- The amplification is more accurate for low spatial frequencies
  - Images are smoother
  - Motions are smaller

- Use the desired $\alpha$ for lower spatial frequencies, and attenuate for the higher spatial frequencies

\[
(1 + \alpha)\delta(t) < \frac{\lambda}{8}
\]
Motion Magnification Results

Source

Motion-amplified (x10)
Source (500 fps)

* The subtle flickering is due to the light source resonating at 120Hz

Motion magnified x150 (30-50Hz)
Source (300 FPS)  Motion magnification (x50)  Motion magnification (x50) Large motions unmagnified
Ground Truth Validation

- Induce motion (with hammer)
- Record with accelerometer
Ground Truth Validation
Neck Skin Vibrations

Source (2 KHz)
Source (2 KHz)  

100 Hz Amplified x100  

Fundamental frequency: ~100Hz
The Visual Microphone: Passive Recovery of Sound from Video

Abe Davis    Michael Rubinstein    Neal Wadhwa
Gautham Mysore    Fredo Durand    William T. Freeman

(slides adopted from Siggraph presentation)
Remote Sound Recovery
Can We Recover Sound from Video?
Sound and Motion

Source: mediacollege.com
Analysis

Low Volume

Long Distance
Analysis
Audio

- MaryMIDI_wav.wmv
- MaryMIDI_Chips_wav.wmv
- MaryMIDI_Plant_wav.wmv
Sound Recovered from Video

Source sound in the room

Waveform

Spectrogram

Recovered sound

2200Hz video
Sound Recovered from Video

Source sound in the room

Waveform

Spectrogram

Recovered sound

2200Hz video
Sound Recovered from Video

Source sound in the room

Waveform
Spectrogram

Recovered sound

(small patch on the chip bag)

20 kHz video
The Visual Microphone
The Visual Microphone
The Visual Microphone

Input

Air pressure (Pa) → Object response (A) → Object motion (mm disp.)
The Visual Microphone

Input

Air pressure (Pa)

Object response (A)

Object motion (mm disp.)

Camera (Projection)

Video (pixels)

[Diagram showing the process of converting air pressure into visual output]
The Visual Microphone

Air pressure (Pa) → Object response (A) → Object motion (mm disp.) → Video (pixels) → Recovered Signal (~Pa)

Input

Processing (B)

Video
The Visual Microphone

- **Very subtle motion**
  - Micrometers or smaller
  - Thousandths of a pixel

- **Averaging**
Combining Local Motions

- Telephone frequencies
  - 300Hz to 3.4kHz

- Speed of sound in air
  - ~343 m/s
Combining Local Motions

• Telephone frequencies
  – 300Hz to 3.4kHz

• Speed of sound in air
  – ~343 m/s

3.4 kHz
Wavelength:
>10cm
Combining Local Motions

300 Hz Wavelength:
1.15m
Processing

- Extract local motion signals
- Average and Align
- Post-process
The Visual Microphone

Input

Air pressure (Pa) → Object motion (mm disp.) → Video (pixels) → Recovered Signal (~Pa)
Some materials are better microphones than others
The Visual Microphone: Passive Recovery of Sound from Video

Abe Davis
Michael Rubinstein
Neal Wadhwa
Gautham J. Mysore
Fredo Durand
William T. Freeman