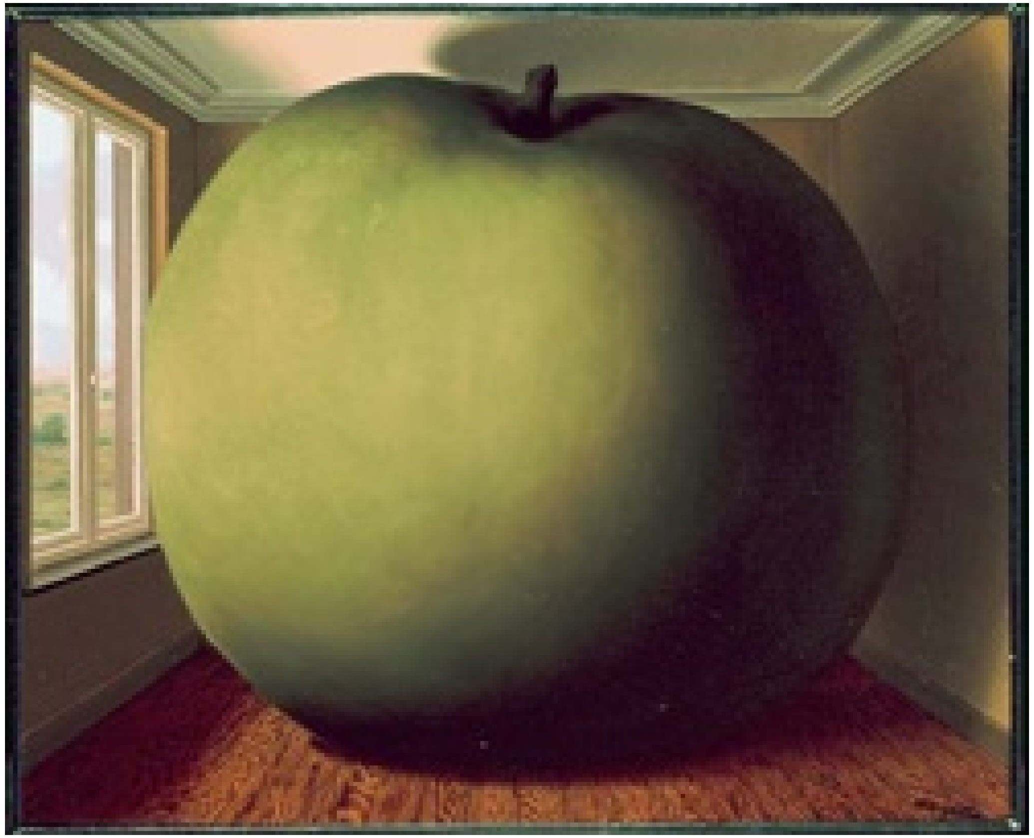
Video Magnification



Magritte, “The Listening Room”

Computational Photography

Derek Hoiem, University of Illinois

This Class

1. Video Magnification
   - Lagrangian (point tracking) approach
   - Eulerian (signal within a pixel) approach

2. Video Microphone

Imperceptible Motions and Changes

[Liu et al. 2005]

[Wu et al. 2012]

MAGNIFIED Imperceptible Motions and Changes






[Liu et al. 2005]

[Wu et al. 2012]

Motion Magnification

Goal: exaggerate selected motions

Ideas?

Approach 1: Point Tracking

Motion Magnification (SIGGRAPH 2005)

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

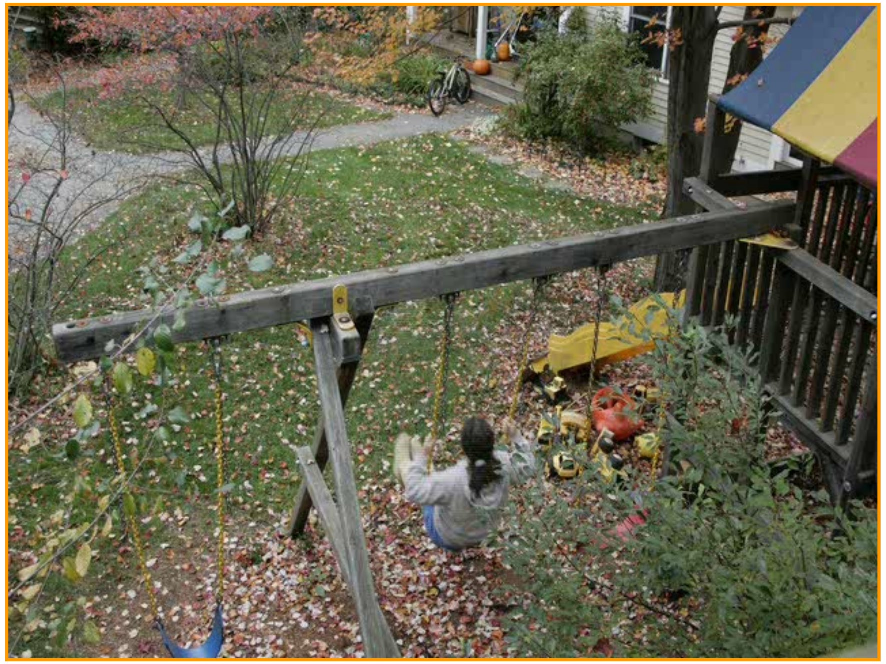
Computer Science and Artificial Intelligence Laboratory

Massachusetts Institute of Technology

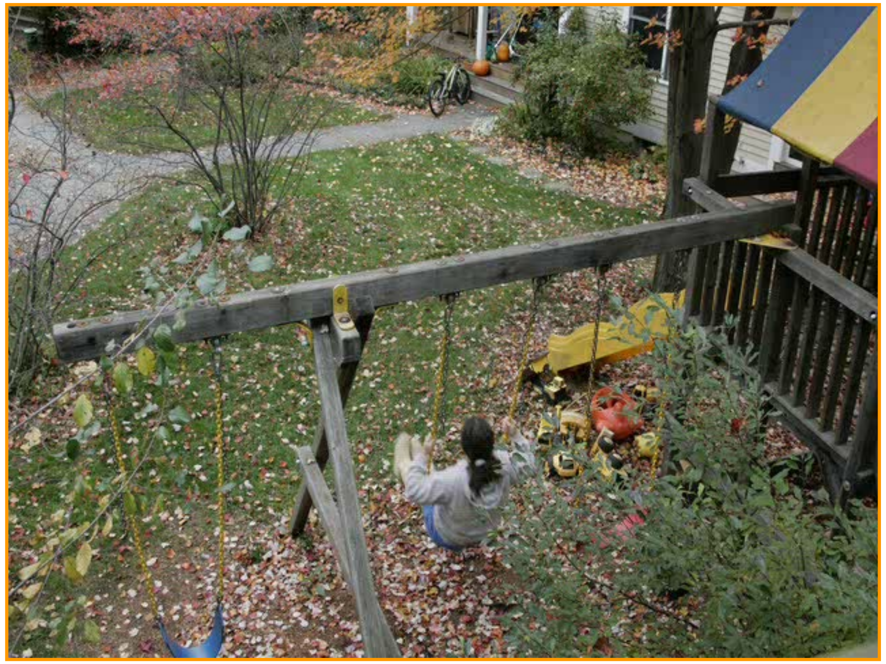
Following slides based on SG 2005 presentation:
http://people.csail.mit.edu/celiu/motionmag/motionmag.html

Naïve Approach

- Magnify the estimated optical flow field
- Rendering by warping

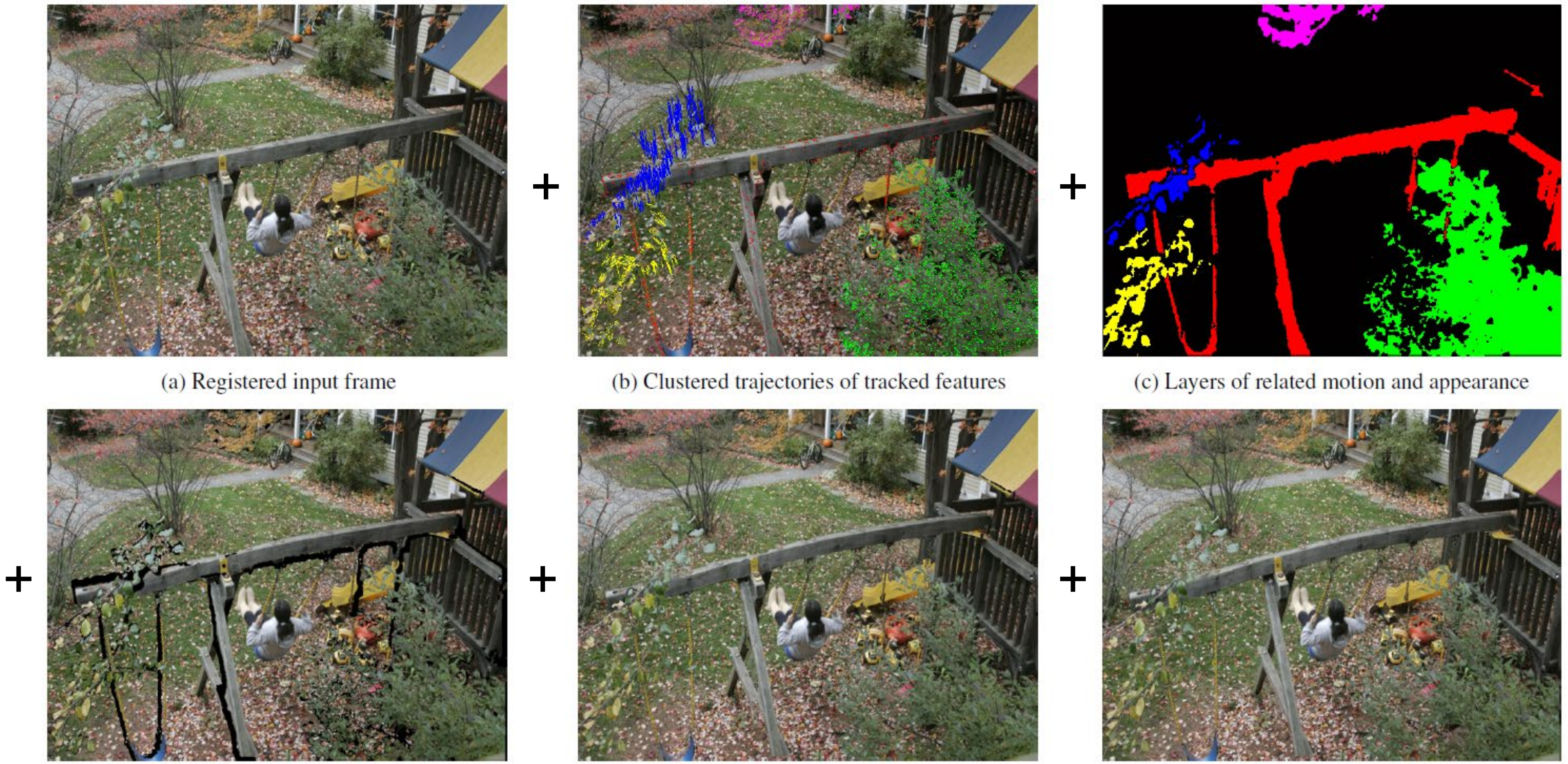


Original sequence



Magnified by naïve approach

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 in-painting to fill holes
(f) After user's modifications to segmentation map in (c)

Liu et al. *Motion Magnification*, 2005

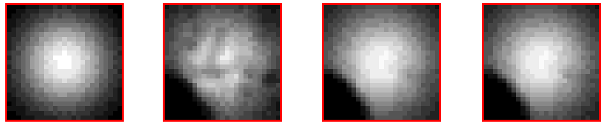
Robust Video Registration

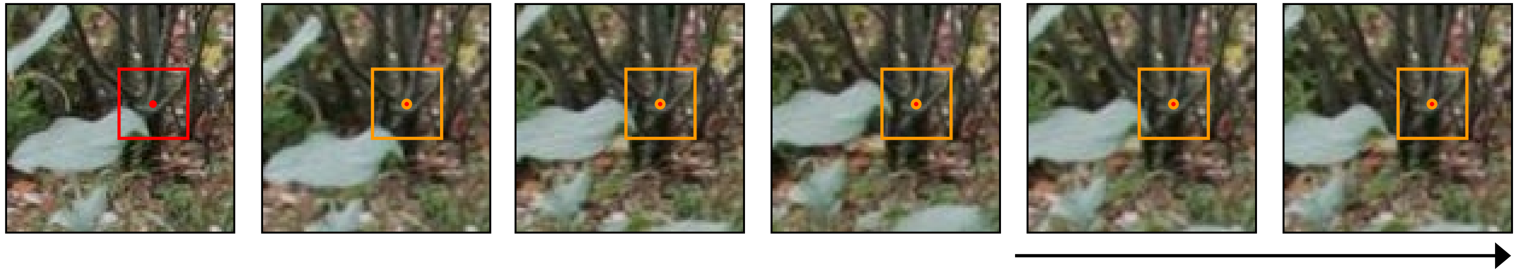
- Find feature points with Harris corner detector on the reference frame
- Track feature points
- Select a set of robust feature points with inlier and outlier estimation (most from the rigid background)
- Warp each frame to the reference frame with a global affine transform

Feature tracking trick 1: Adaptive Region of Support

- SSD patch matching search

- Learn adaptive region of support using expectation-maximization (EM) algorithm



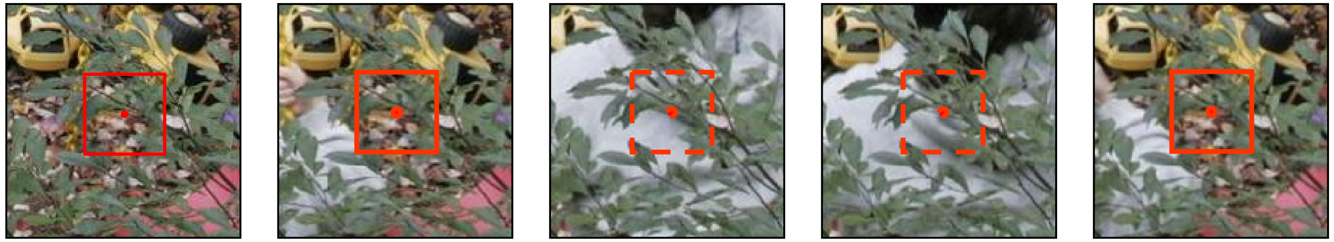


Feature tracking trick 2: trajectory pruning

- Tracking with adaptive region of support

Nonsense at full occlusion!

- Outlier detection and removal by interpolation



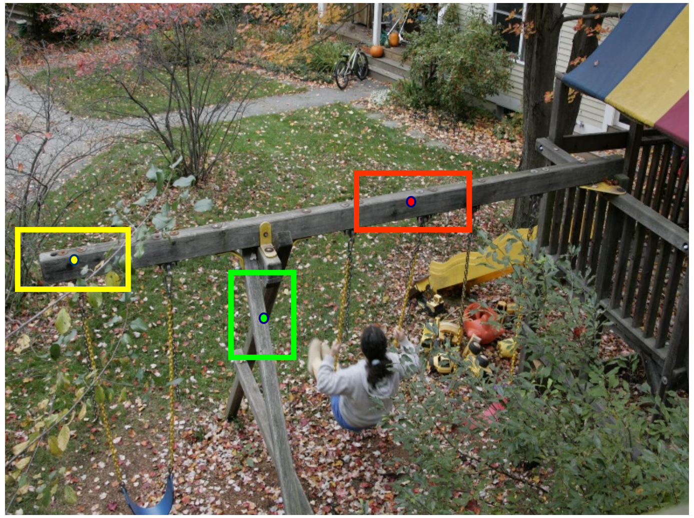
Comparison

Without With adaptive region of support and trajectory pruning

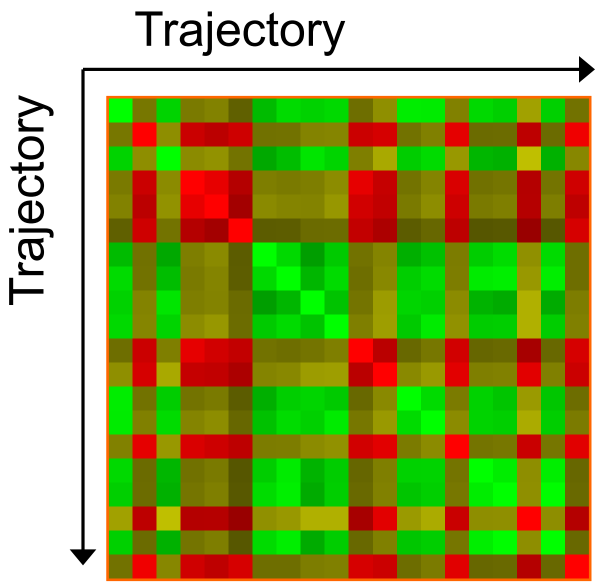
Cluster trajectories based on normalized complex correlation

- The similarity metric should be independent of phase and magnitude
- Normalized complex correlation

$$S(C_1, C_2) = \frac{|\sum_t C_1(t)\overline{C}_2(t)|^2}{\sqrt{\sum_t C_1(t)\overline{C}_1(t)}\sqrt{\sum_t C_2(t)\overline{C}_2(t)}}$$



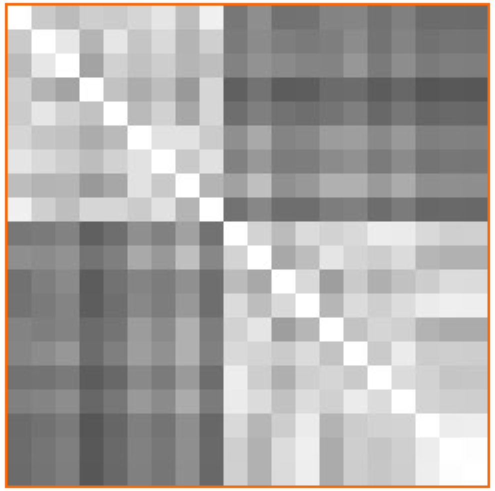
Spectral Clustering



Affinity matrix

Two clusters

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

Flow vectors of clustered sparse feature points

Dense optical flow field of cluster 2 (swing)

(leaves)

Cluster 1: leaves

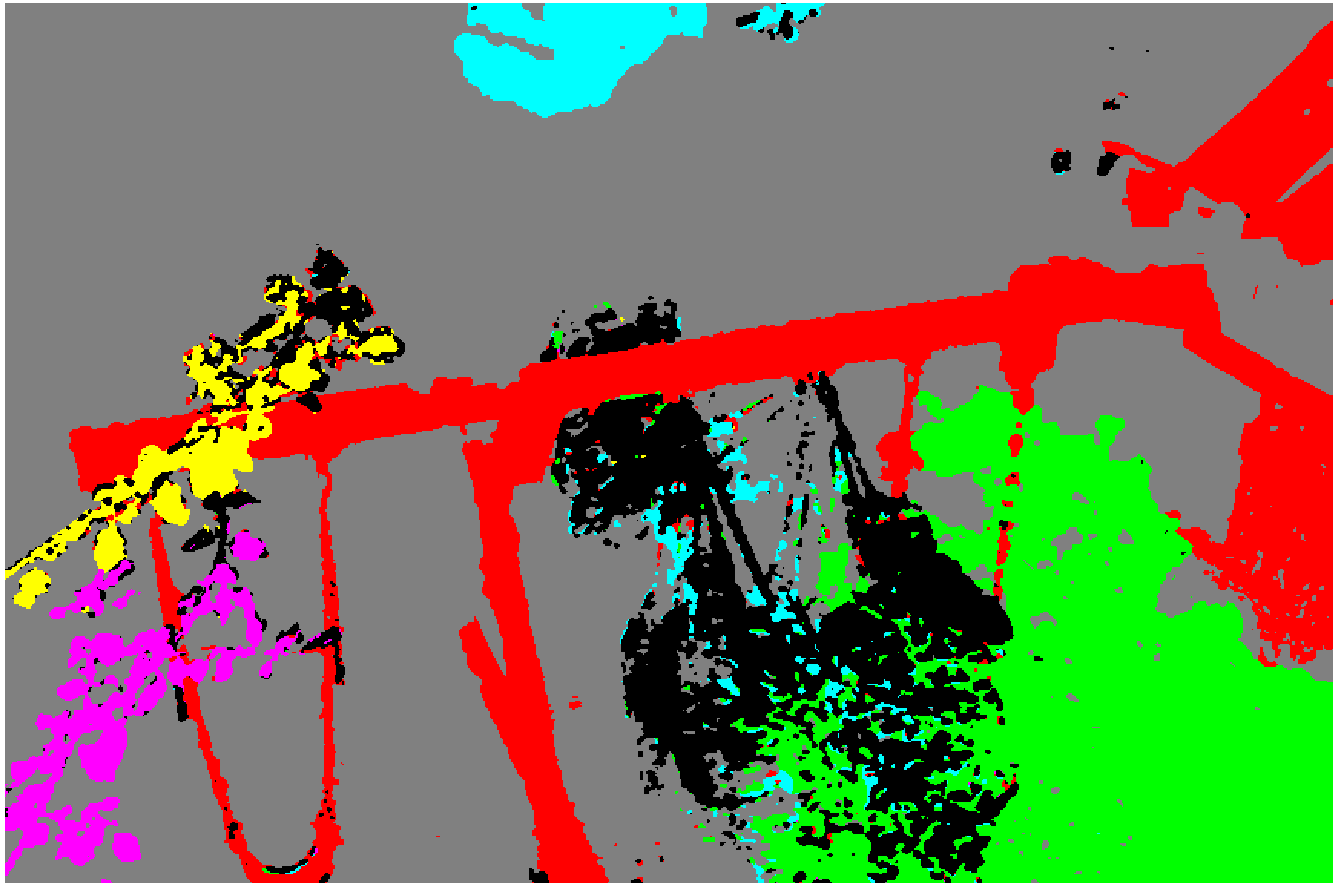
Cluster 2: swing

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

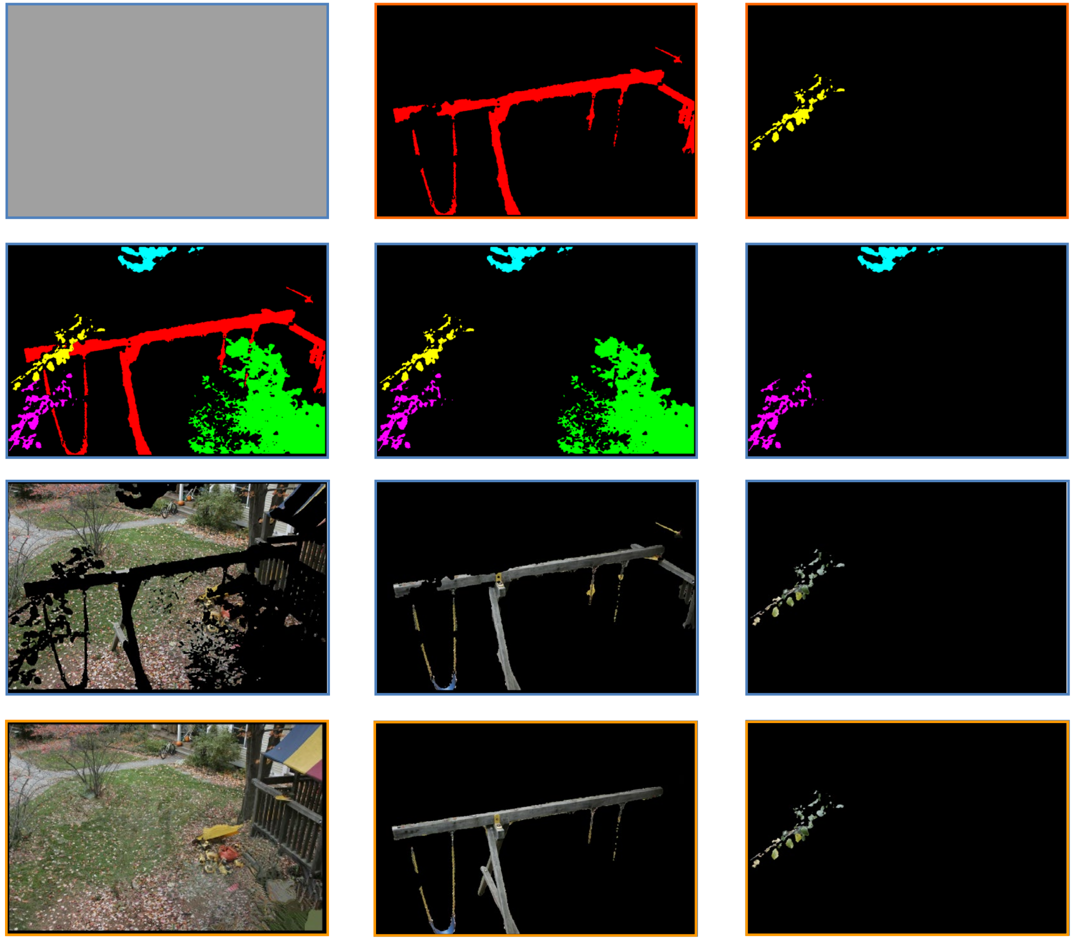
Segmentation Results

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

after user editing


Original Sequence

Discussion of point tracking approach

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

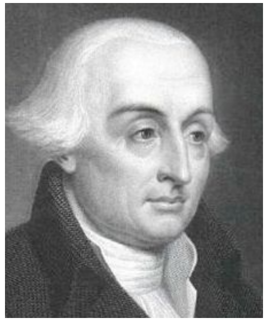
Approach 2: pixelwise processing

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

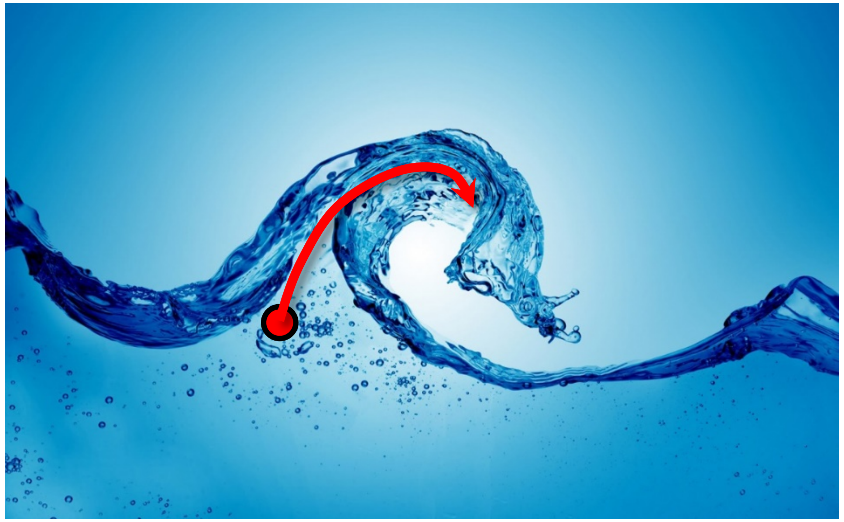
Following slides based on Siggraph presentations:
http://people.csail.mit.edu/mrub/vidmag/
http://people.csail.mit.edu/nwadhwa/phase-video/

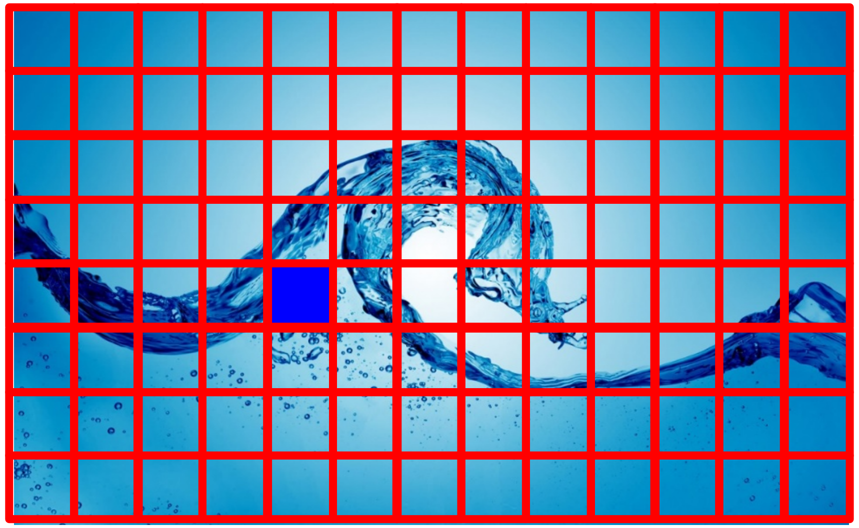
Lagrangian and Eulerian Perspectives: Fluid Dynamics



Lagrangian

Eulerian

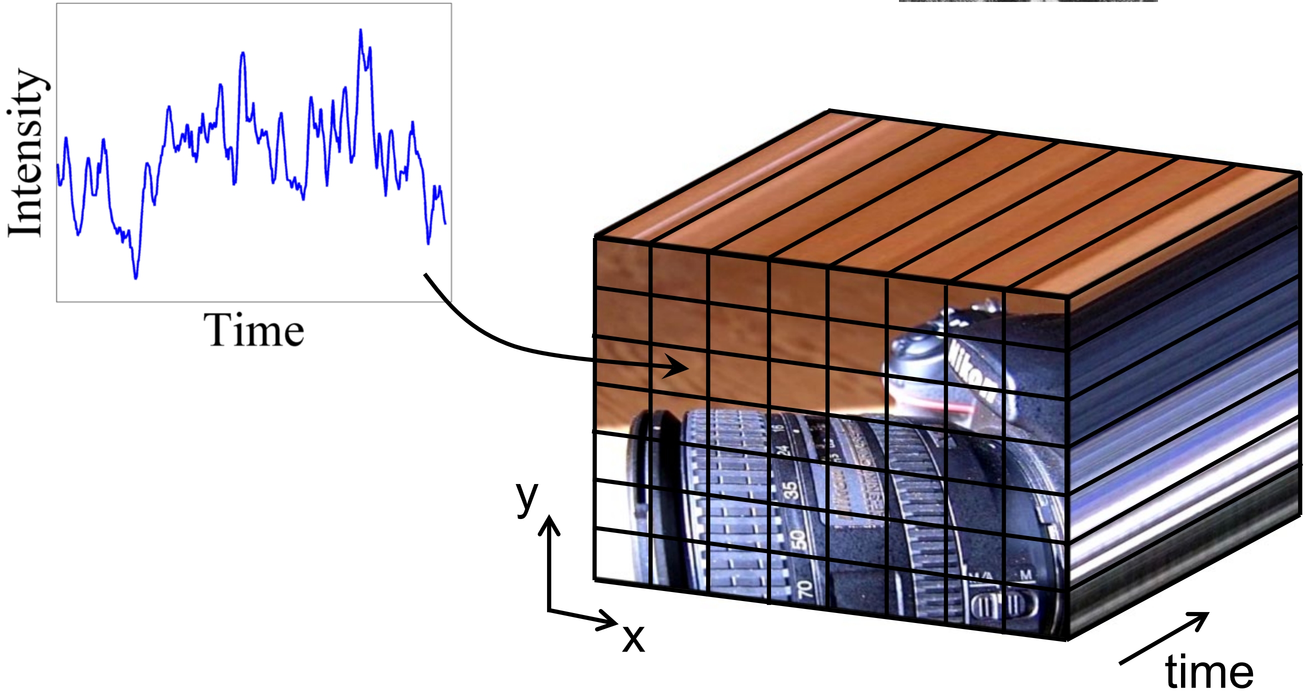




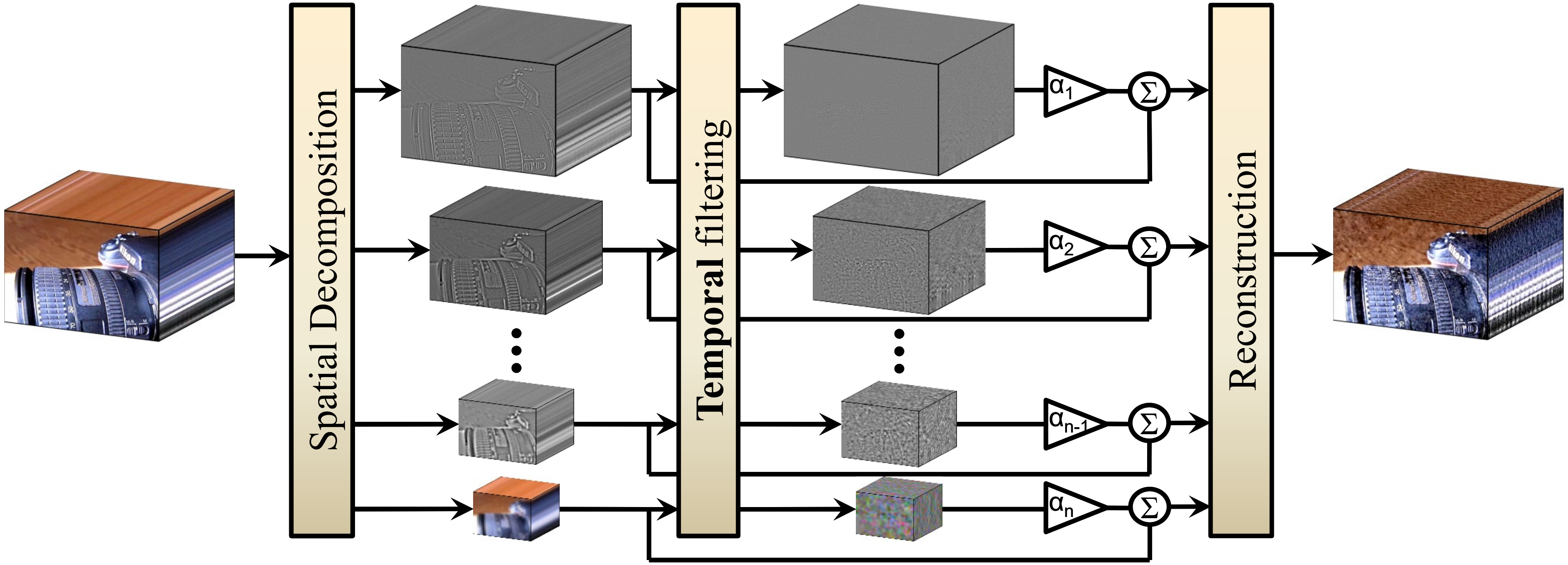
Eulerian Perspective: Videos

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

Eulerian



Method Overview



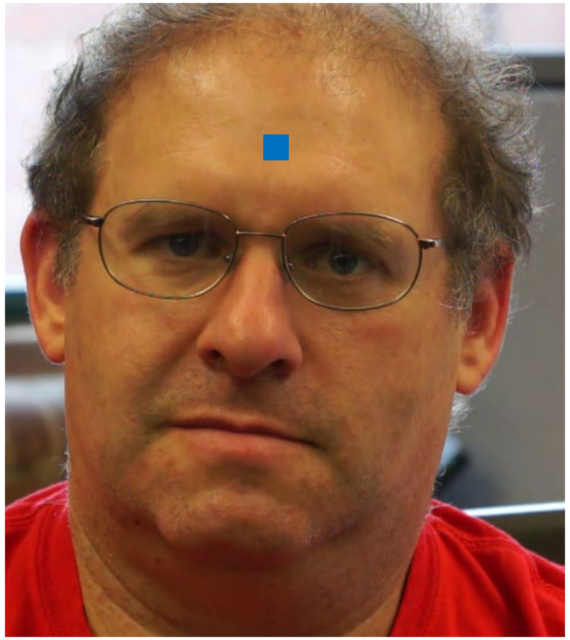
Laplacian Pyramid

Bandpass filter intensity at each pixel over time

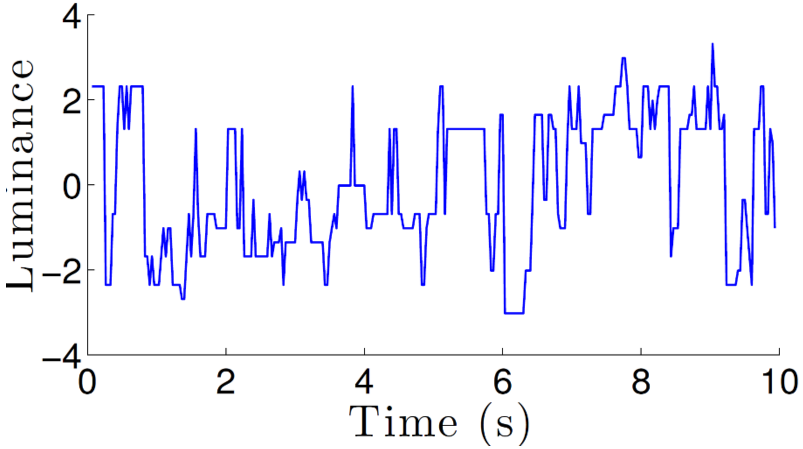
Amplify bandpassed signal and add back to original

Subtle Color Variations

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



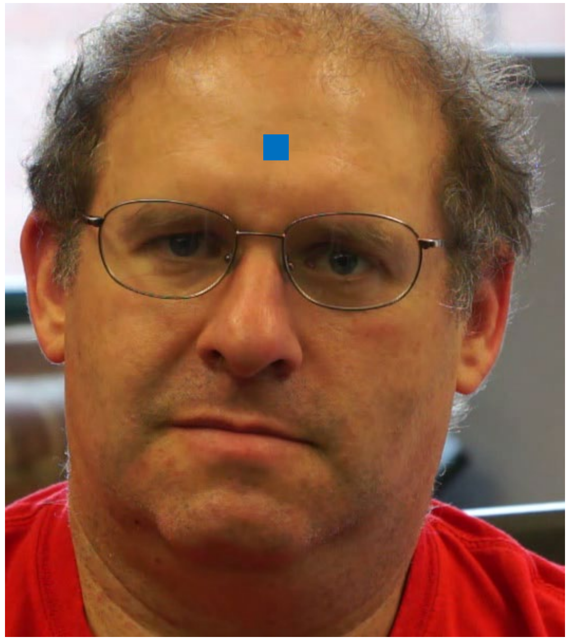
Input frame



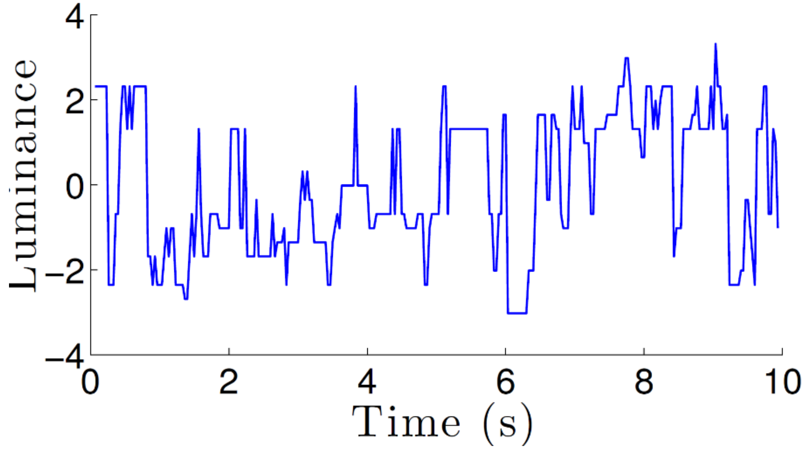
Luminance trace (zero mean)

Subtle Color Variations

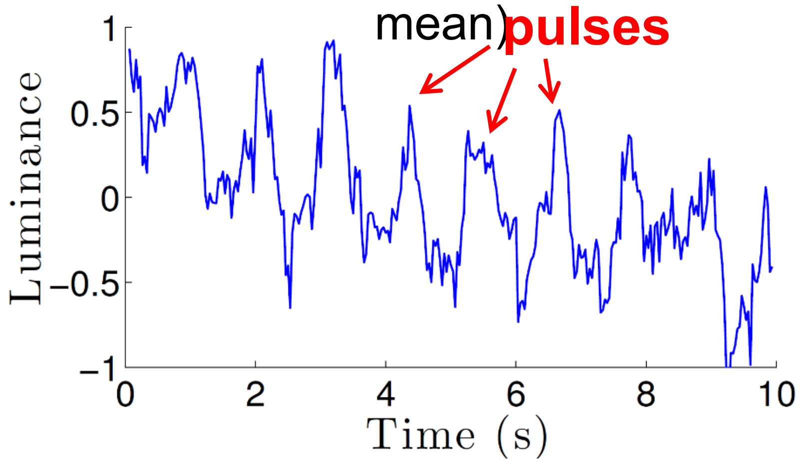
1. Average spatially to overcome sensor and quantization noise



Input frame



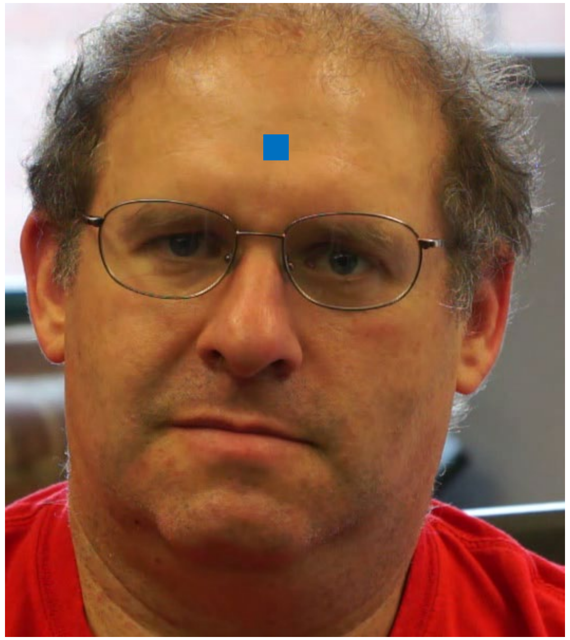
Luminance trace (zero mean)



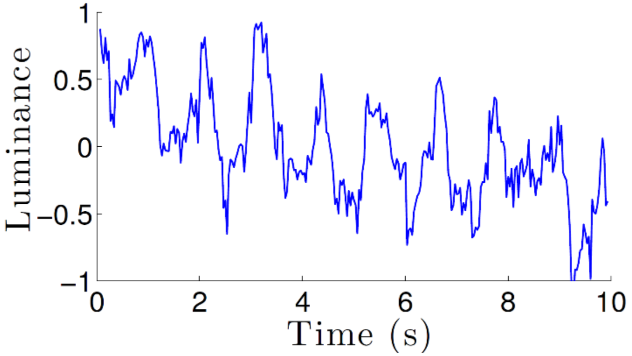
Spatially averaged luminance trace

Amplifying Subtle Color Variations

2. Filter temporally to extract the signal of interest

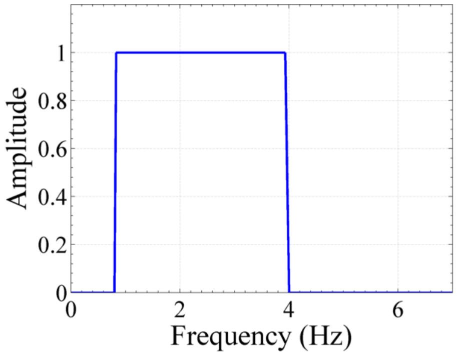


Input frame



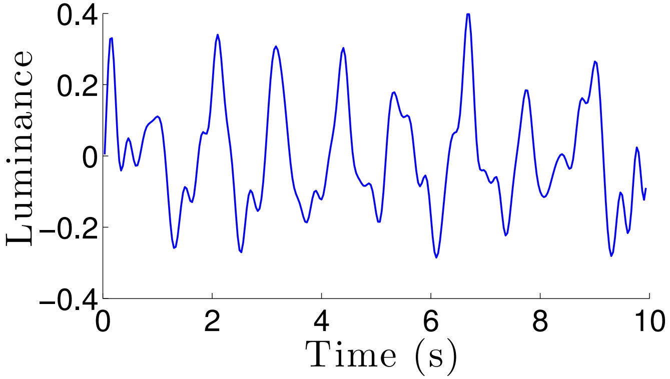
Spatially averaged luminance trace

⊗



Temporal filter

=



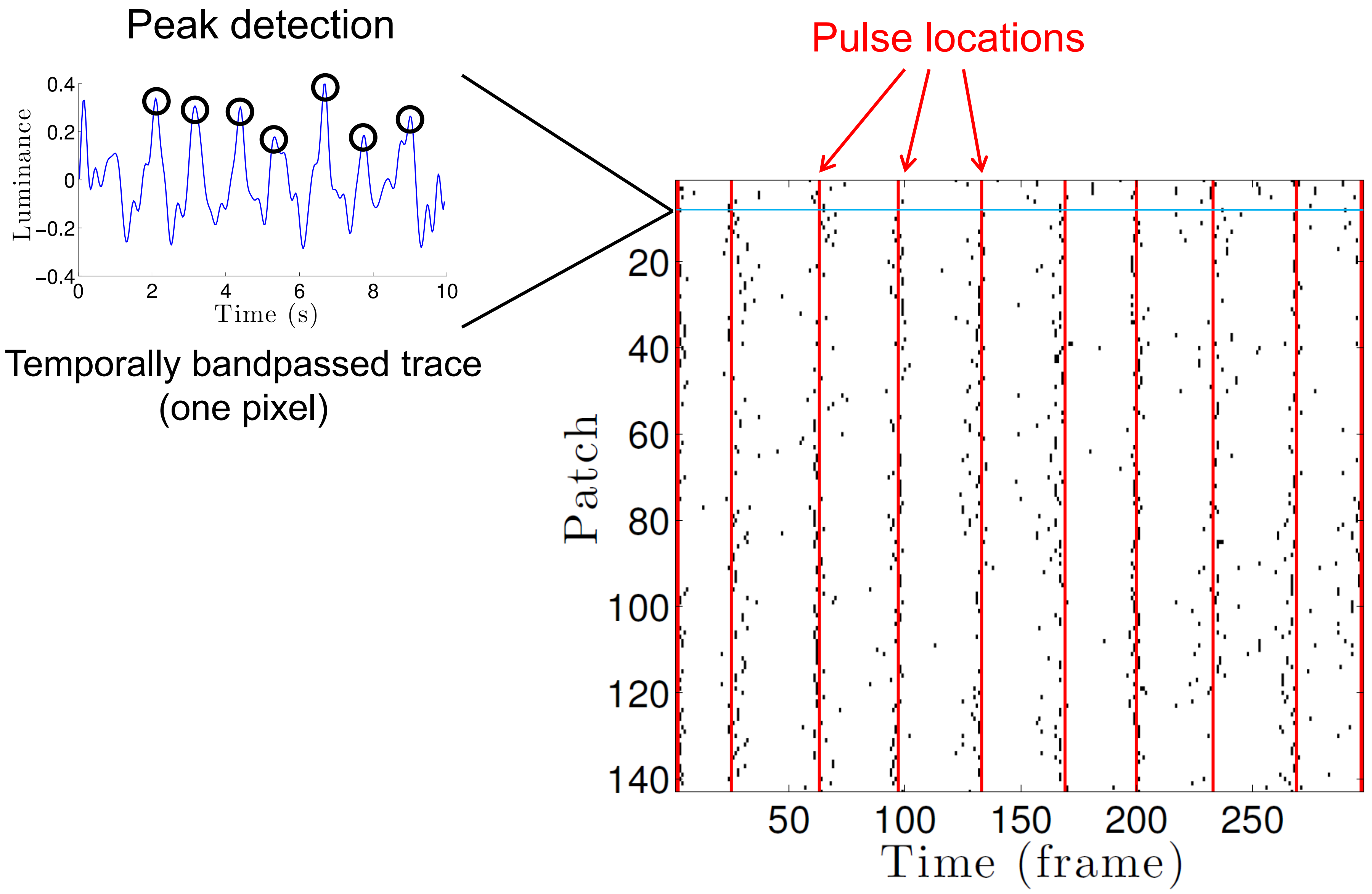
Temporally bandpassed trace

Color Amplification Results

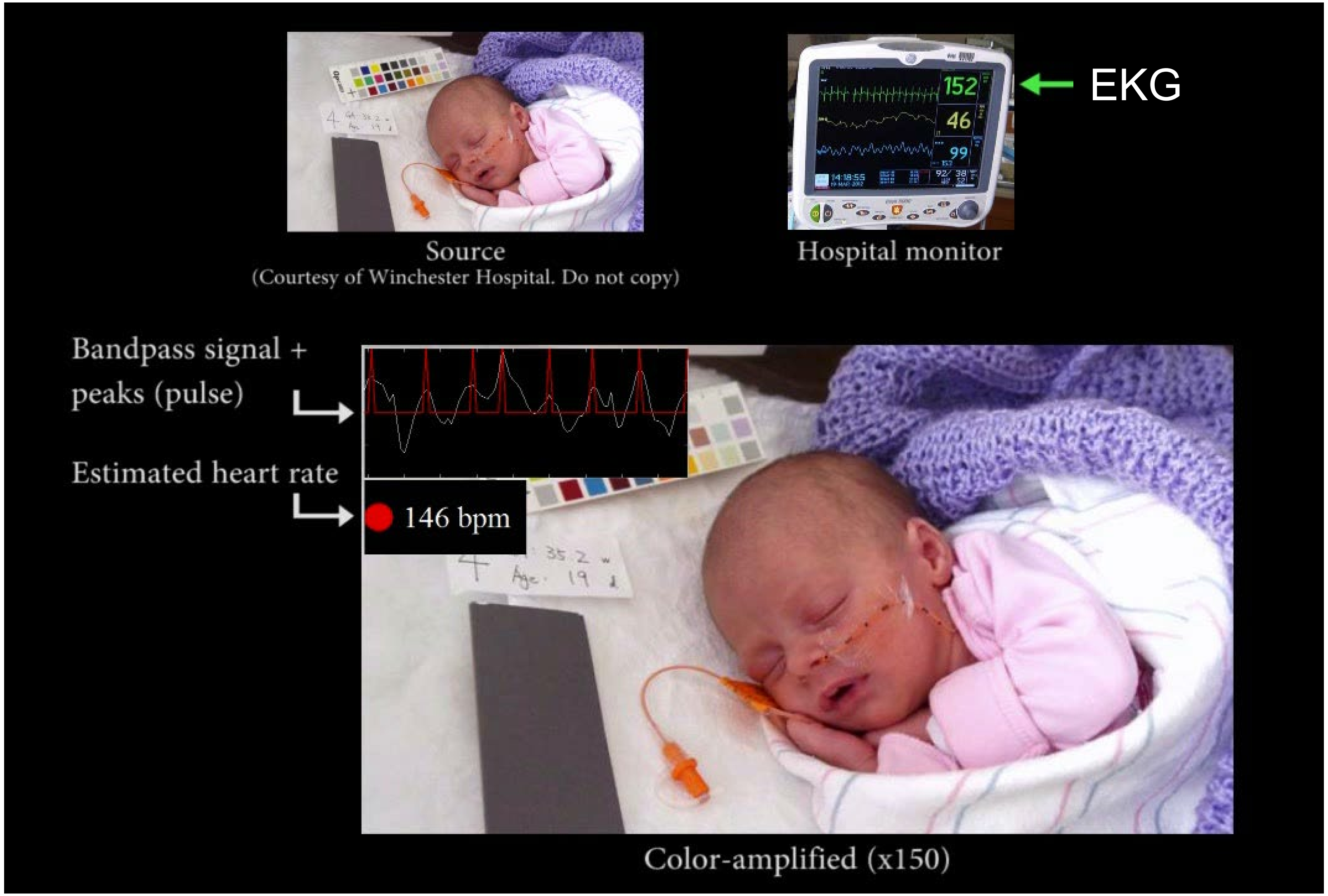
Source

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

Heart Rate Extraction

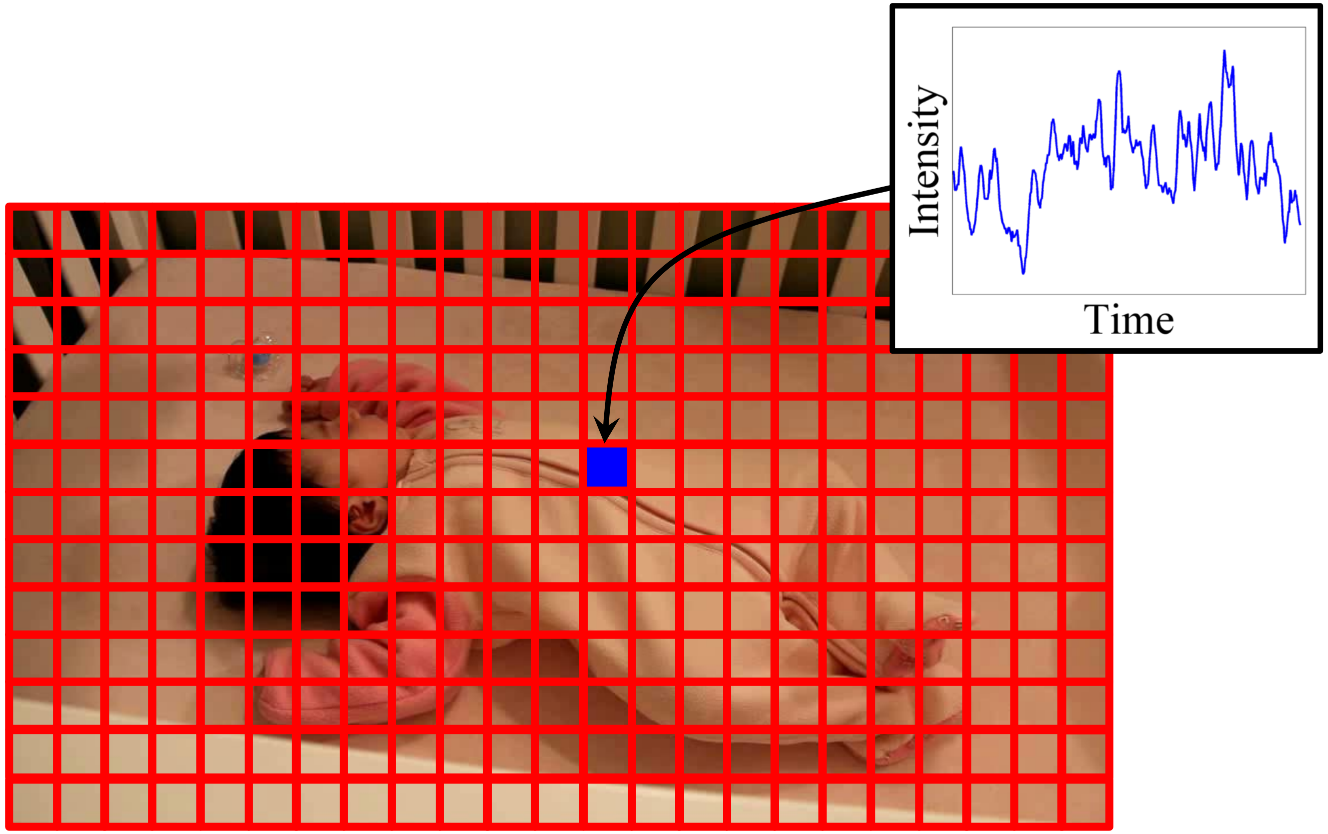


Heart Rate Extraction

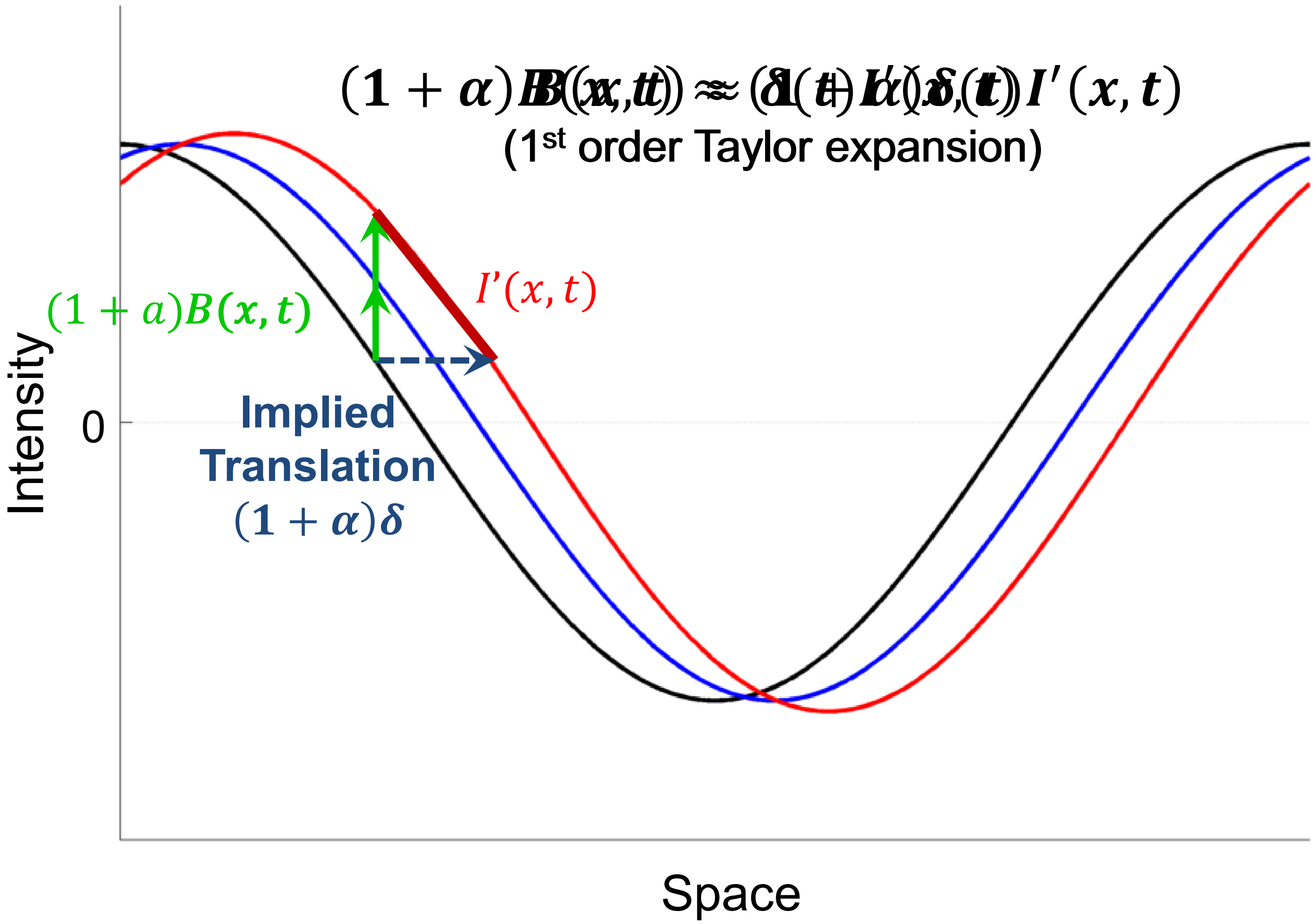


Thanks to Dr. Donna Brezinski and the Winchester Hospital staff 2.33-2.67 Hz (140-160 bpm)

Why It Amplifies Motion

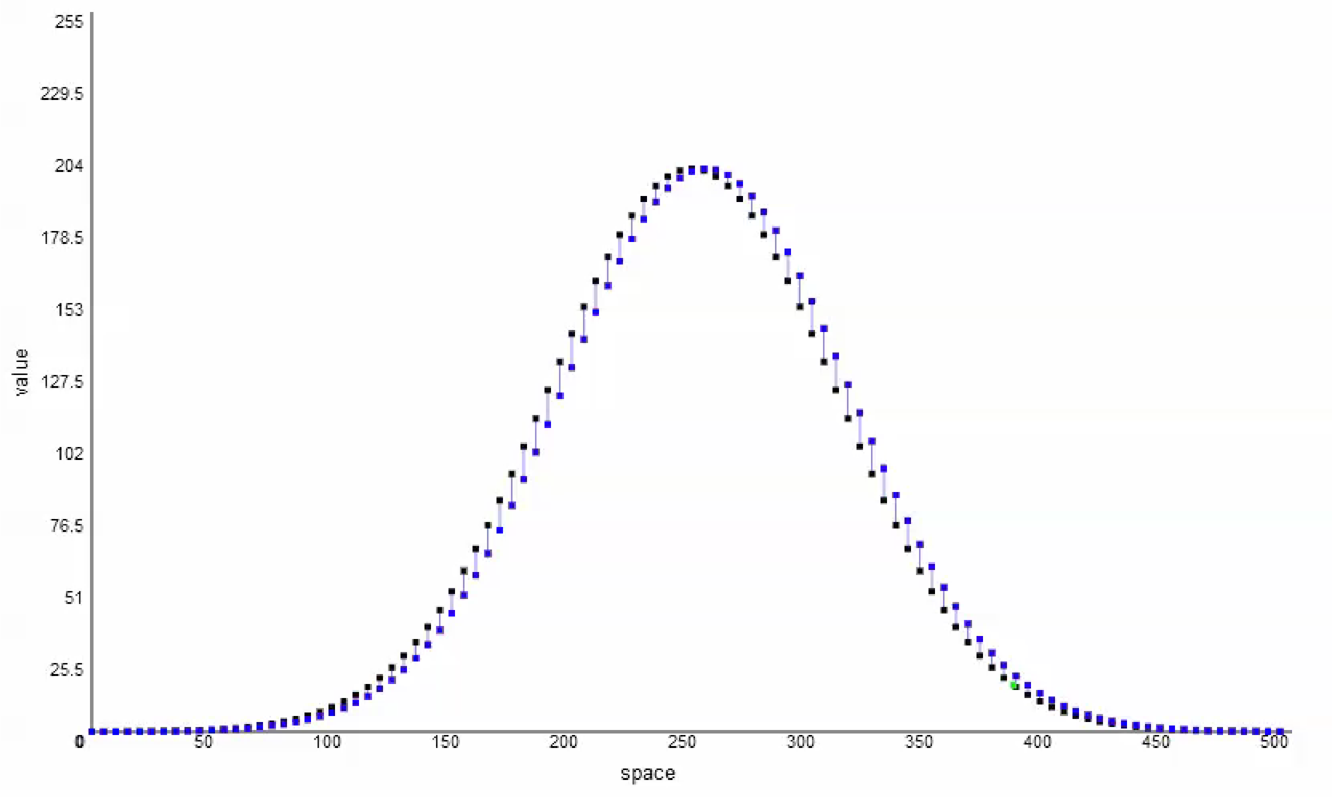


Relating Temporal and Spatial Changes



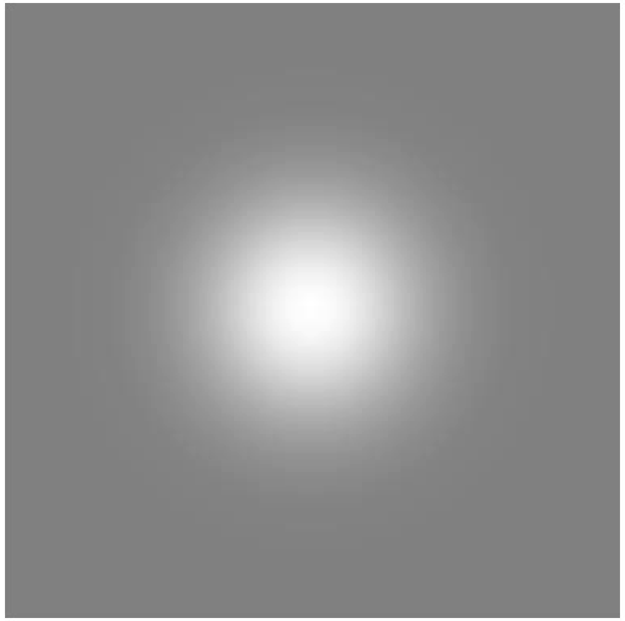
Relating Temporal and Spatial Changes

- Signal at time t
- Signal at time $t + 1$
- Motion-magnified



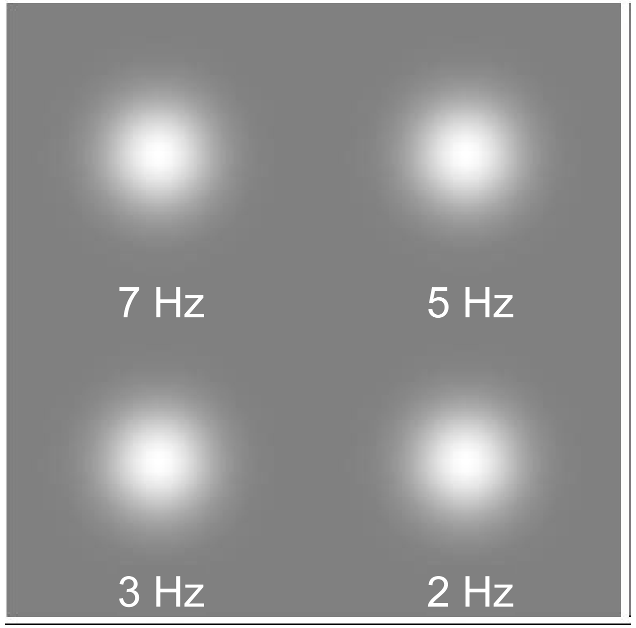
Courtesy of Lili Sun

Synthetic 2D Example



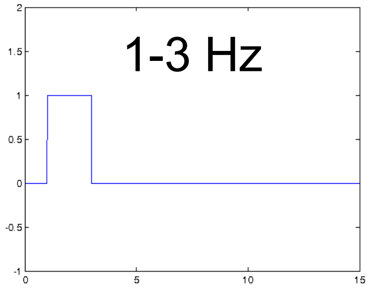
Source

Selective Motion Magnification

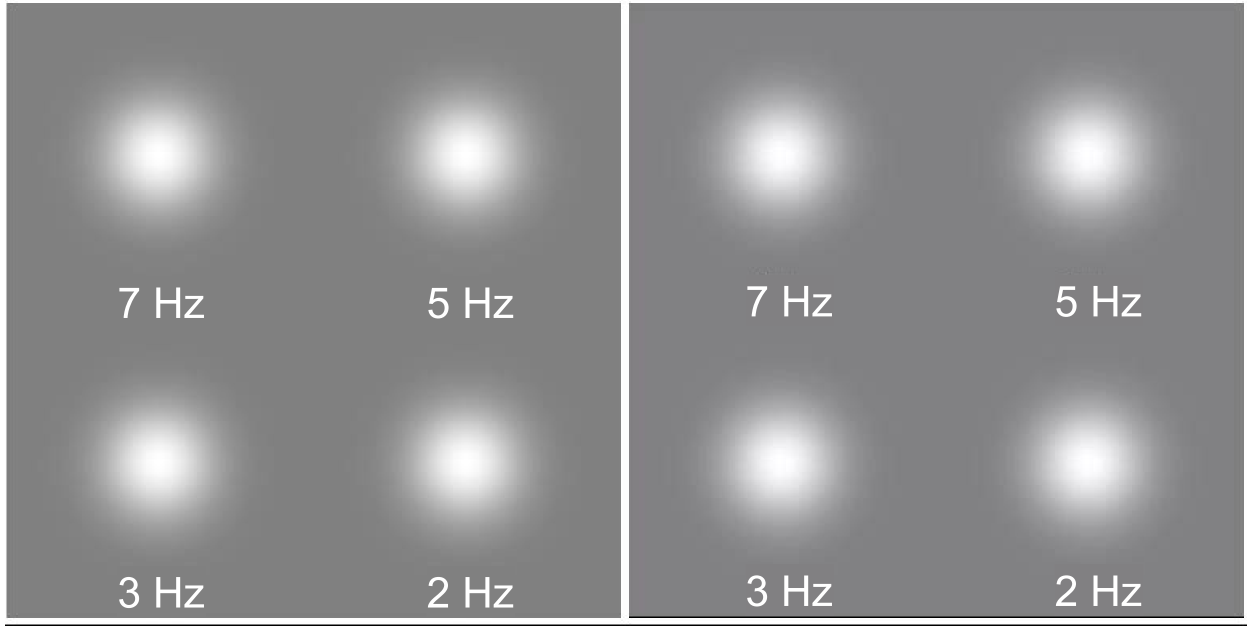


Source
(Single video with 4 blobs)

Temporal filter:



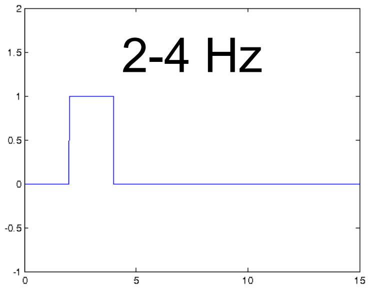
Selective Motion Magnification



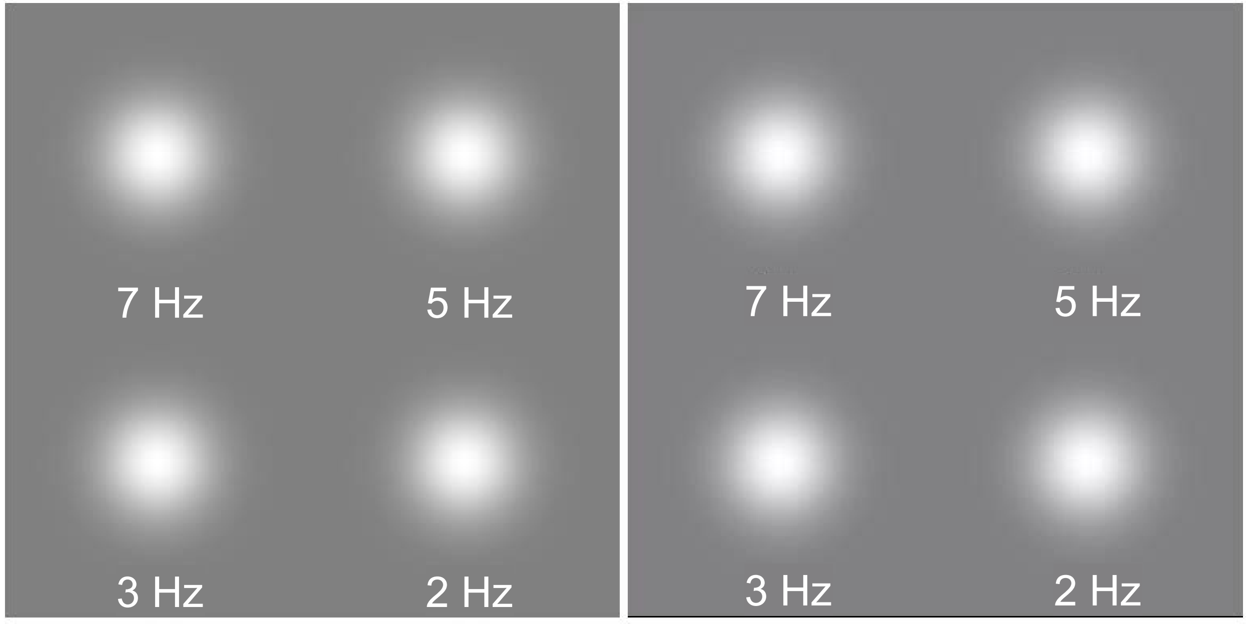
Source
(Single video with 4 blobs)

Motion-magnified (3 Hz)

Temporal filter:



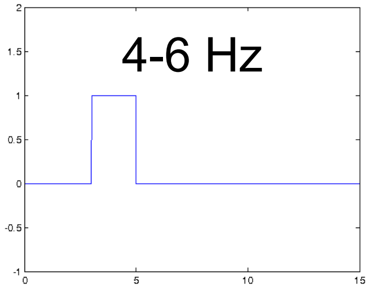
Selective Motion Magnification



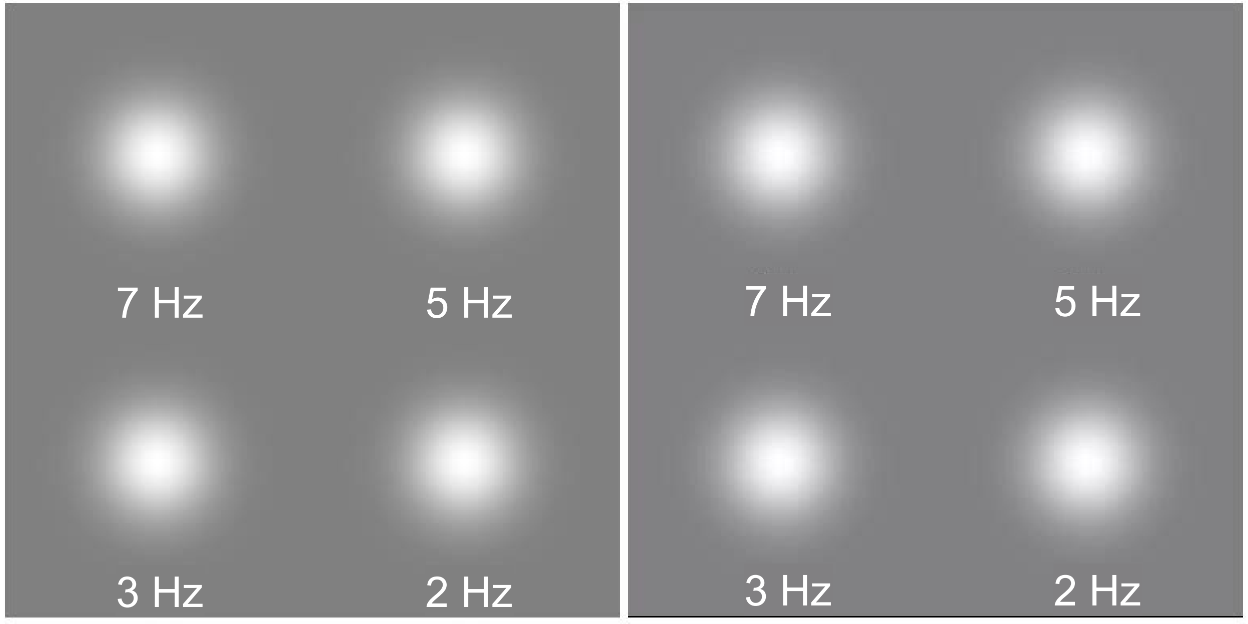
Source
(Single video with 4 blobs)

Motion-magnified (5 Hz)

Temporal filter:



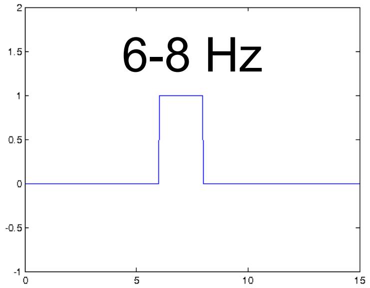
Selective Motion Magnification



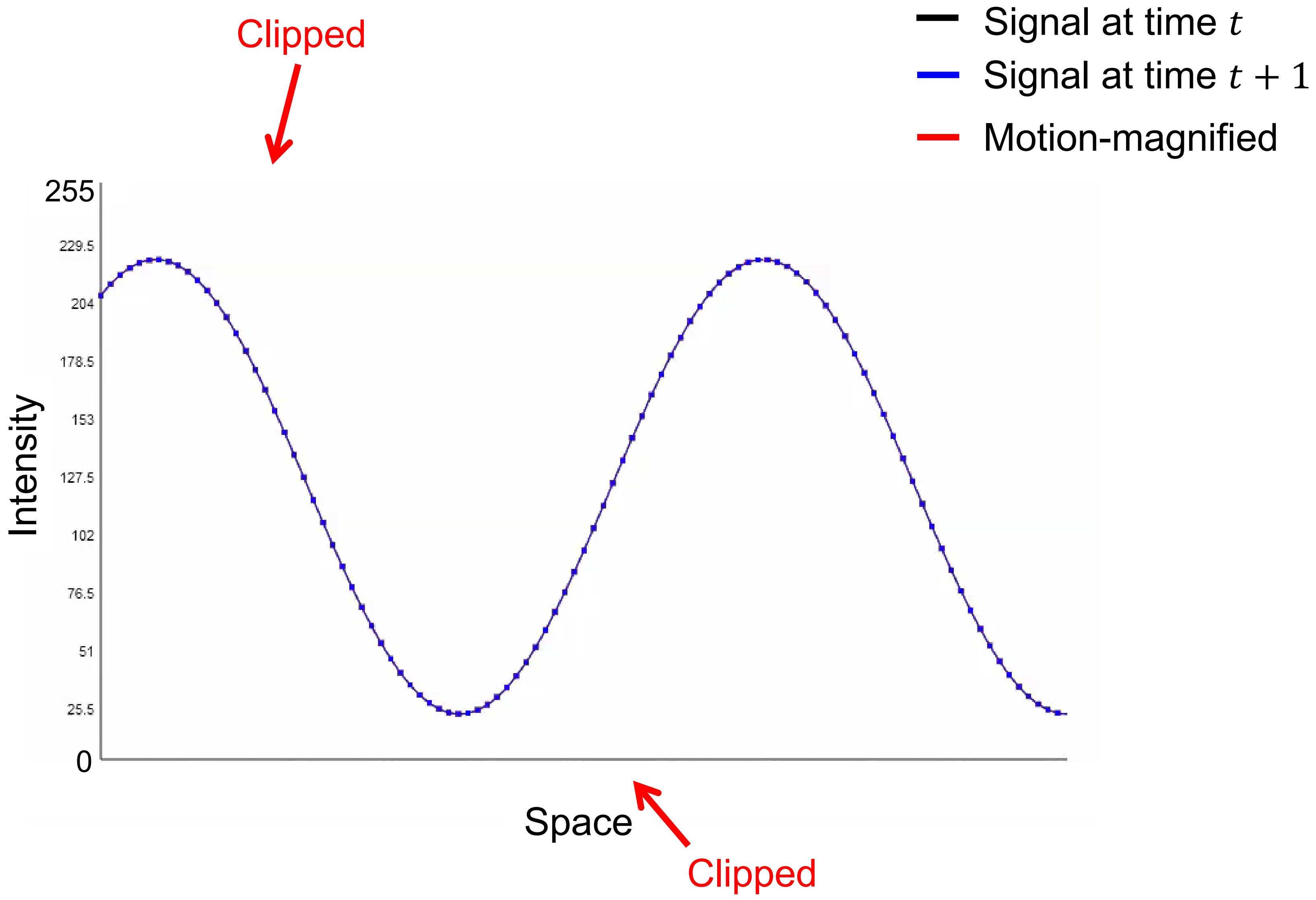
Source
(Single video with 4 blobs)

Motion-magnified (7 Hz)

Temporal filter:



When Does It Break?



Motion Magnification Artifacts

Artifact

Source

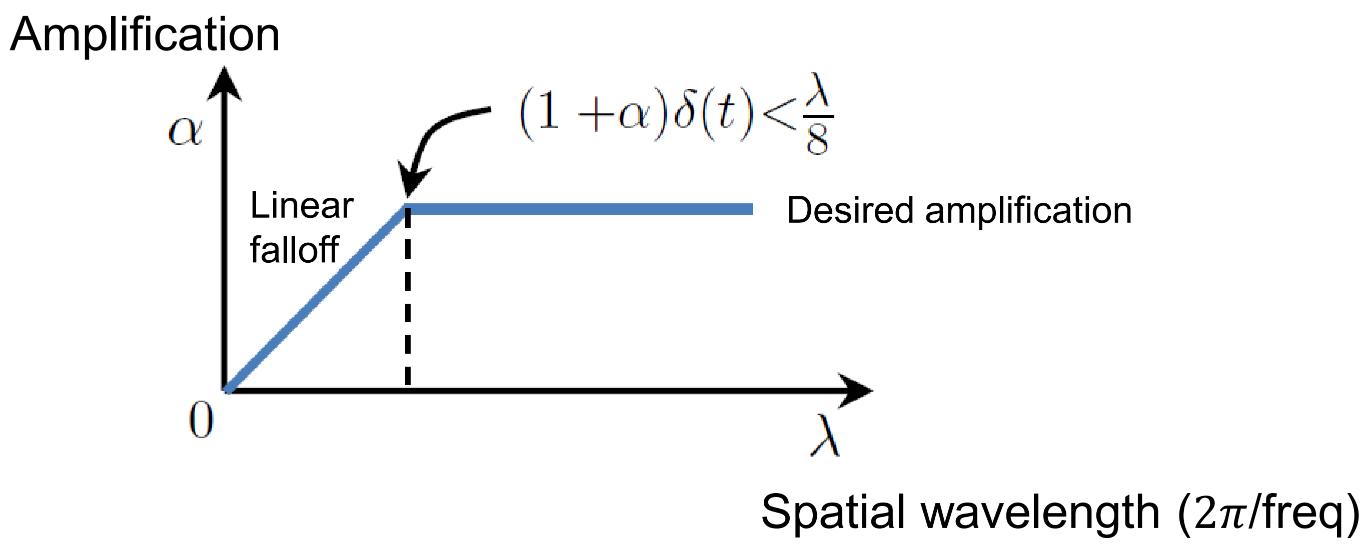
Motion-magnified (3.6-6.2 Hz, x60)

Artifact

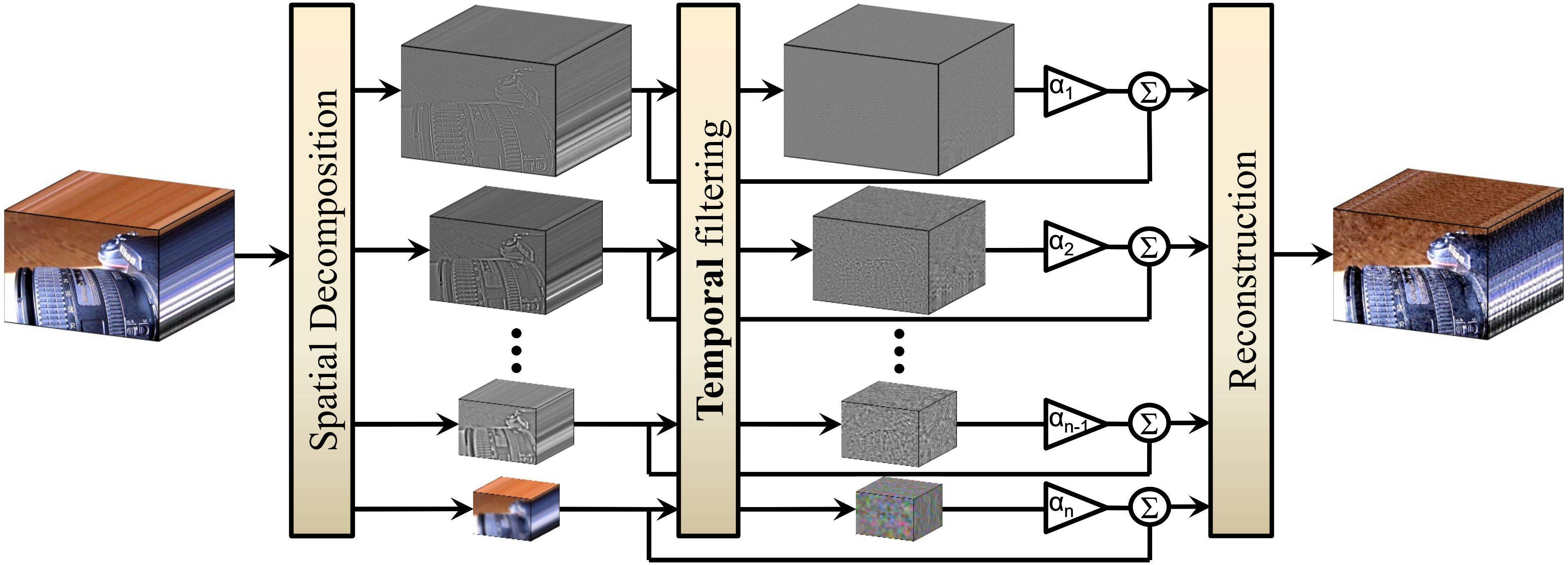
Scale-varying Amplification

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

- Use the desired α for lower spatial frequencies, and attenuate for the higher spatial frequencies



Method Recap



Laplacian Pyramid

Bandpass filter intensity at each pixel over time

Amplify bandpassed signal and add back to original

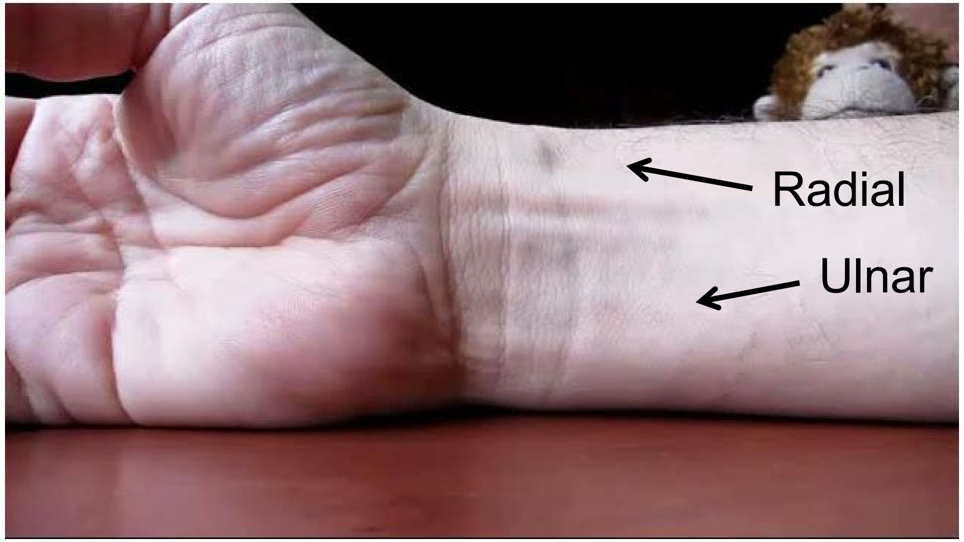
Motion Magnification Results

Source

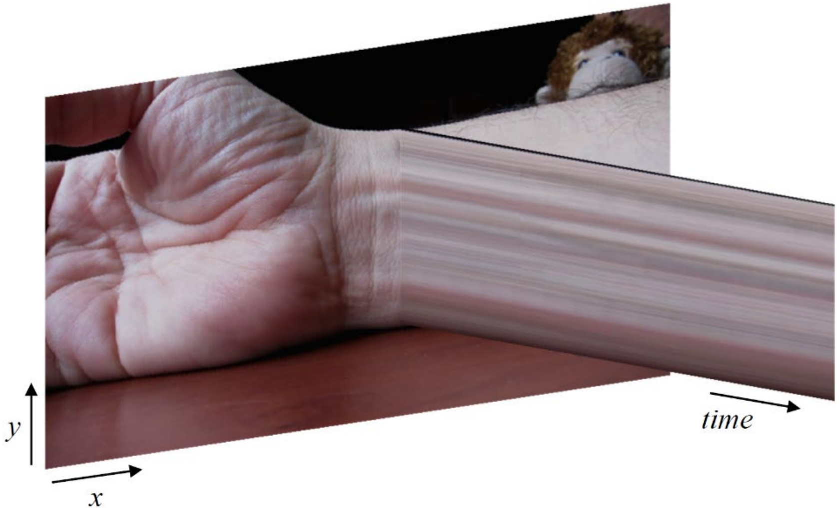
Motion-magnified (0.4-3 Hz, x10)

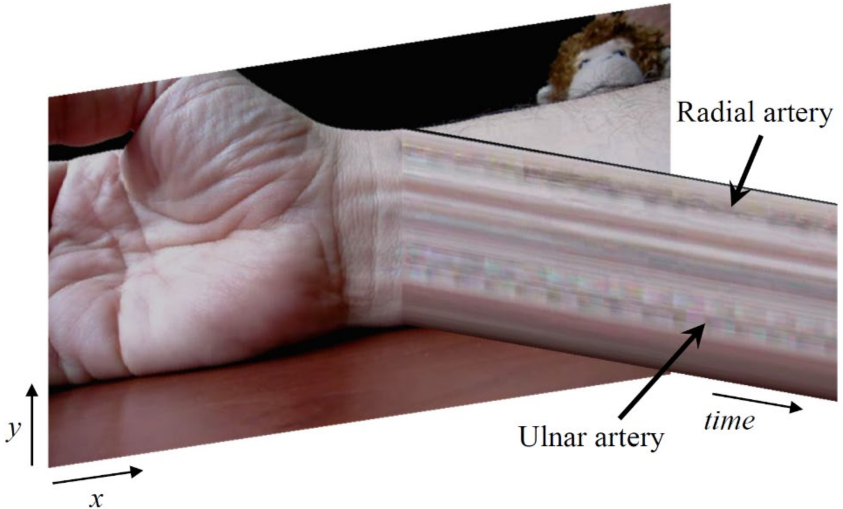
Motion Magnification

Source



Motion-magnified (0.4-3 Hz, x10)





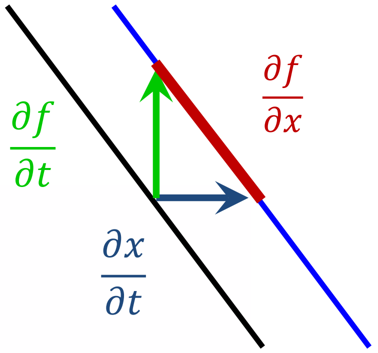
Discussion of pixelwise intensity amplification approach

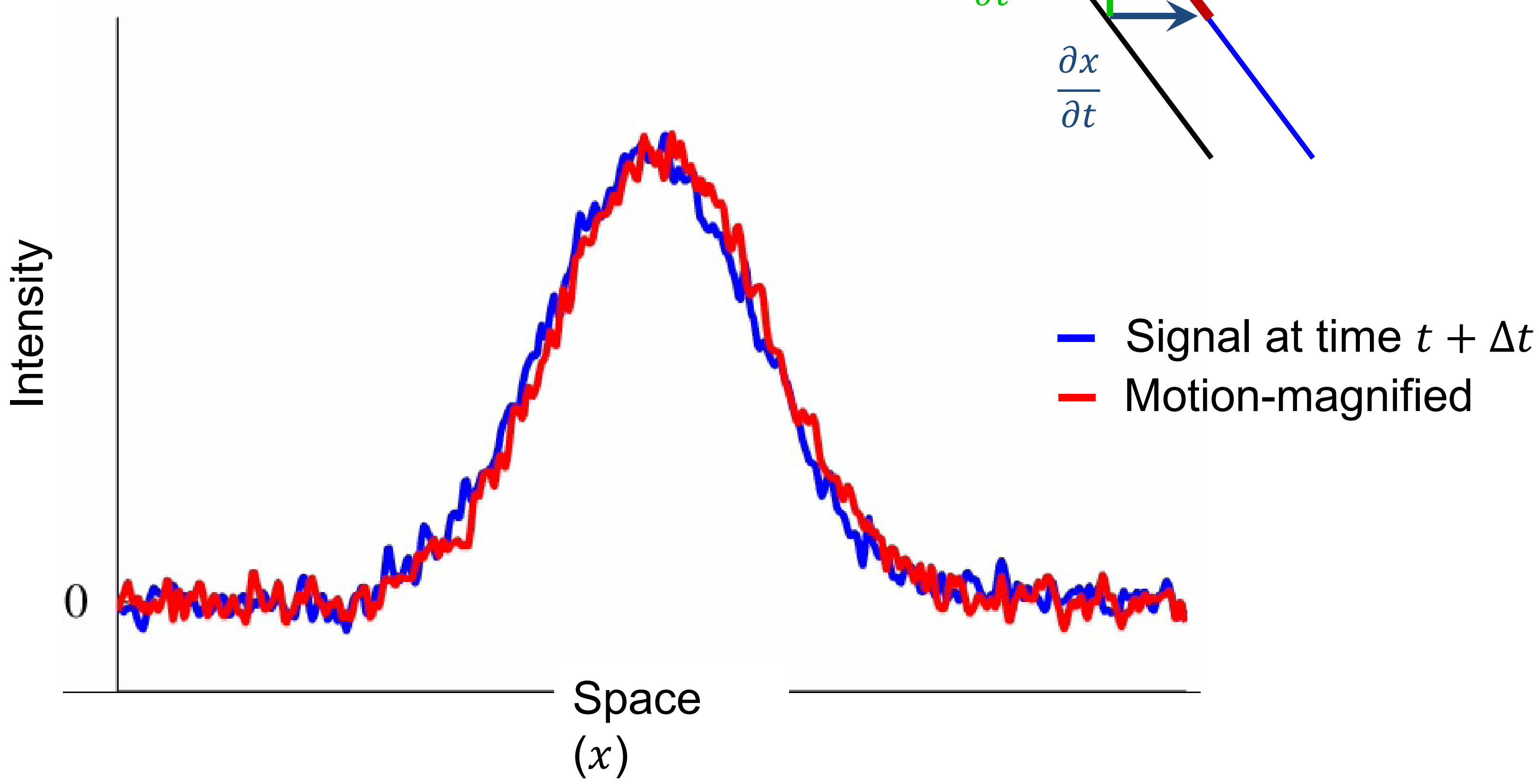
- Good:
  - Does not require explicit motion estimation or texture synthesis (robust)
  - Very fast (real time)

- Bad:
  - Can only handle very small motions
  - Amplifies noise

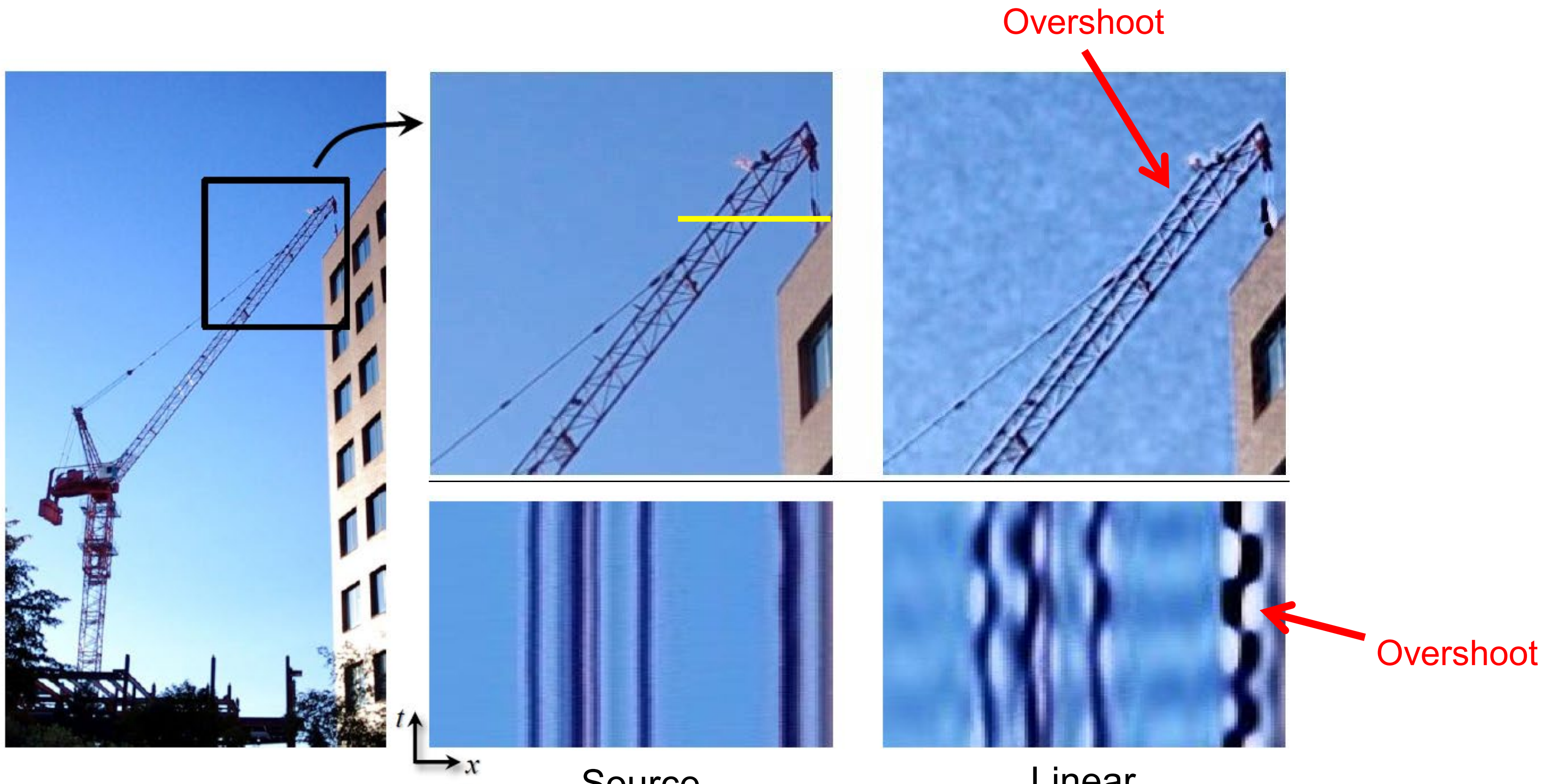
Limitations of Linear Motion Processing

- Noise amplified with signal





Limitations of Linear Motion Processing



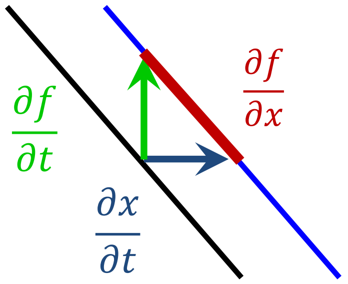
Source

Linear
SIGGRAPH'12

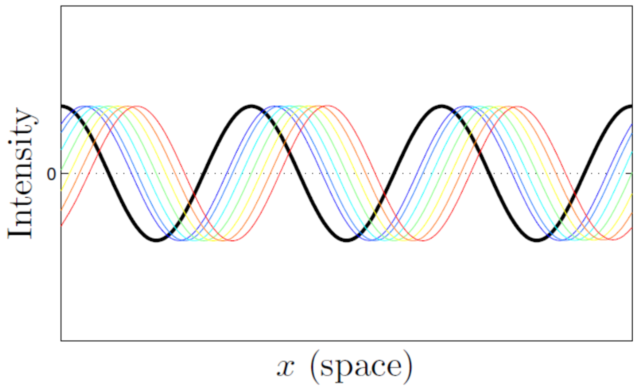
Eulerian approach part 2: shift phase instead of amplifying intensity

Translation in space is equivalent to a shift in phase

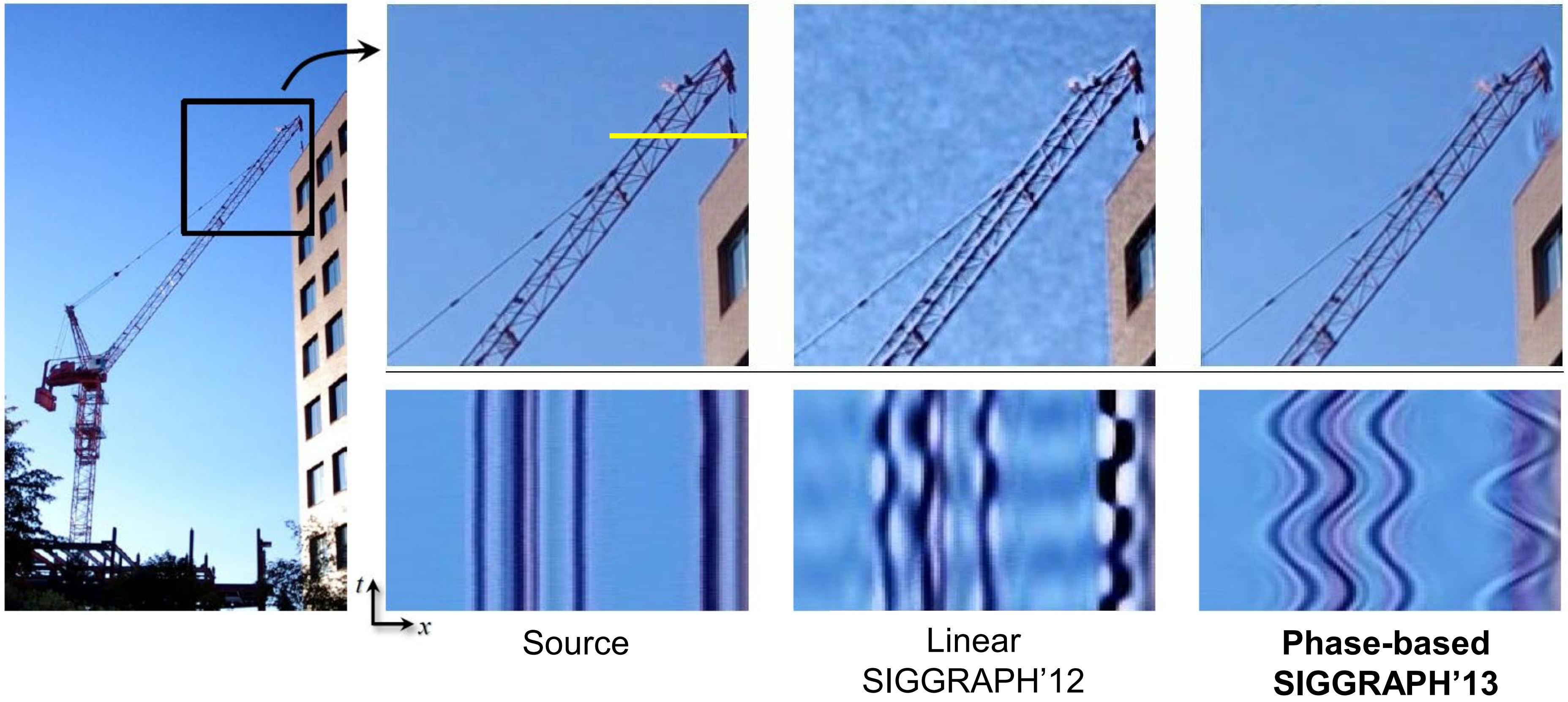
- Linear Motion Processing
  - Assumes images are locally linear
  - Translate by **changing intensities**



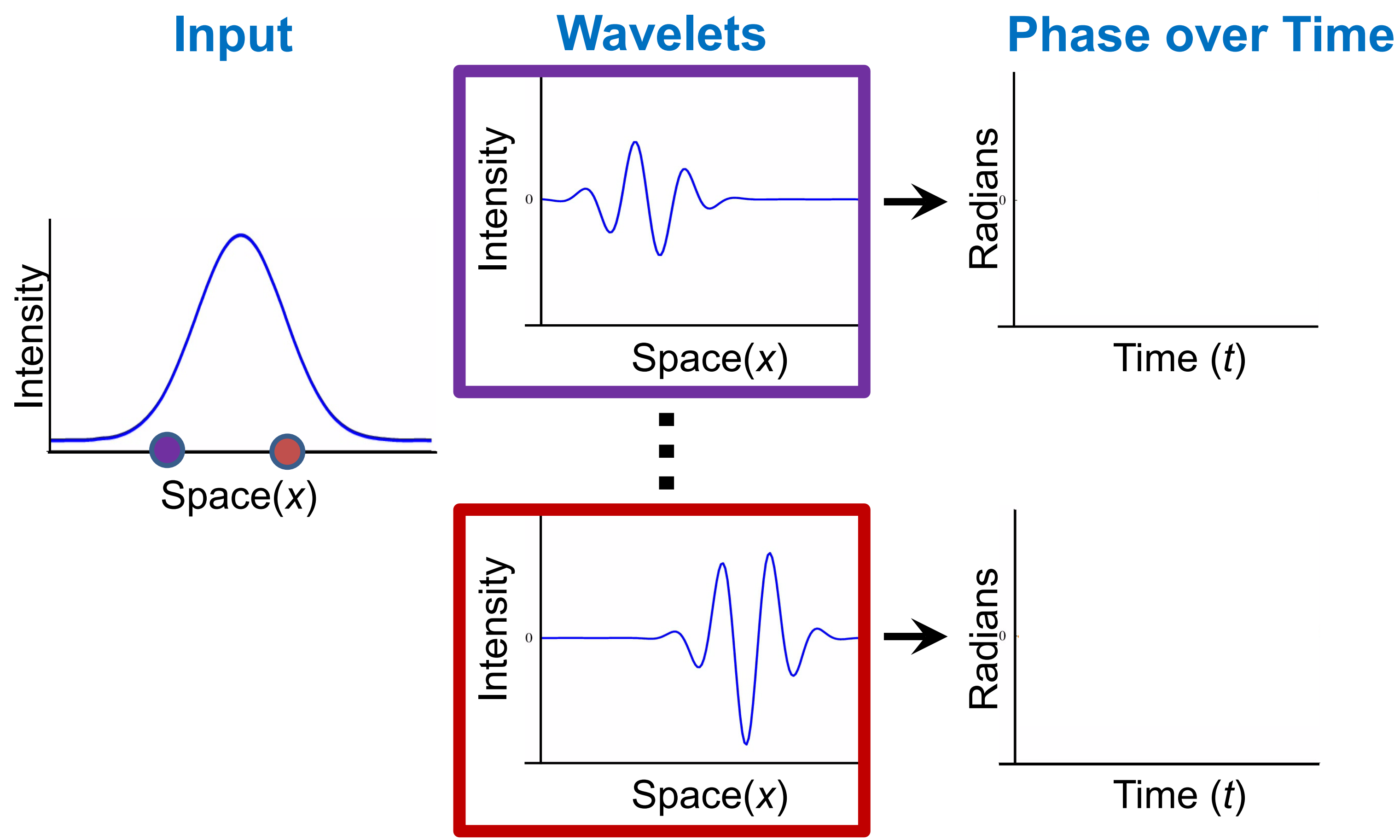
- Phase-Based Motion Processing
  - Represents images as collection of local sinusoids
  - Translate by **shifting phase**



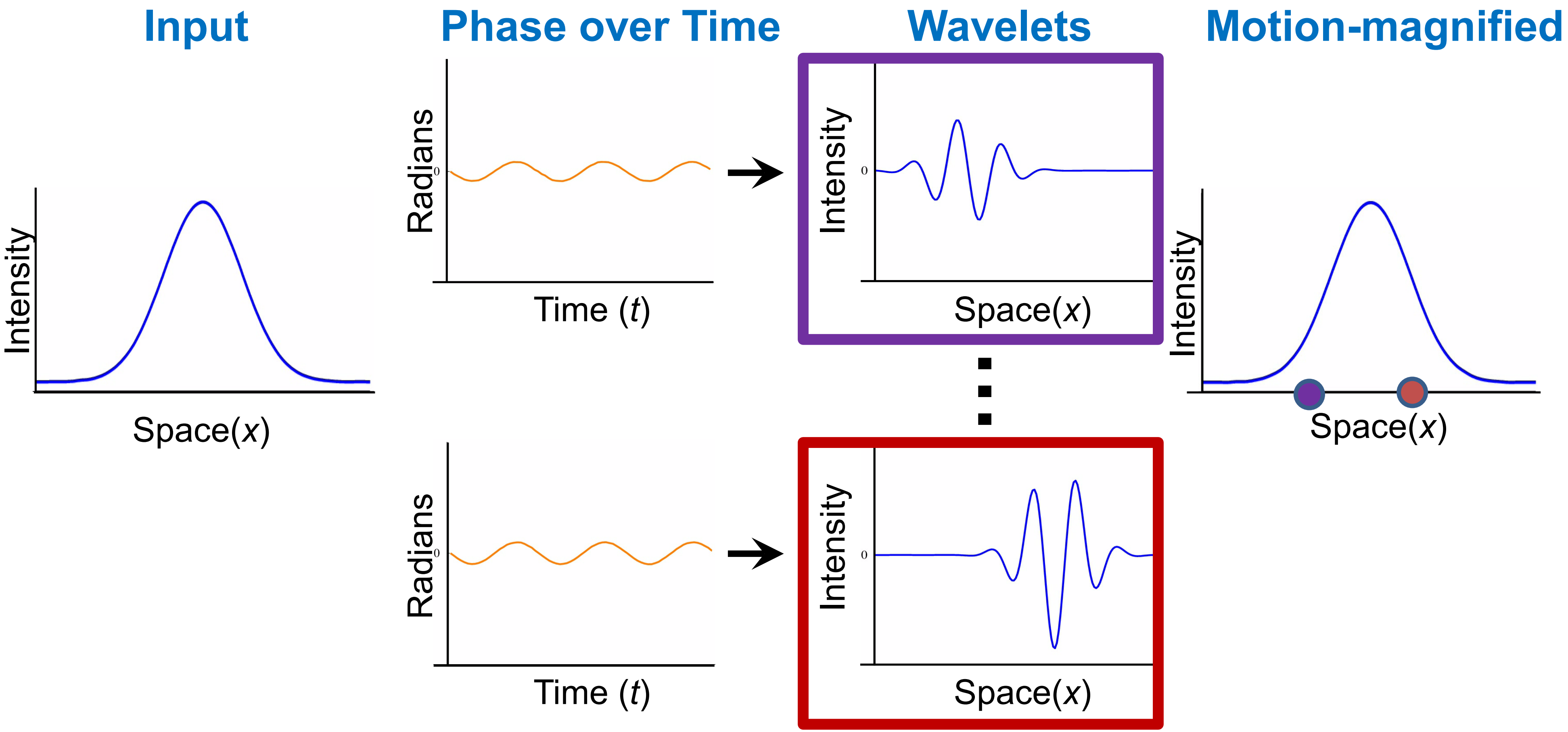
Linear vs. Phase-Based Motion Processing



Phase over Time

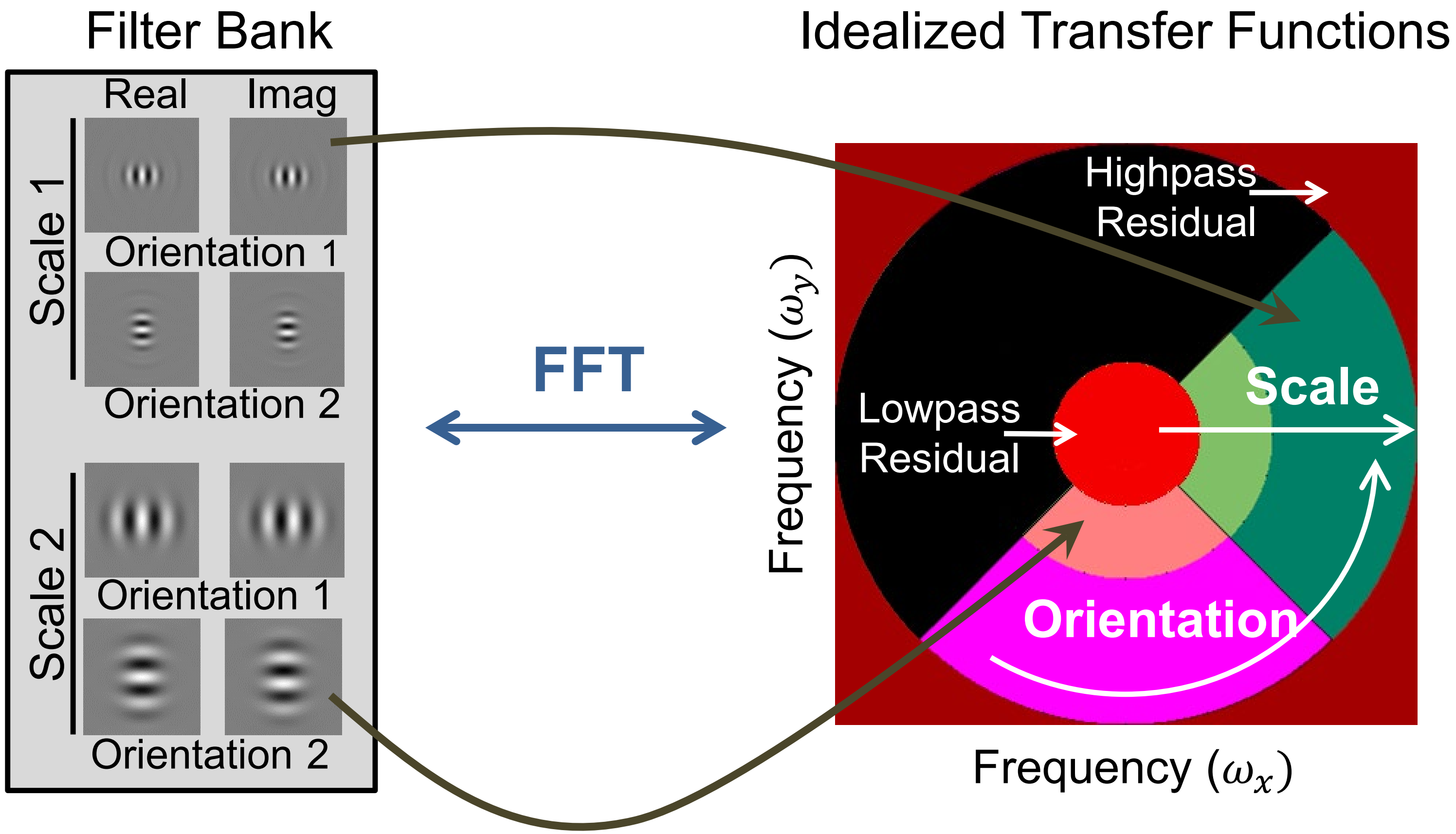


Phase over Time

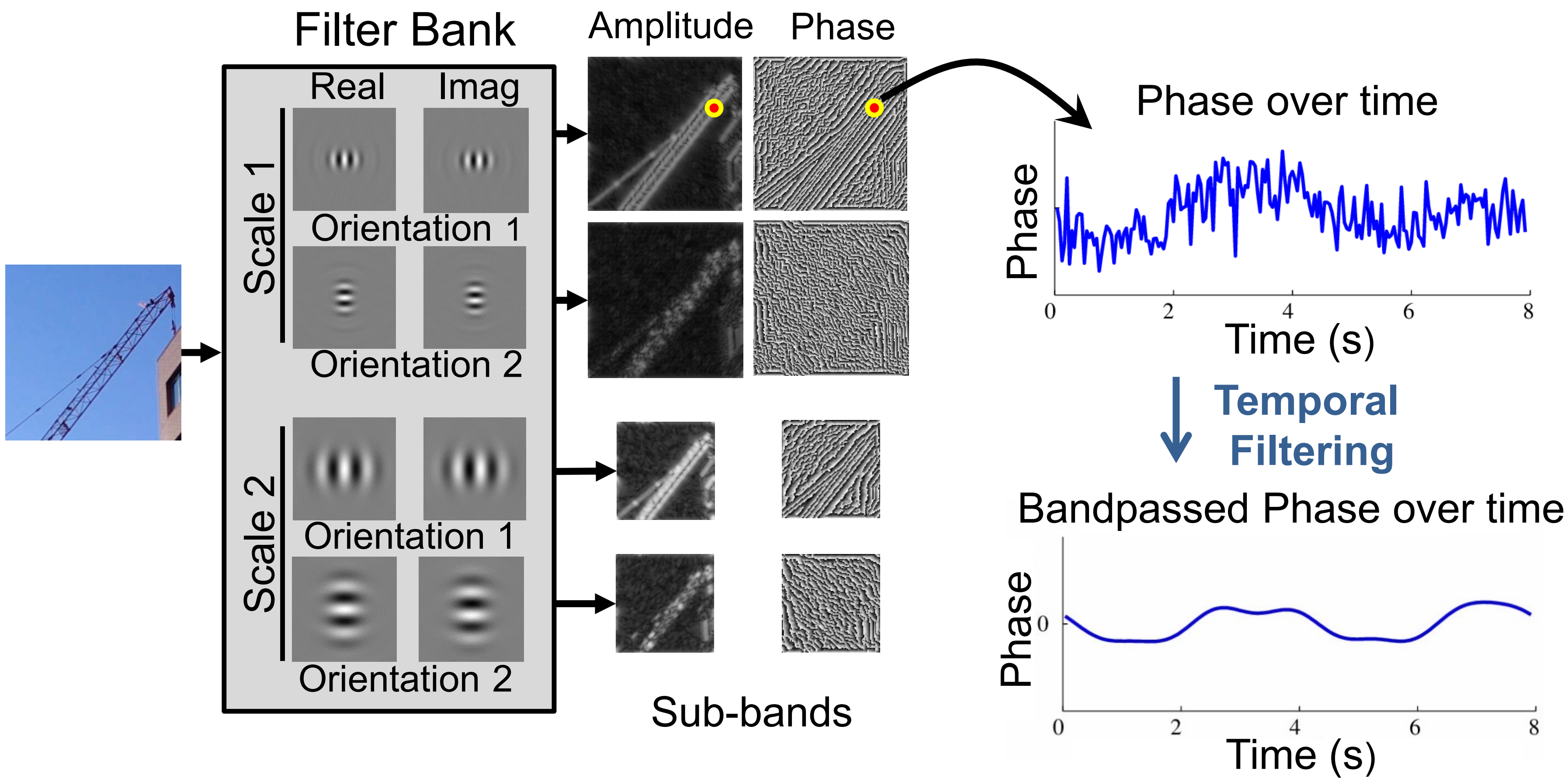


2D Complex Steerable Pyramid

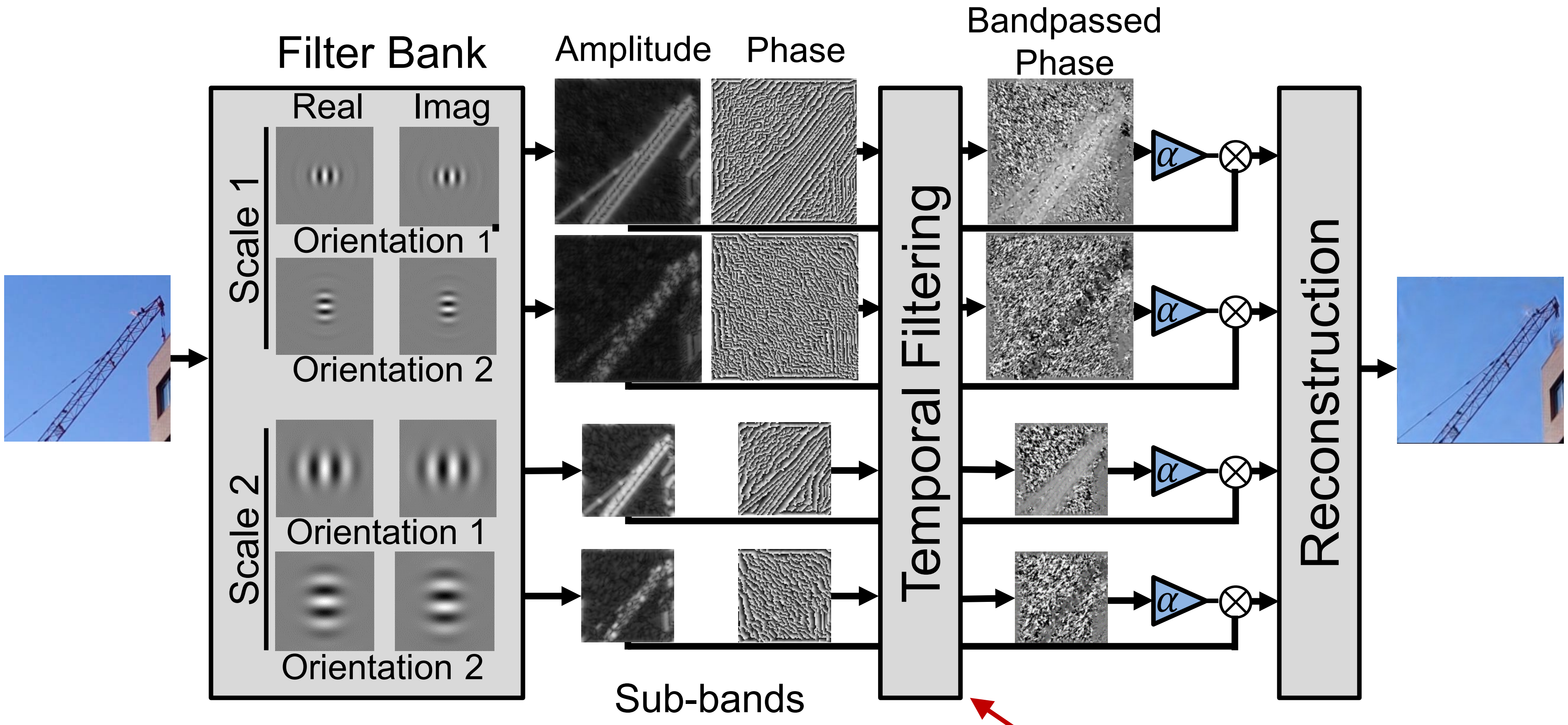
[Simoncelli et al. 1992]



Phase over Time

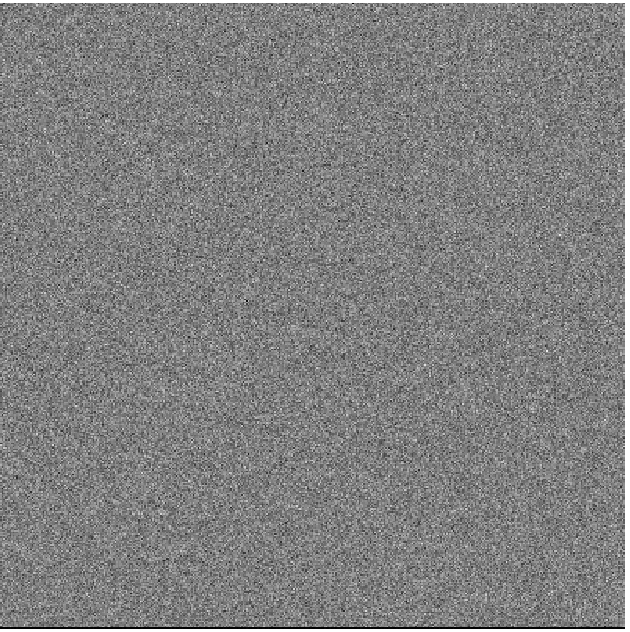


New Phase-Based Pipeline

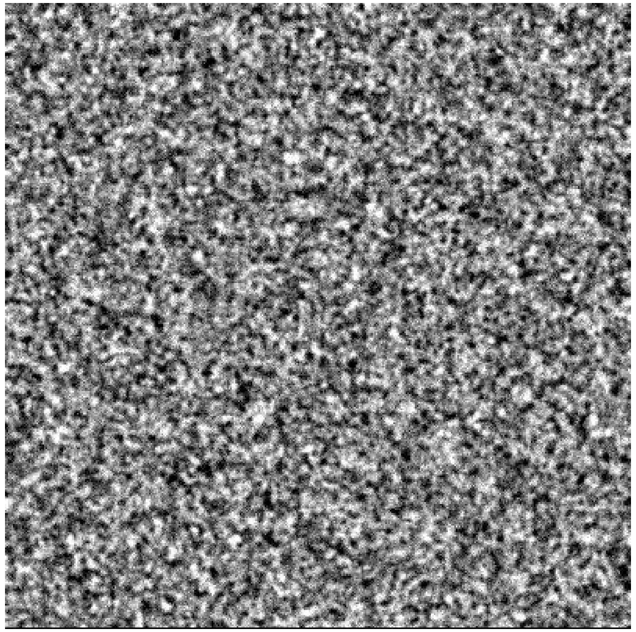


Temporal filtering on **phases**

Improvement #1: Less Noise

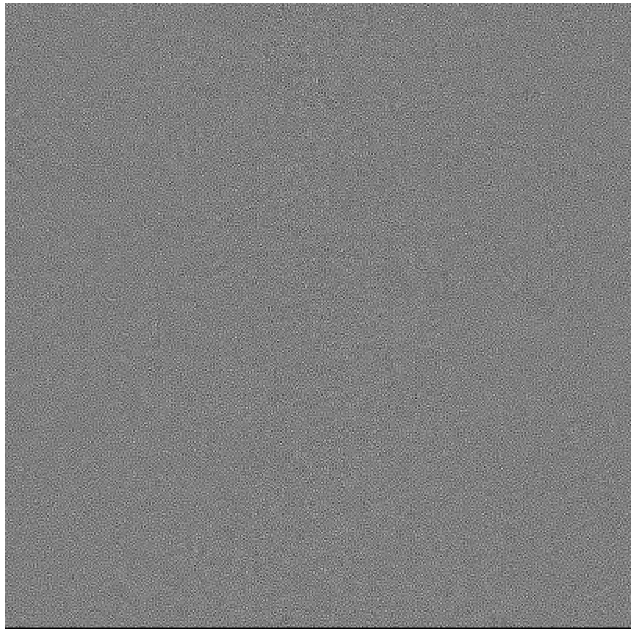


Source (IID Noise, std=0.1)



Linear [Wu et al. 2012] (x50)

Noise **amplified**

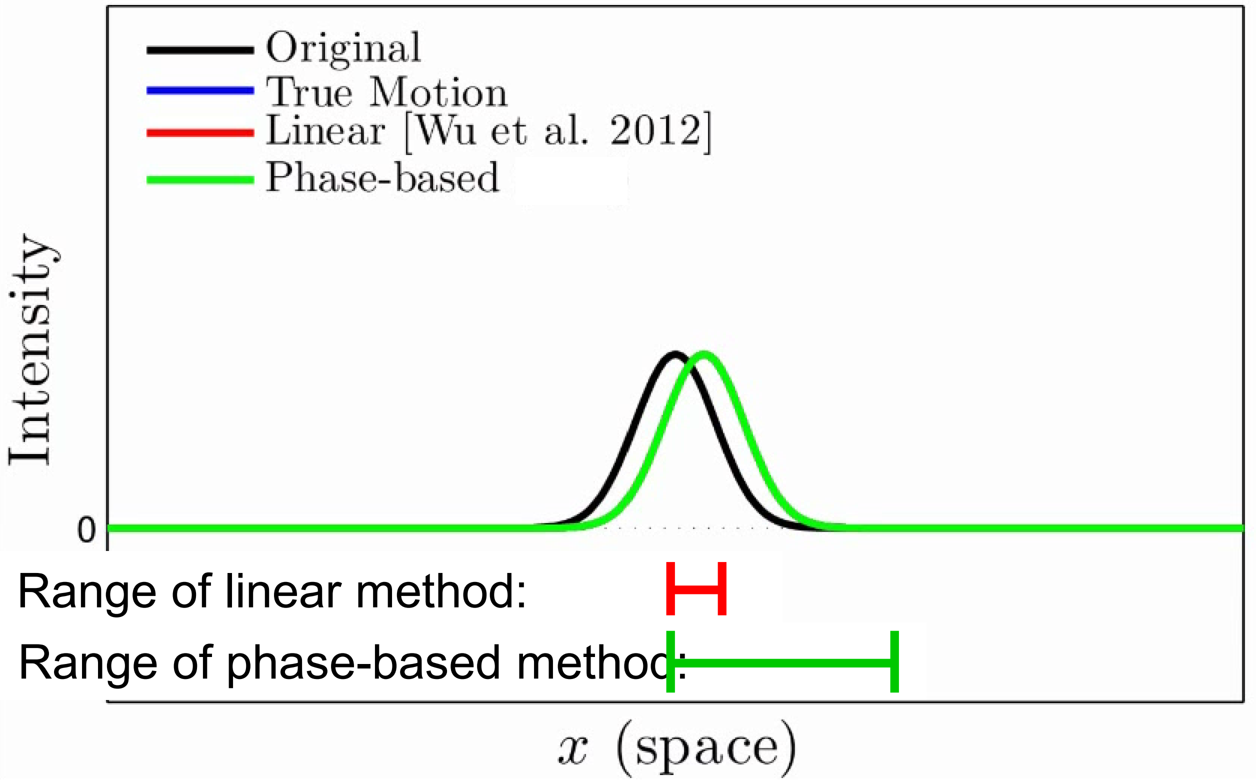


Phase-based (x50)

Noise **translated**

Improvement #2: More Amplification

Amplification factor → $\alpha = 0.0, \ \delta = 0.1$ ← Motion in the sequence



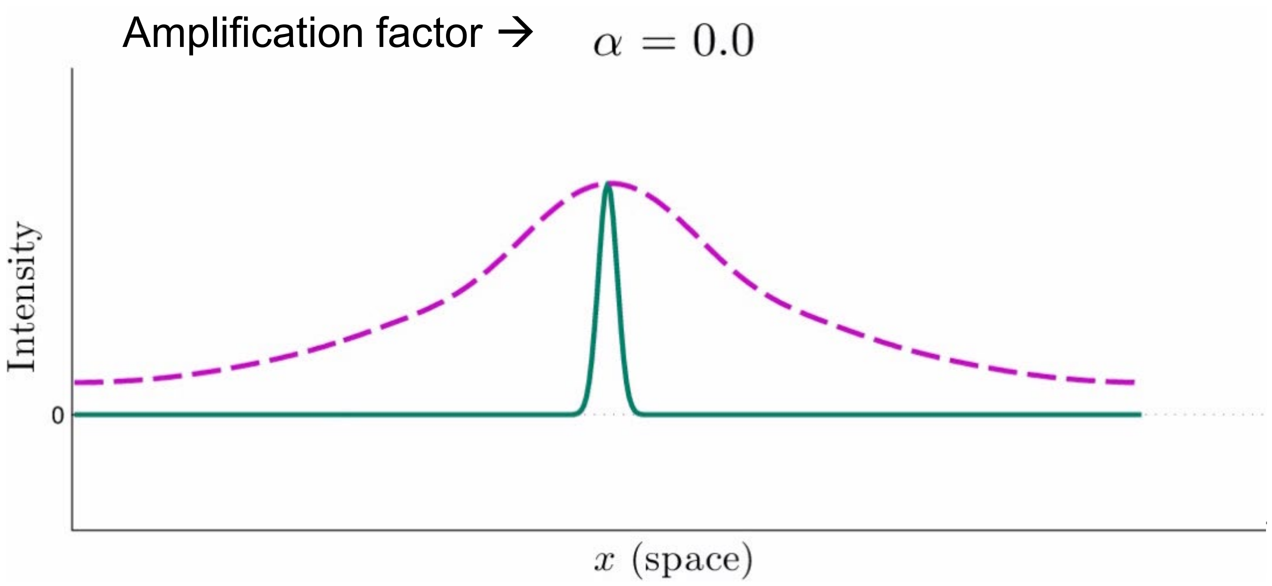
Range of linear method:

Range of phase-based method:

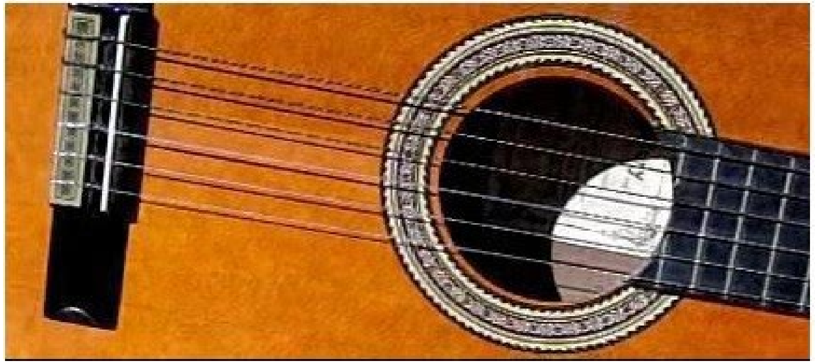
4 times the amplification!

Limits of Phase Based Magnification

- Local phase can move image features, but only within the filter window

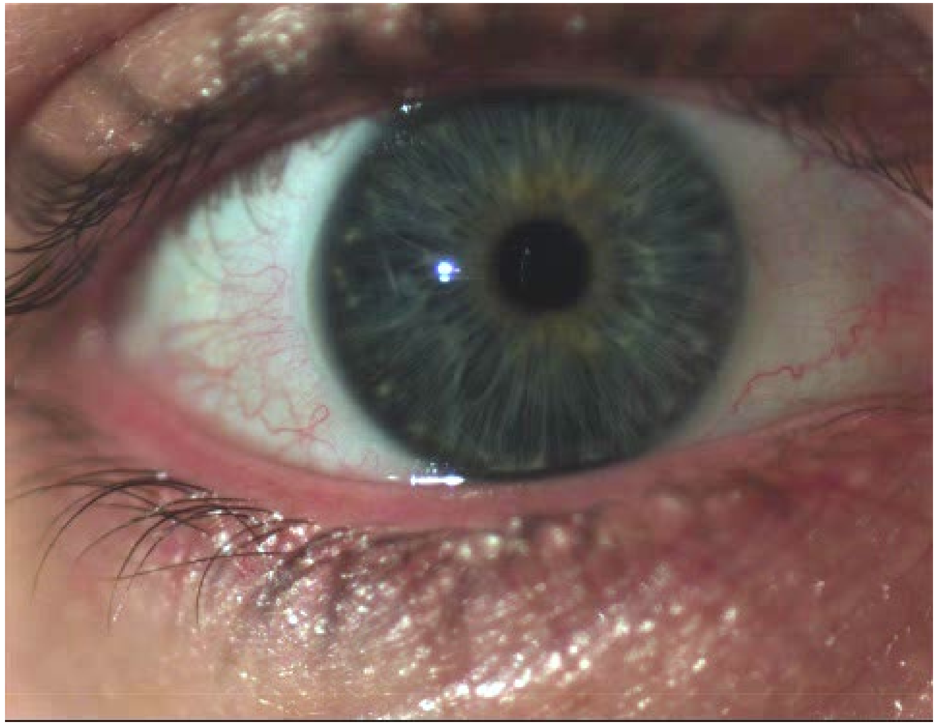


Comparison with [Wu et al. 2012]



Wu et al. 2012

Eye Movements



Source (500FPS)

Expressions

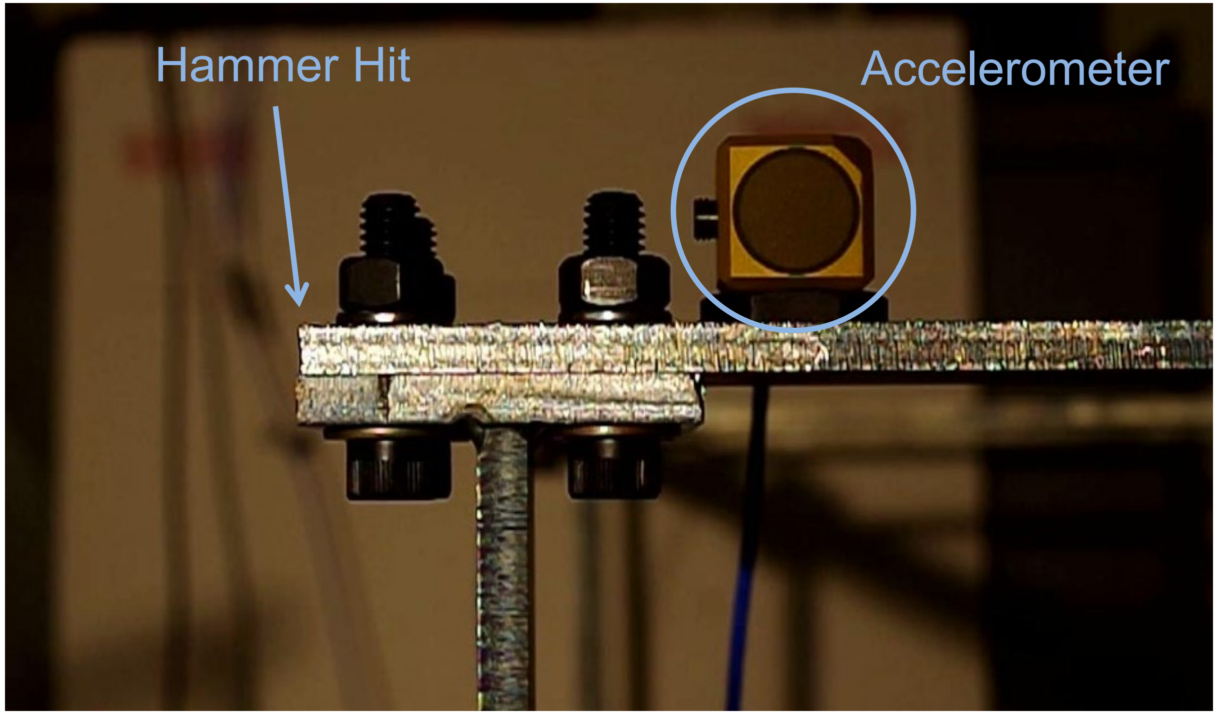
Source

Low frequency motions

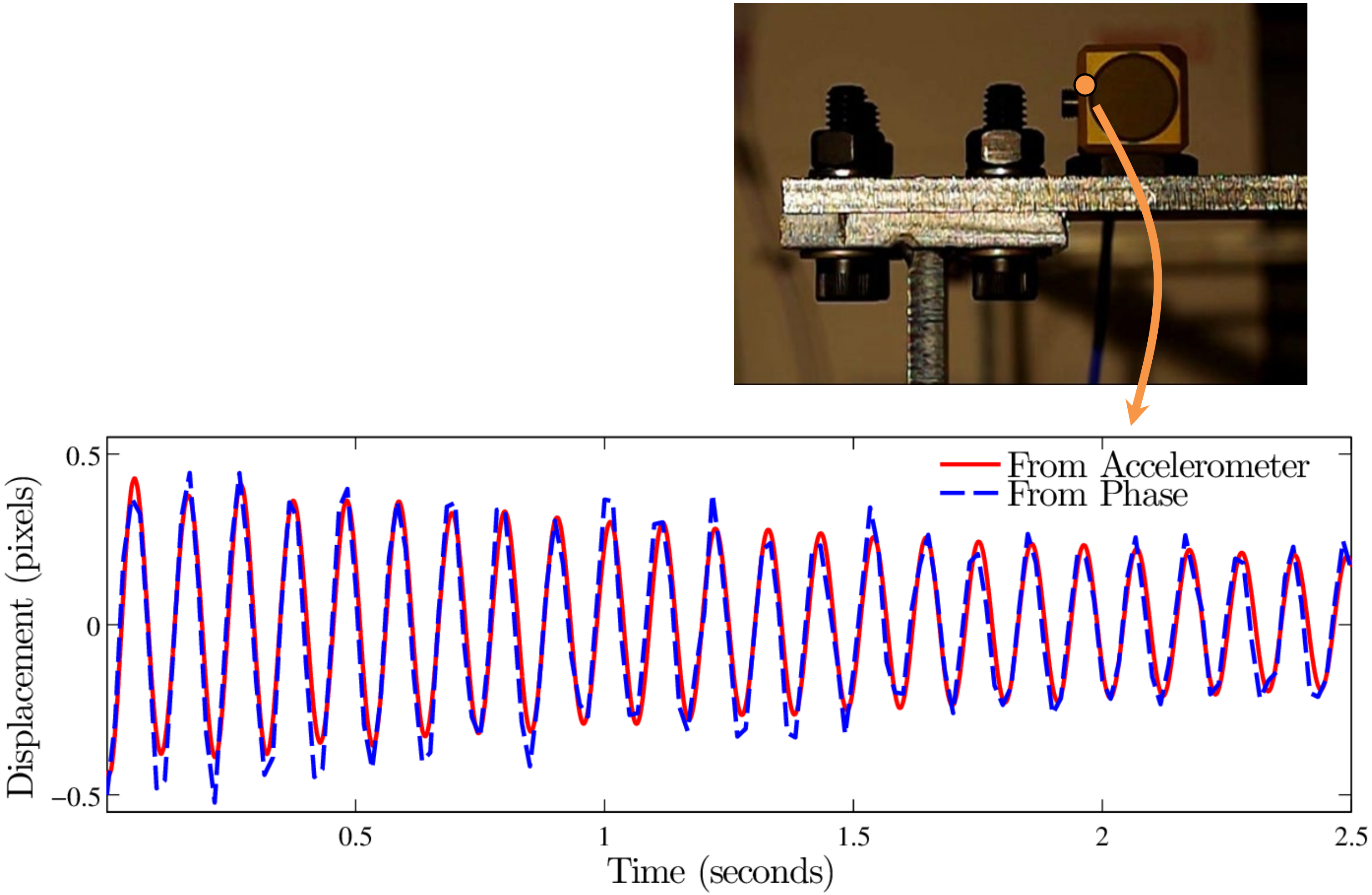
Mid-range frequency motions

Ground Truth Validation

- Induce motion (with hammer)
- Record with accelerometer



Ground Truth Validation



Motion Attenuation

Source

Sequence courtesy Vimeo user Vincent Laforet

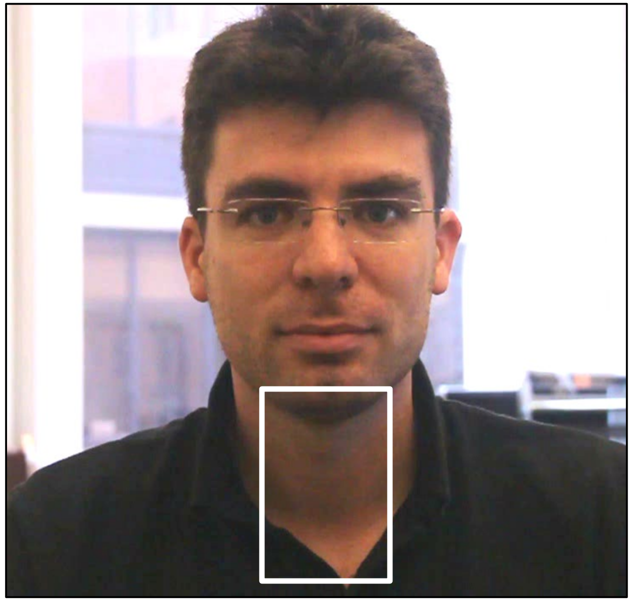
Car Engine

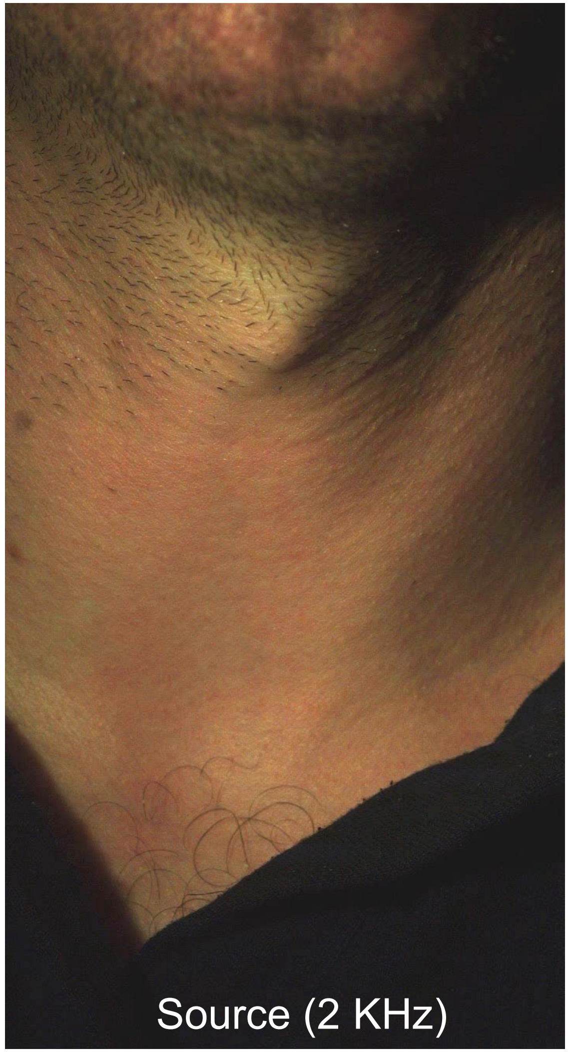
Car Engine

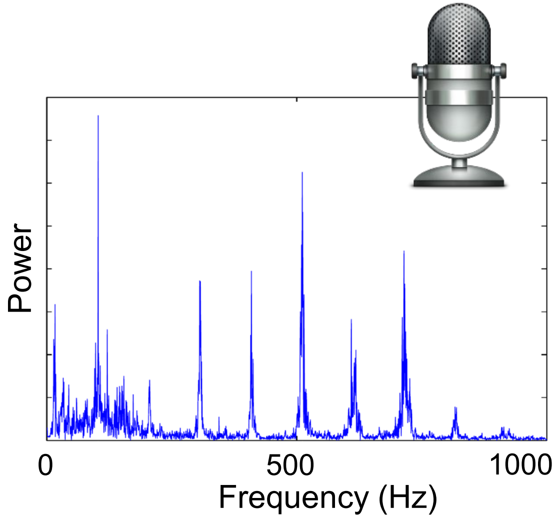
Car Engine

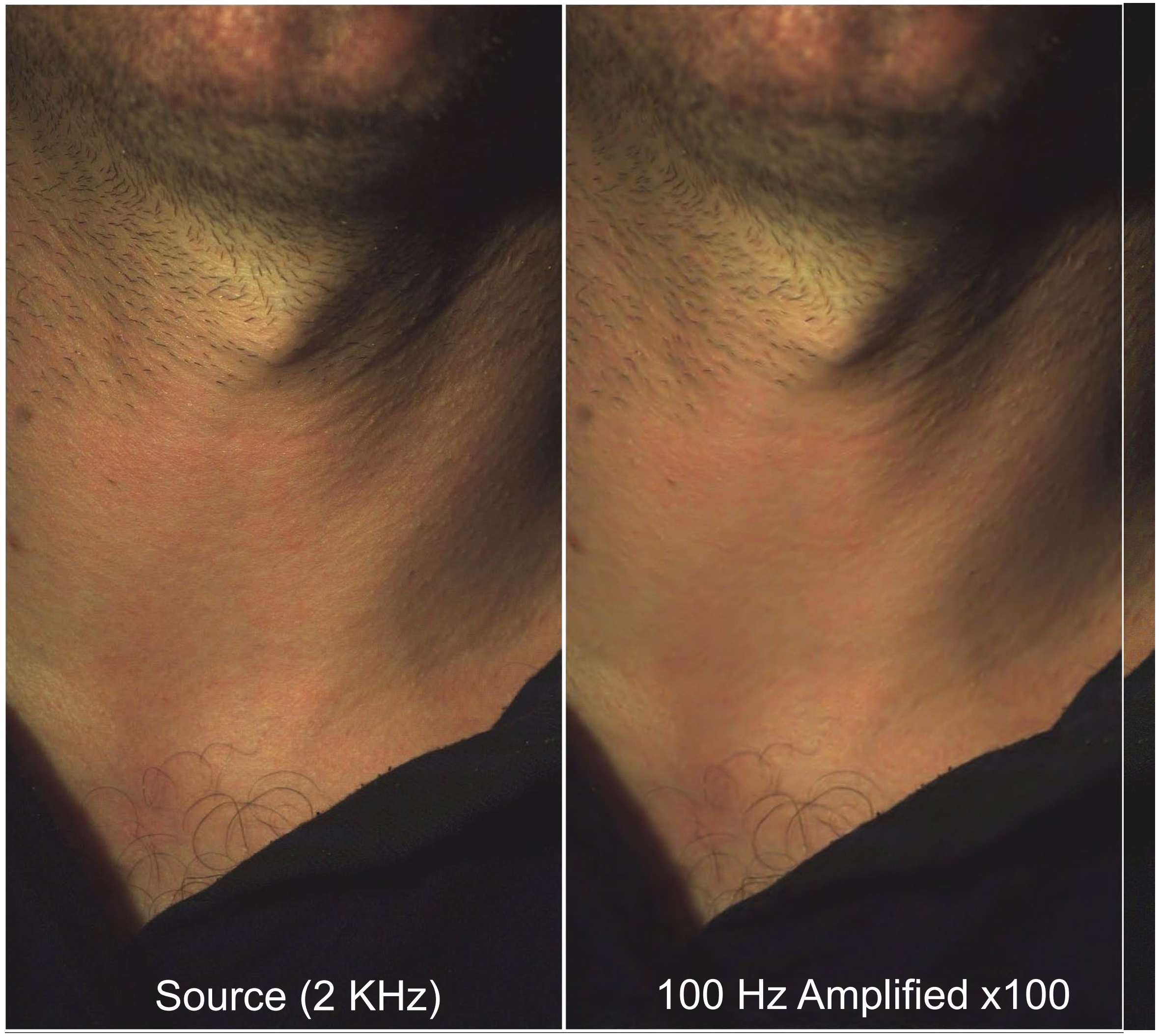
Car Engine

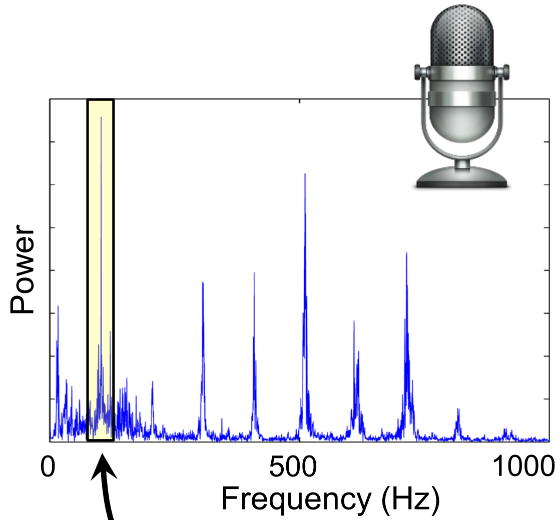
Neck Skin Vibrations



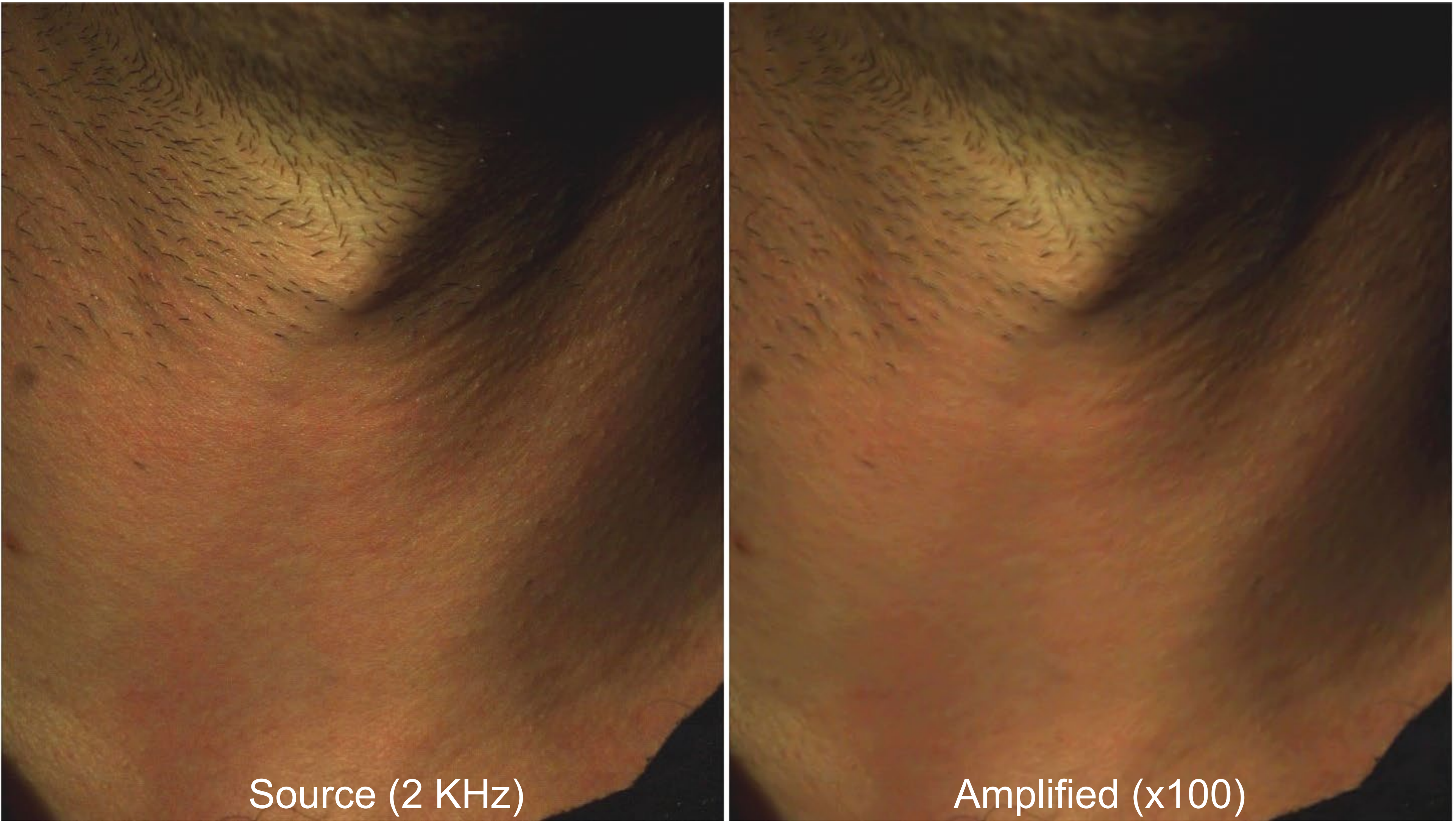








Fundamental frequency: ~100Hz


Source (2 KHz)
Amplified (x100)

Discussion of pixelwise phase magnification approach

- Good:
  - Does not require explicit motion estimation
  - Produces more direct translations (instead of perceived motion)
  - Does not amplify noise
- Bad:
  - Limited in range of amplication (compared to pointwise approach)
  - May have difficulty with non-periodic motion and large motions

Non-periodic Motions and Large Motions

Source (300 FPS)

Motion Magnification x50

Non-periodic motion

Motion Magnification x50
Large Motions Unmagnified

SIGGRAPH2014

The Visual Microphone:

Passive Recovery of Sound from Video

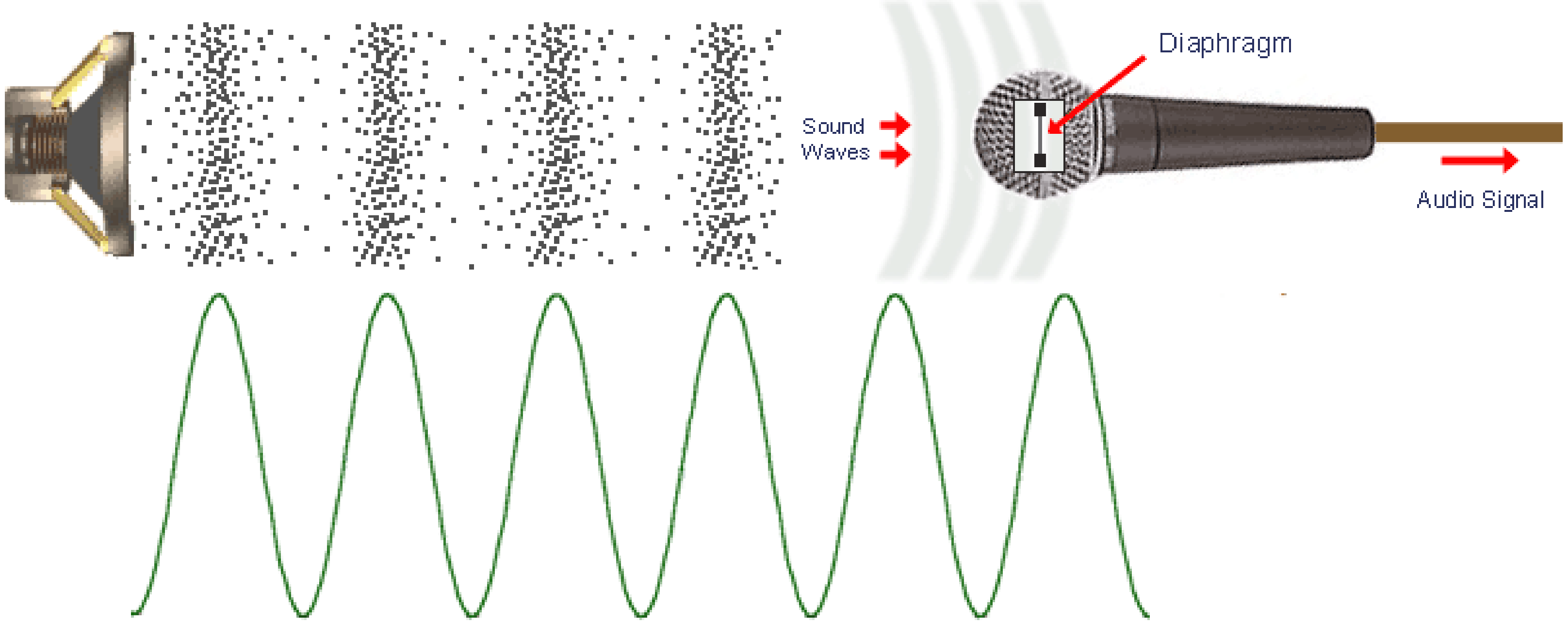
Abe Davis Michael Rubinstein Neal Wadhwa
Gautham Mysore Fredo Durand William T. Freeman

MIT CSAIL Microsoft Research Adobe

(slides adopted from Siggraph presentation)

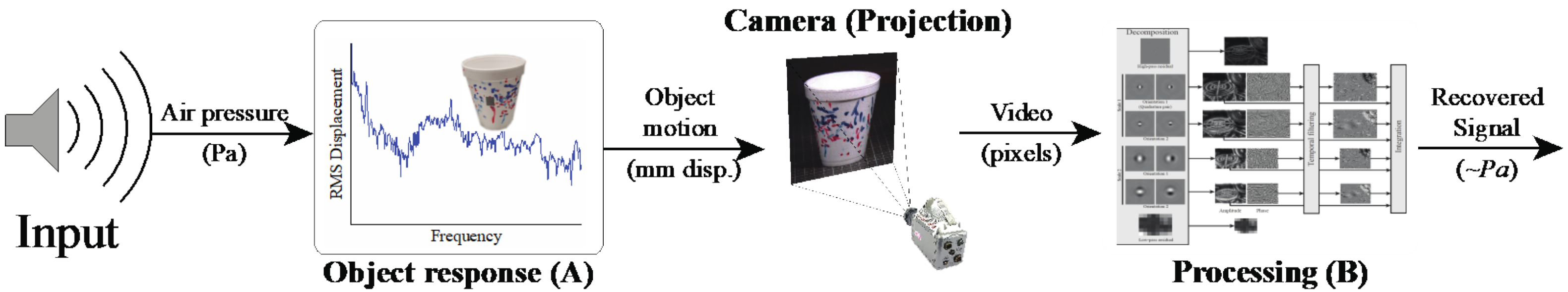
Remote Sound Recovery

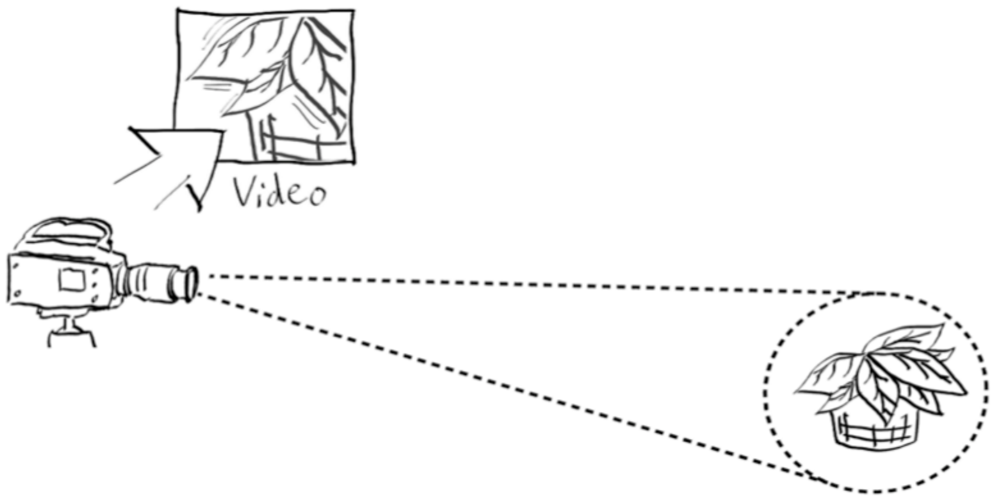
Sound and Motion



Source: mediacollege.com

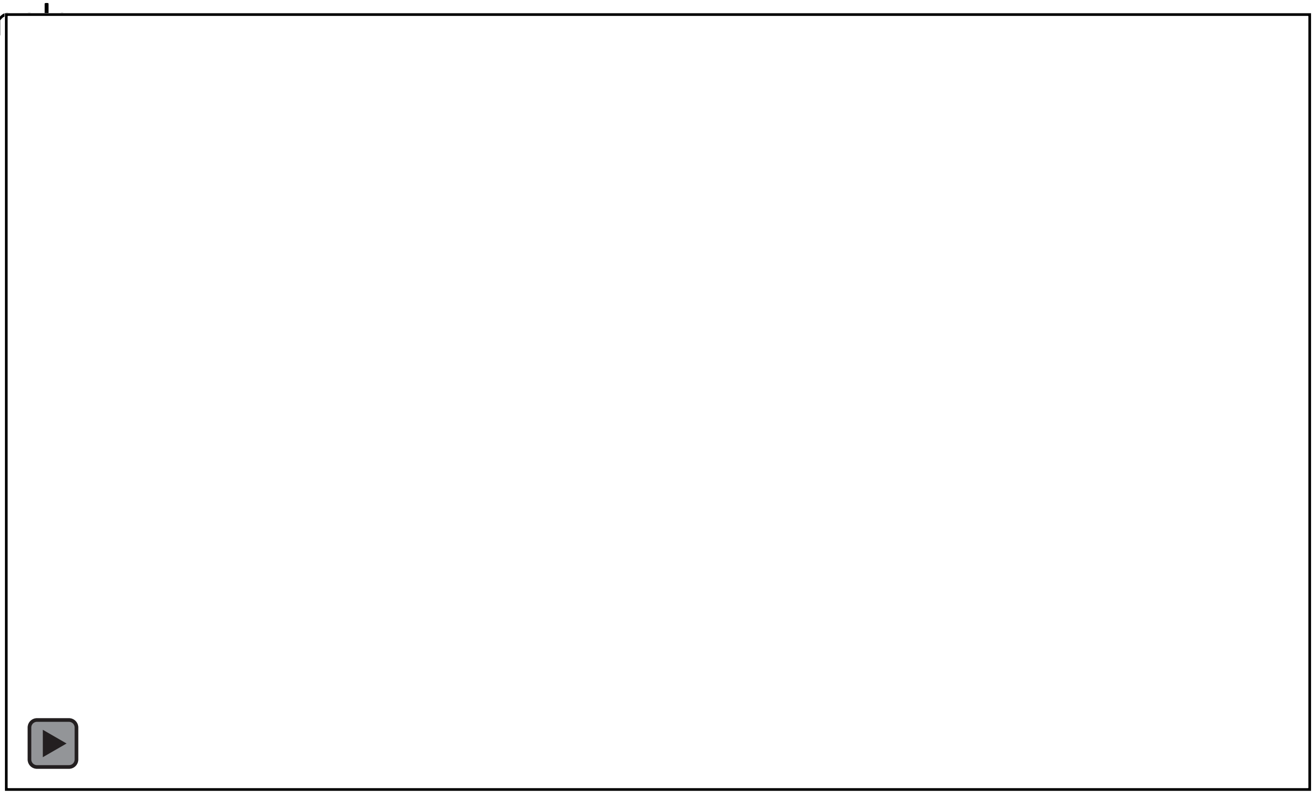
The Visual Microphone



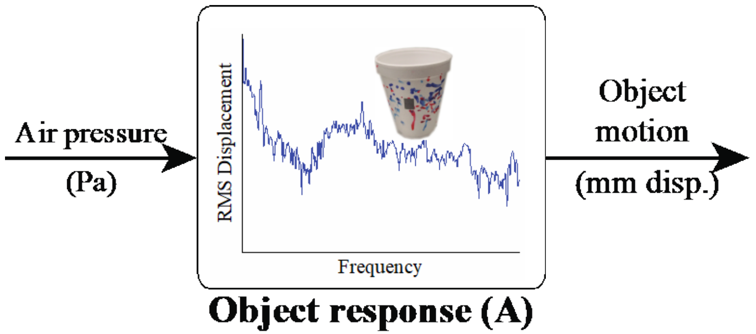


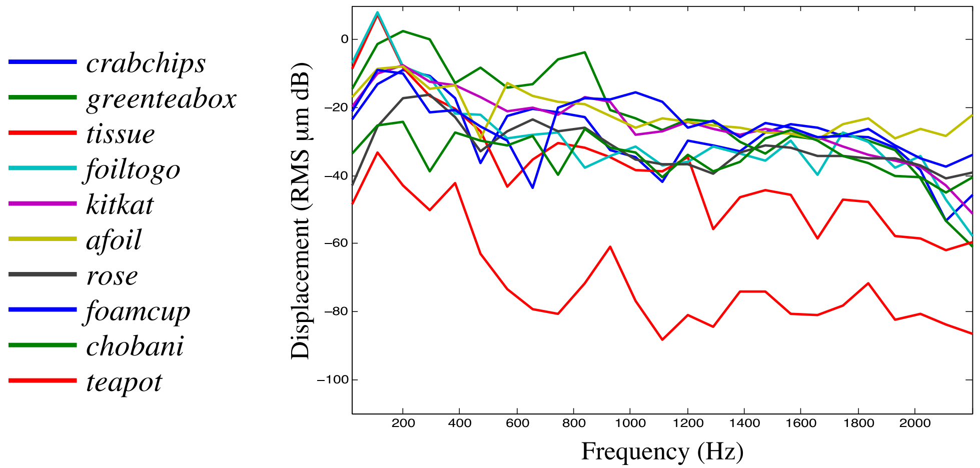
Processing

- Extract local motion signal
- Average and Align
- Post-process



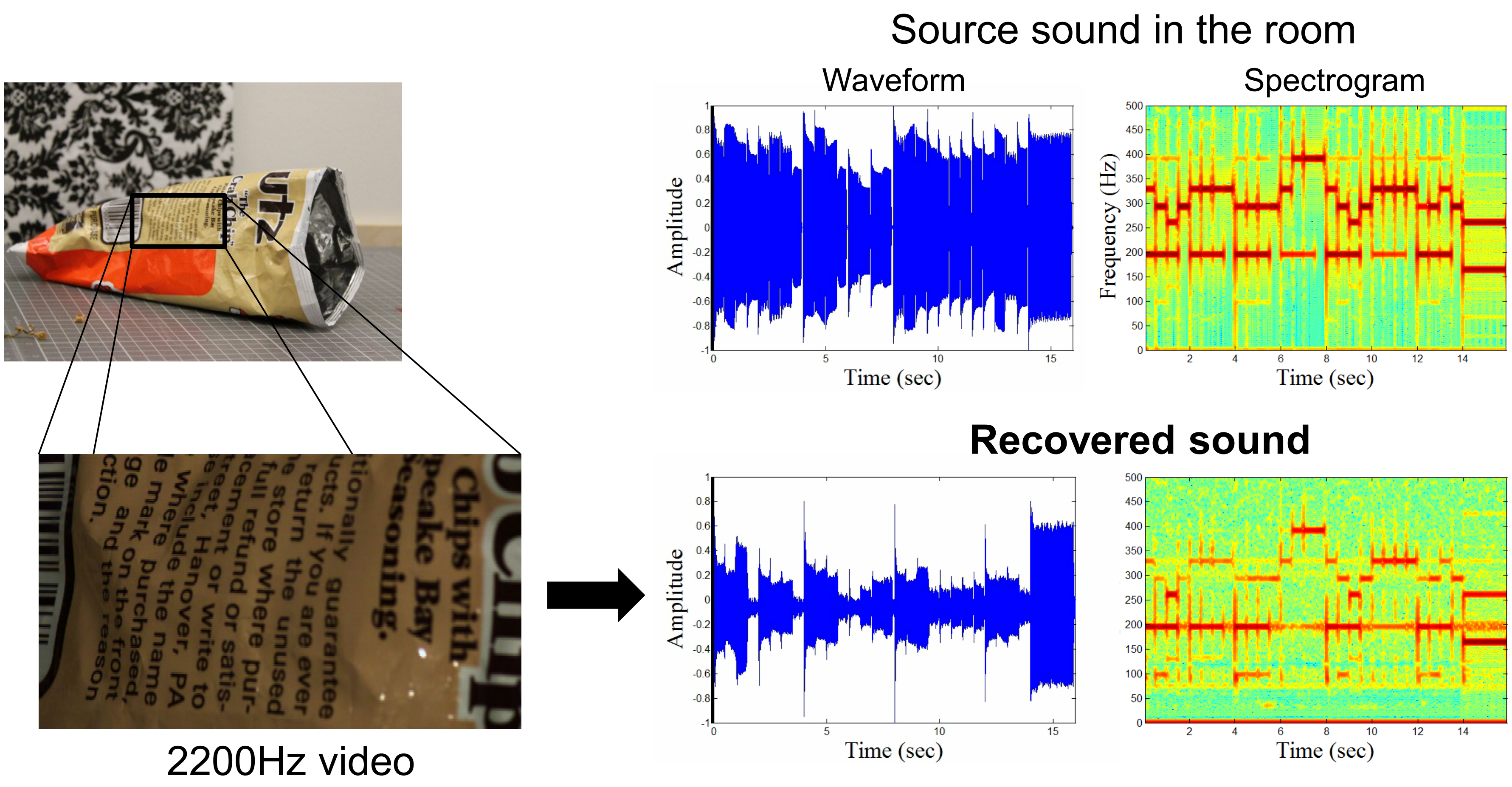
Some materials are better microphones than others





(c) Frequency responses

Sound Recovered from Video

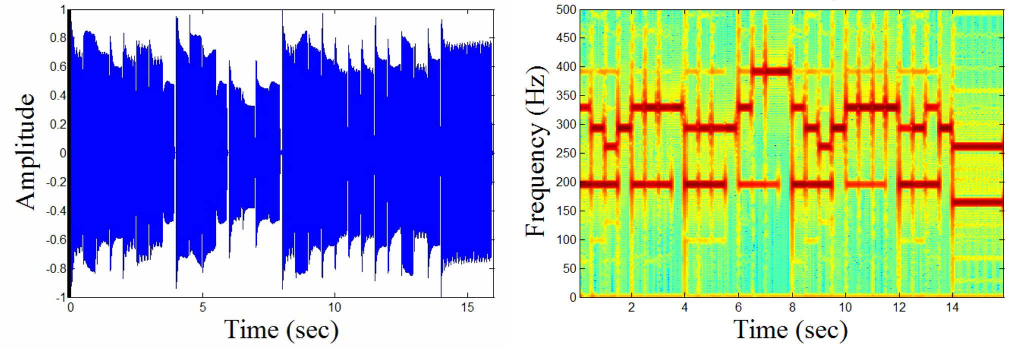


Sound Recovered from Video

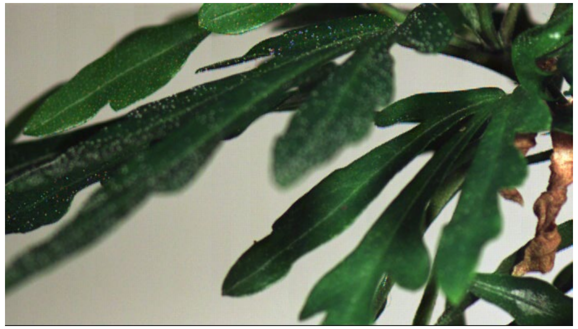
Source sound in the room

Waveform

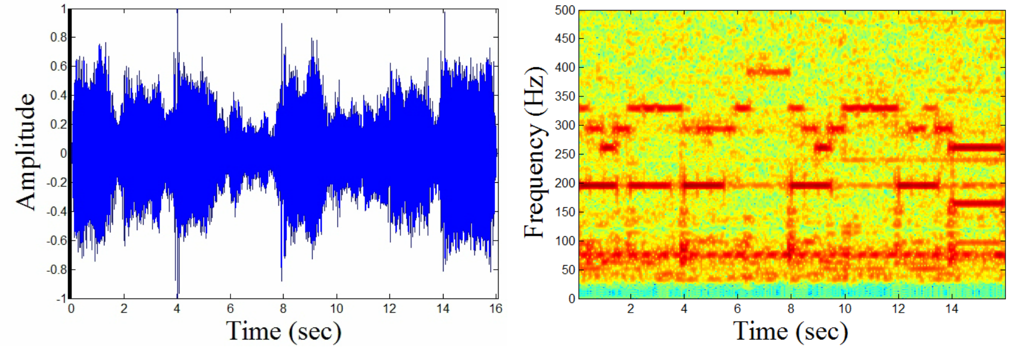
Spectrogram



Recovered sound



2200Hz video



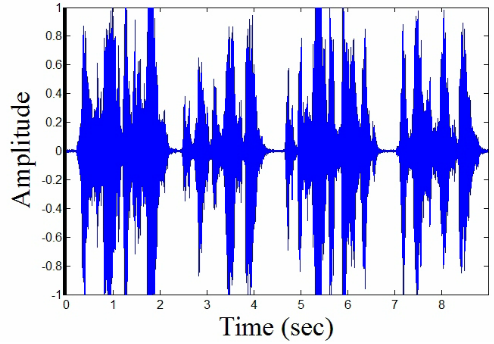
Sound Recovered from Video

(small patch on the chip bag)

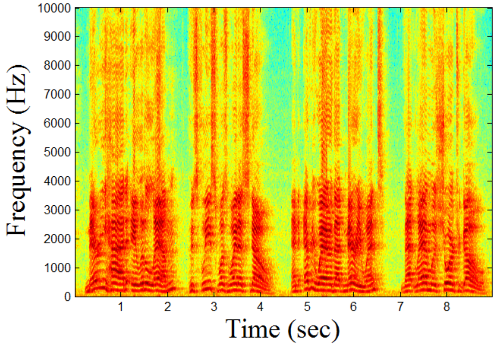
20 kHz video

Source sound in the room

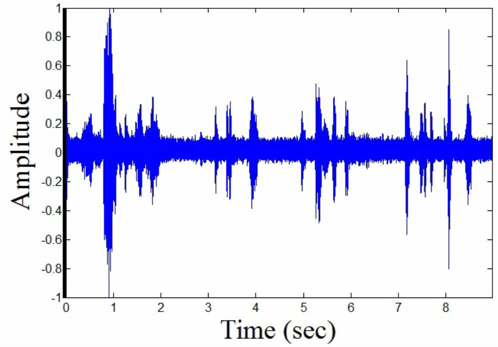
Waveform

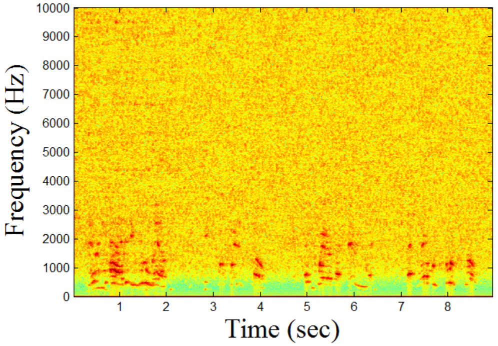


Spectrogram



Recovered sound

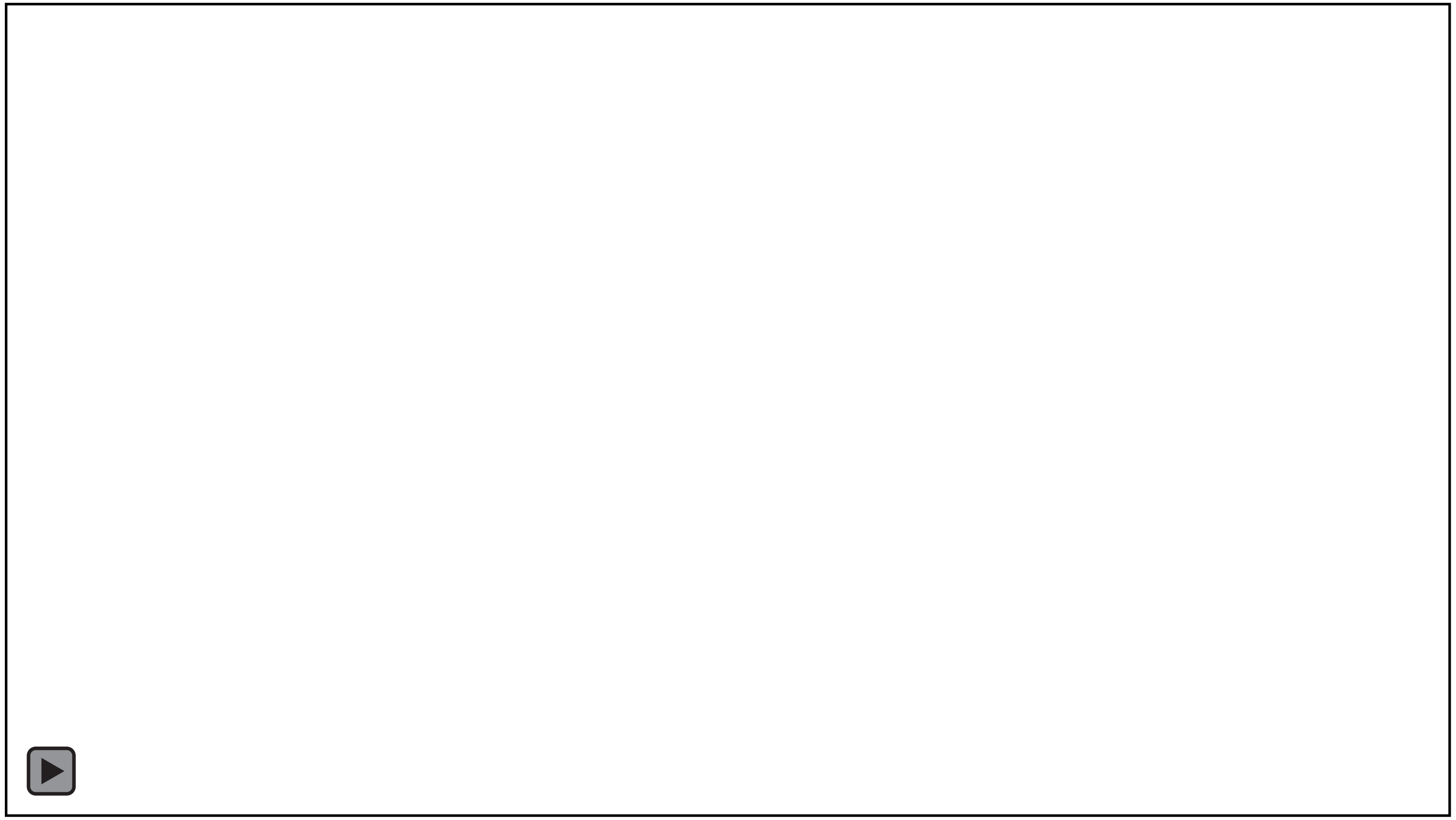




Rolling Shutter

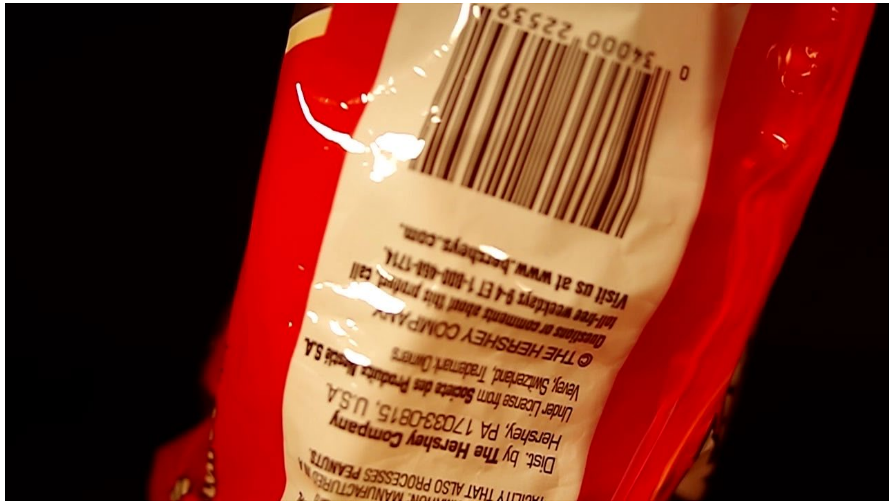
https://www.flickr.com/photos/sorenragsdale/3904937619/
http://www.flickr.com/photos/boo66/5730668979/

Rolling Shutter

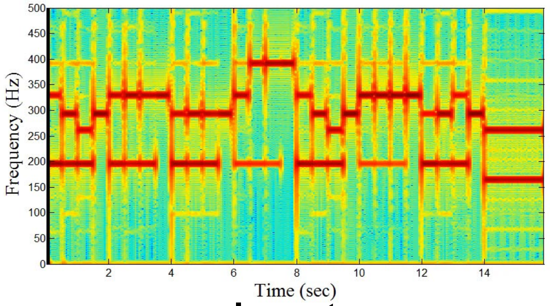


Rolling Shutter

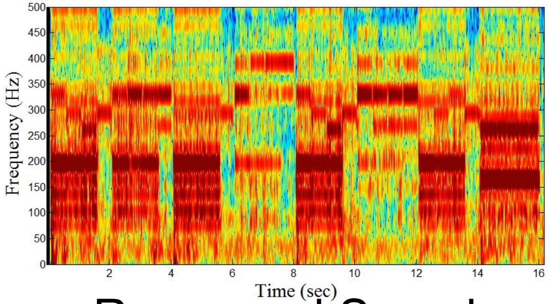
Rolling Shutter



Input video (60 fps)

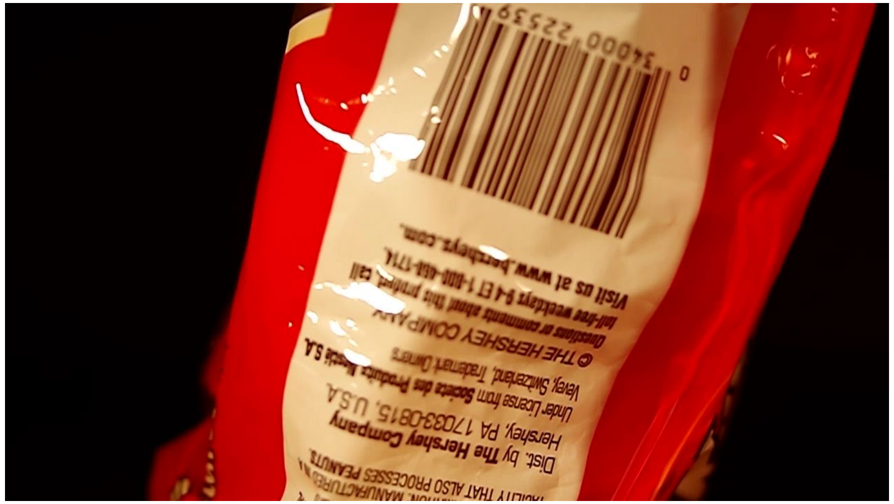


Input

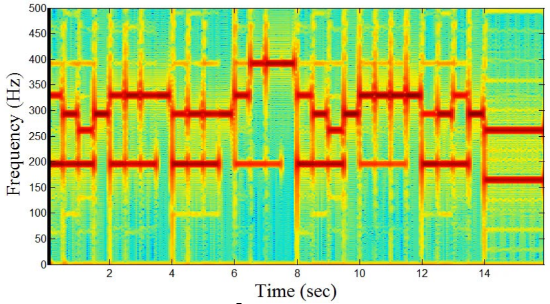


Recovered Sound

Rolling Shutter



Input video (60 fps)



Input

400Hz!

Recovered Sound

Summary

- Several ways to magnify motion
  - Directly measure and exaggerate point motions
  - Amplify intensity changes after temporal filtering (creating apparent motion)
  - Amplify local phase variations after temporal filtering

- Micro-motion estimates can be used to measure sound

Next week

- Final class
  - A few examples of cutting edge applications
  - Where to learn more
  - Course feedback (important for me)