

Video Magnification



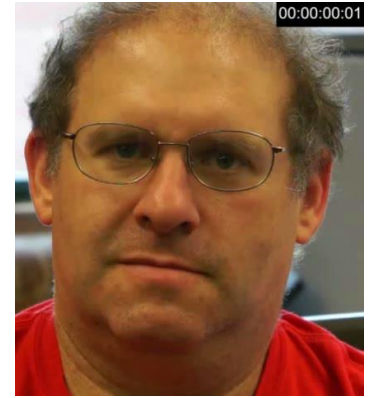
Magritte, "The Listening Room"

Computational Photography
Derek Hoiem, University of Illinois

Today

1. Go over common mistakes in midterm
2. Video Magnification
 - Lagrangian (point tracking) approach
 - Eulerian (signal within a pixel) approach
- 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



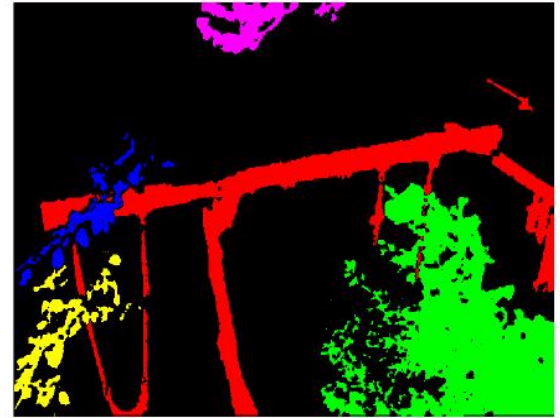
(a) Registered input frame

+



(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)

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

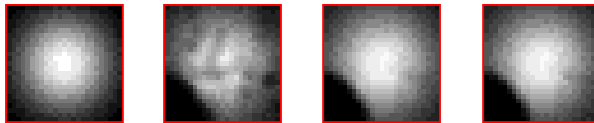
- SSD patch matching search



Confused by
occlusion !

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

region of
support



time

Feature tracking trick 2: trajectory pruning

- Tracking with adaptive region of support Nonsense at full occlusion!



- Outlier detection and removal by interpolation



Comparison

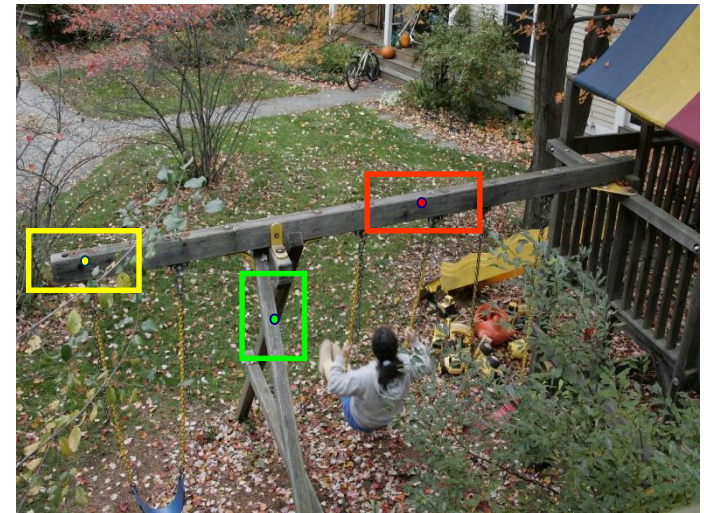


With adaptive region of support and trajectory pruning

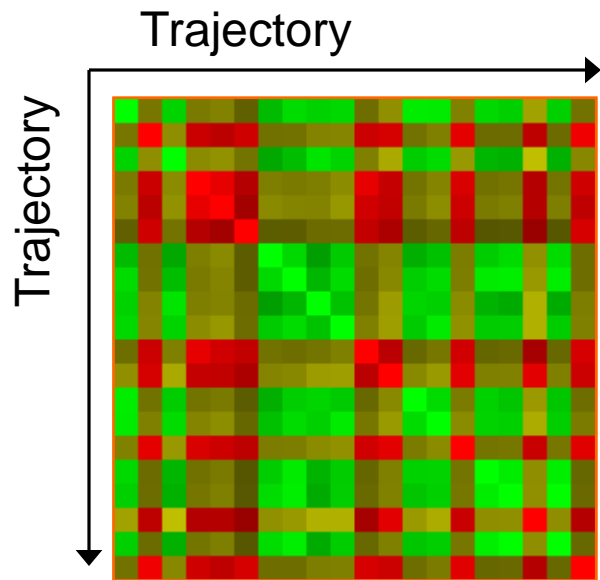
Cluster trajectories based on normalized complex correlation

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

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



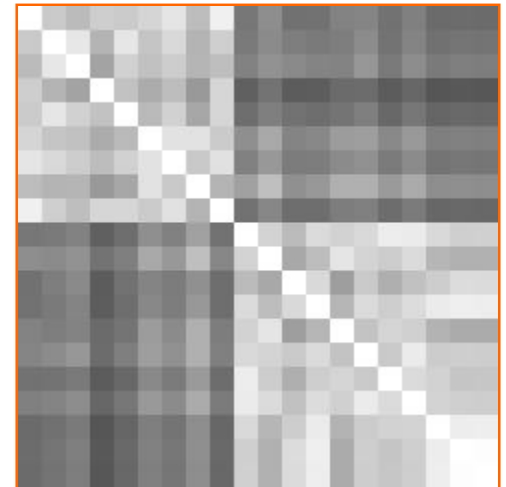
Spectral Clustering



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

Dense optical flow field of cluster 2 (swing) points

Cluster 1: leaves

Cluster 2: swing

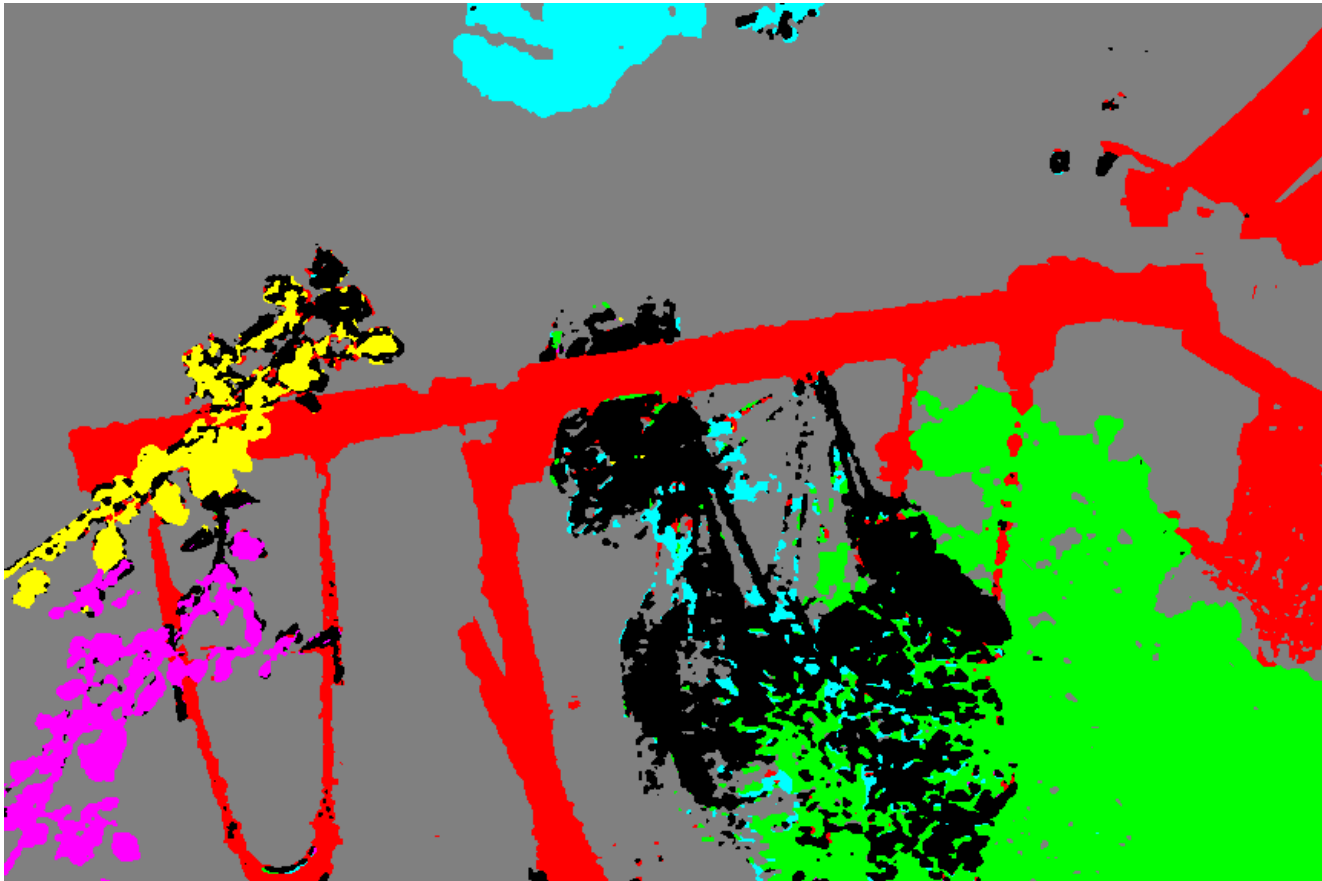


Motion Layer Assignment

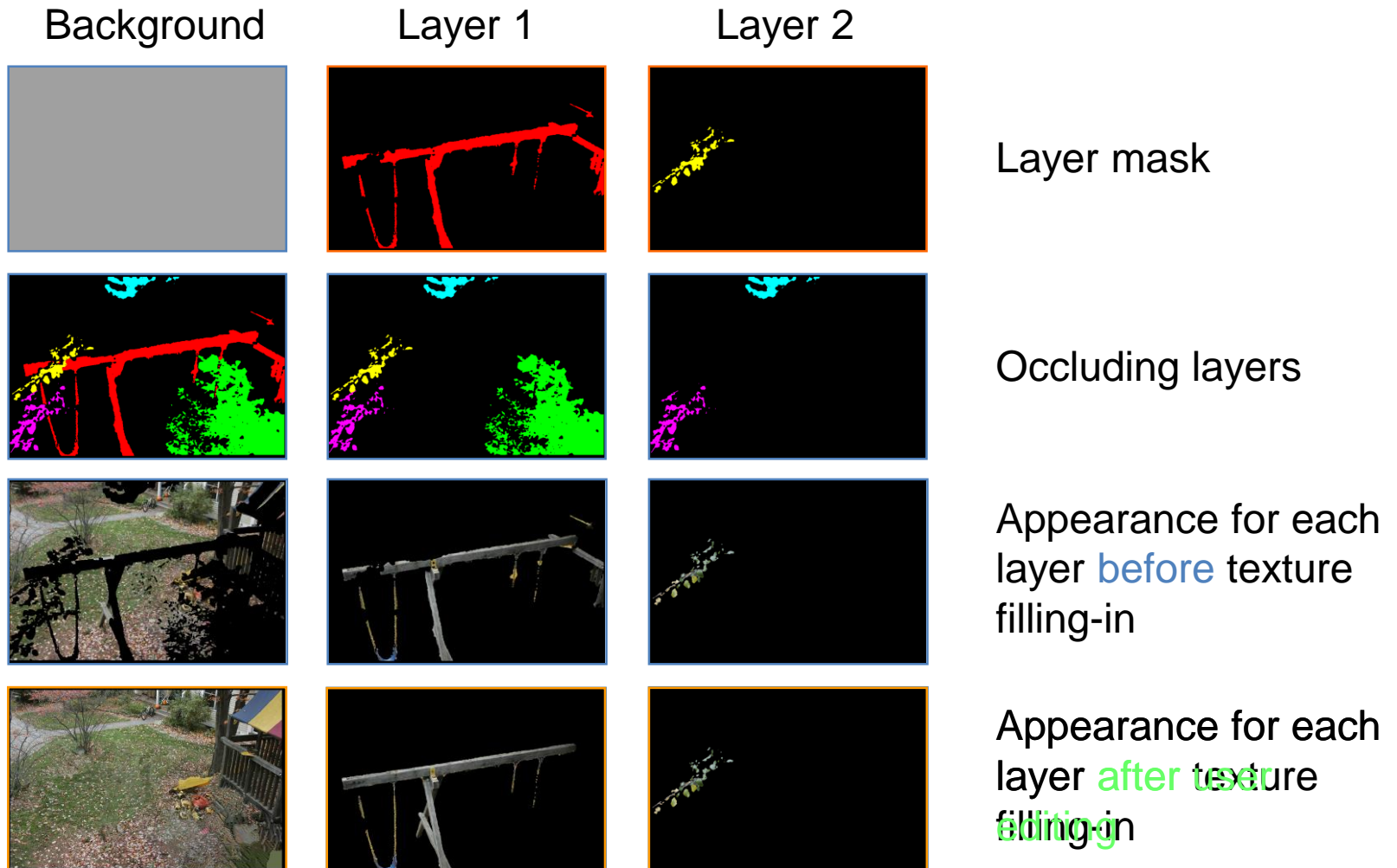
- Assign each pixel to a motion cluster layer, using four cues:
 - **Motion likelihood**—consistency of pixel's intensity if it moves with the motion of a given layer (dense optical flow field)
 - **Color likelihood**—consistency of the color in a layer
 - **Spatial connectivity**—adjacent pixels favored 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



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

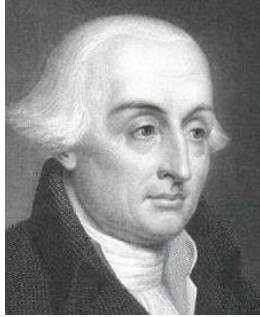
Neal Wadhwa, Michael Rubinstein, Fredo Durand, William T. Freeman
ACM Transactions on Graphics, Volume 32, Number 4 (Proc. SIGGRAPH) 2013

Following slides based on Siggraph presentations:

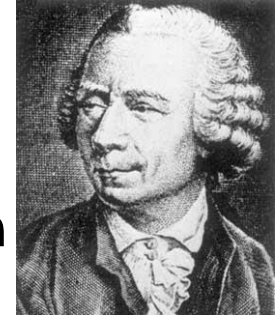
<http://people.csail.mit.edu/mrub/vidmag/>

<http://people.csail.mit.edu/nwadhwa/phase-video/>

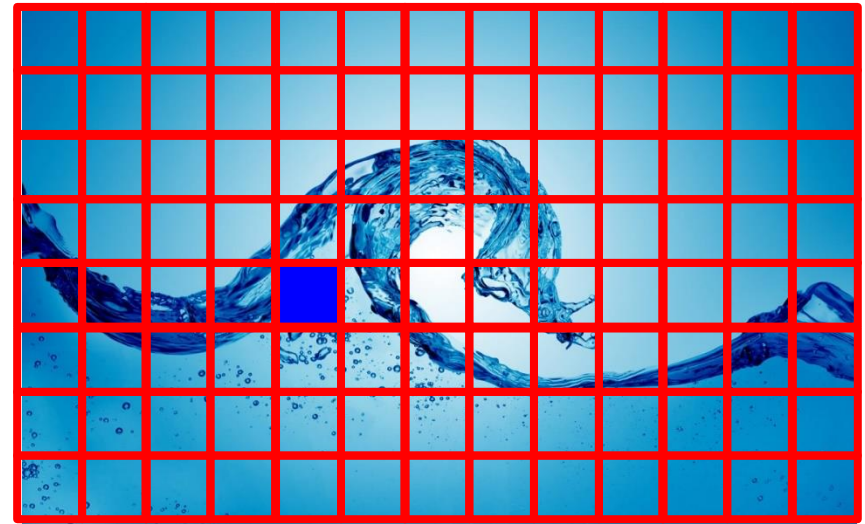
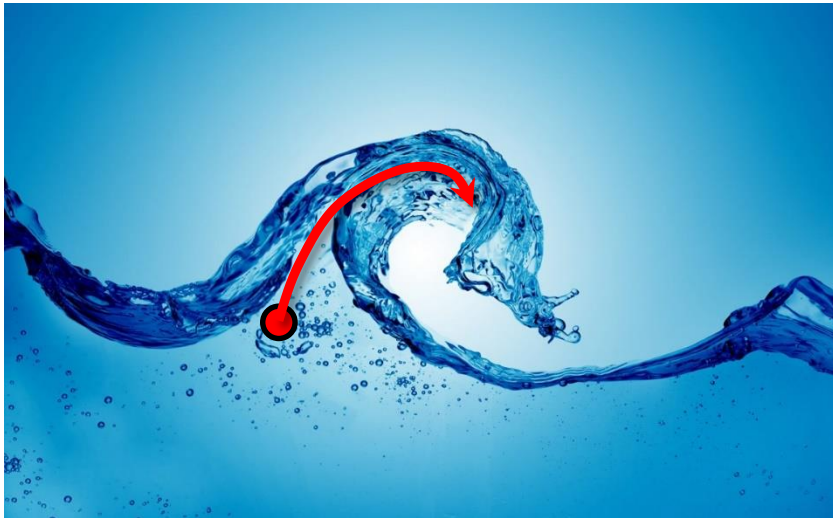
Lagrangian and Eulerian Perspectives: Fluid Dynamics



Lagrangian



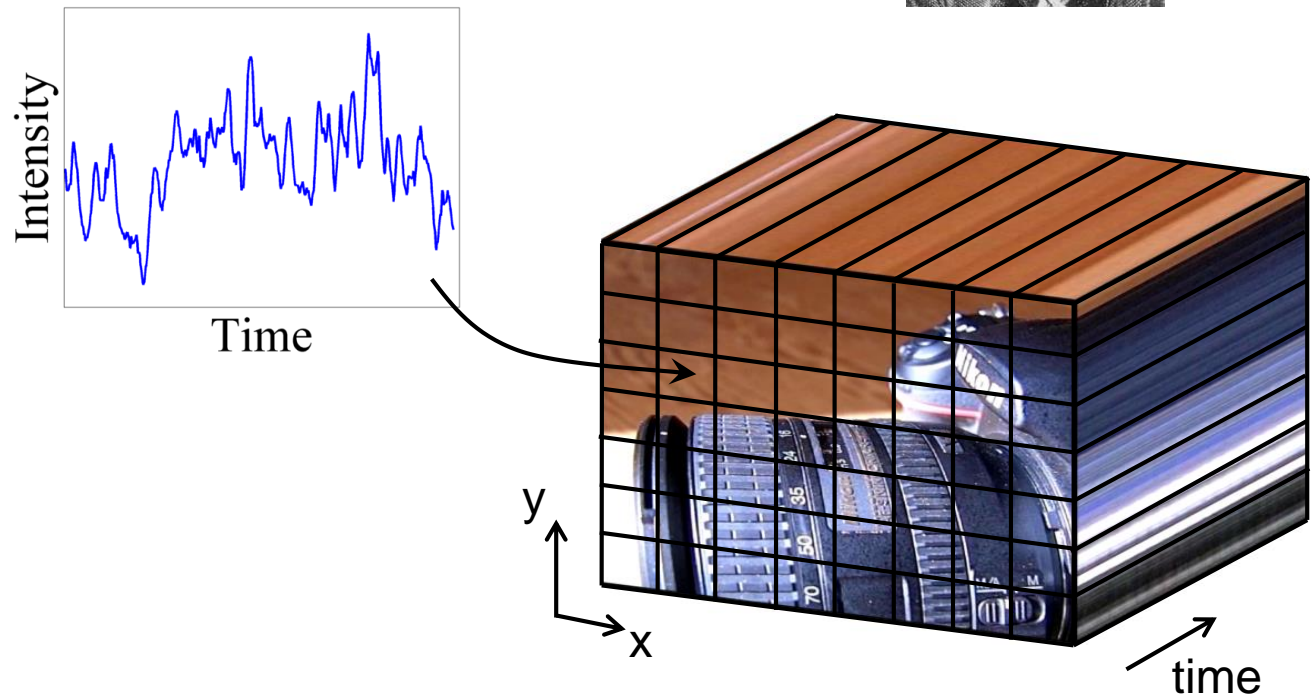
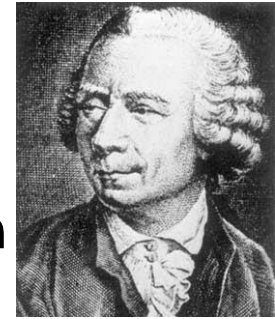
Eulerian



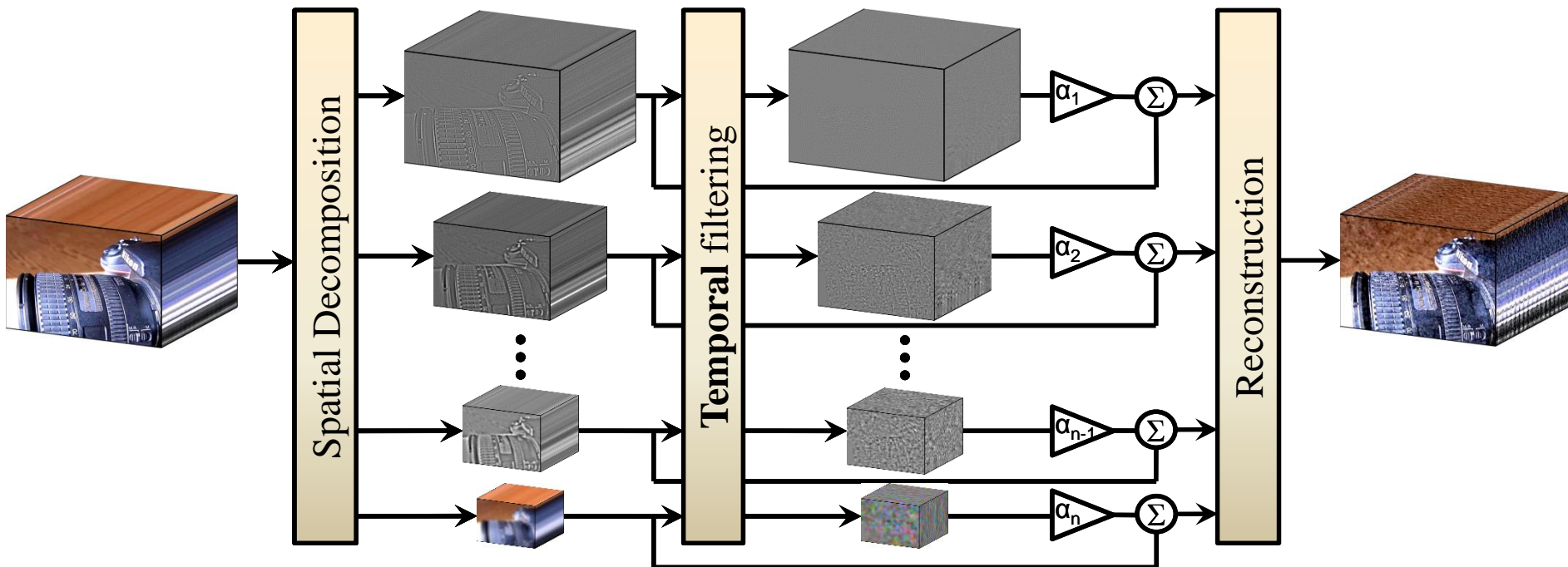
Eulerian Perspective: Videos

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

Eulerian



Method Overview



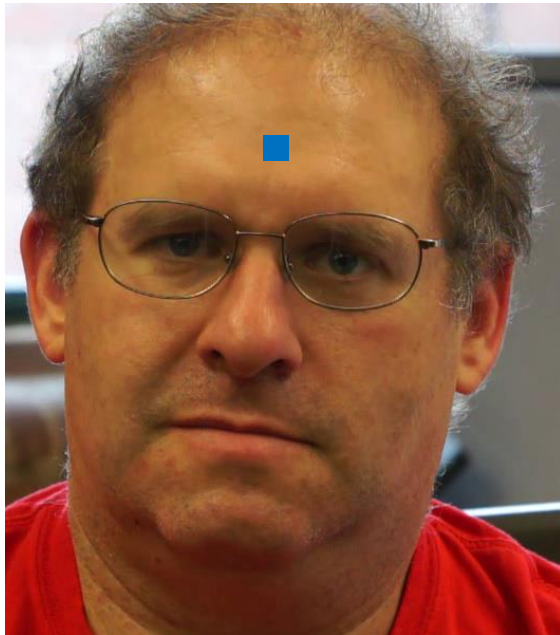
Laplacian
Pyramid

Bandpass filter
intensity at each
pixel over time

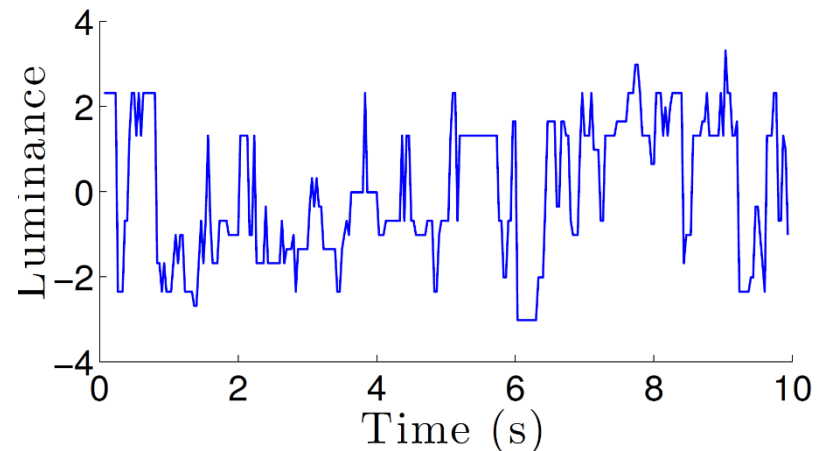
Amplify
bandpassed
signal and add
back to original

Subtle Color Variations

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



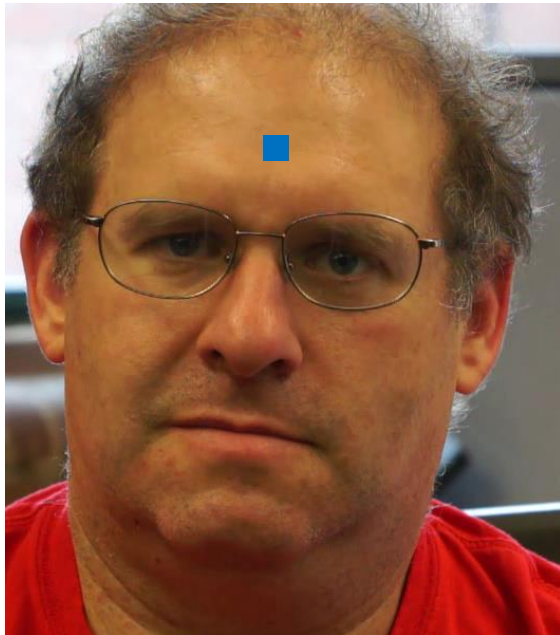
Input frame



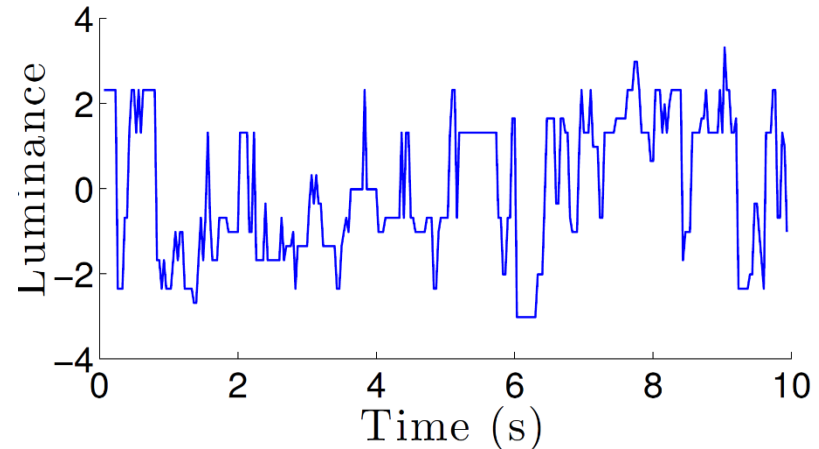
Luminance trace (zero mean)

Subtle Color Variations

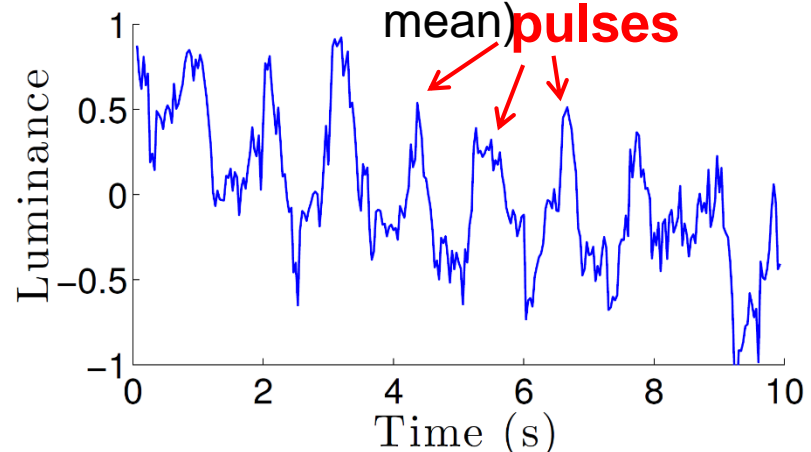
1. Average spatially to overcome sensor and quantization noise



Input frame



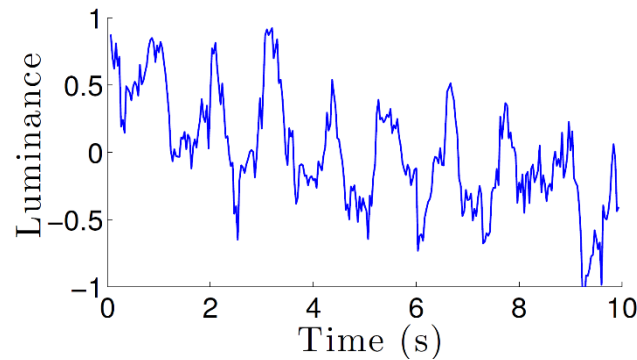
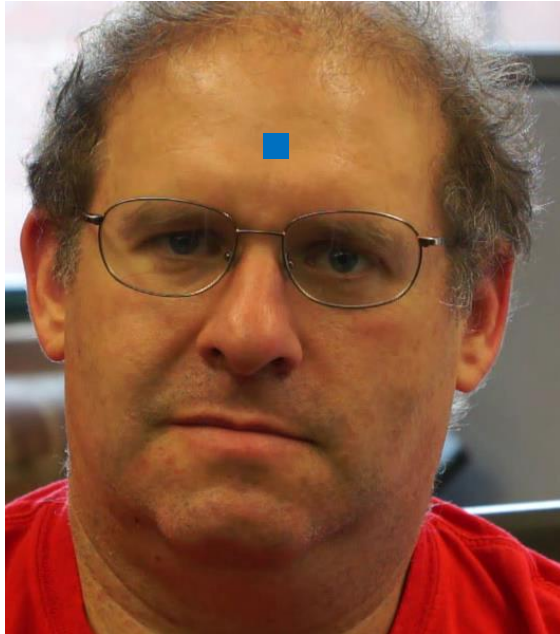
Luminance trace (zero mean)



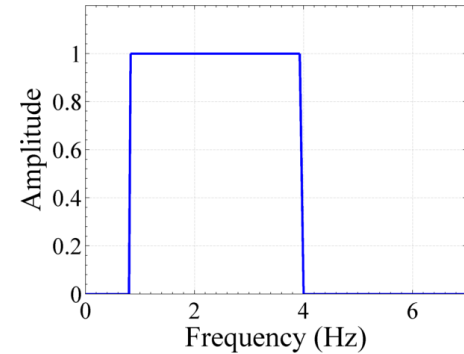
Spatially averaged luminance trace

Amplifying Subtle Color Variations

2. Filter temporally to extract the signal of interest

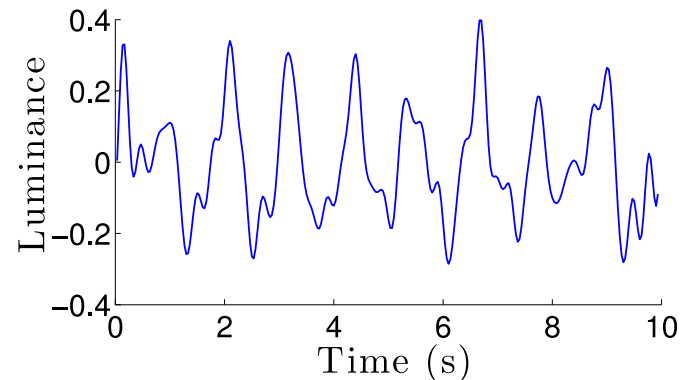


Spatially averaged luminance trace



Temporal filter

=



Temporally bandpassed trace

Color Amplification Results



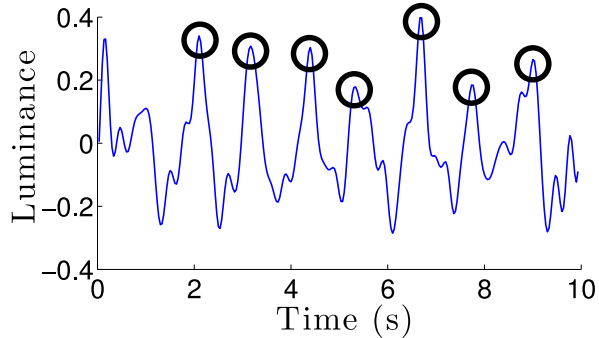
Source



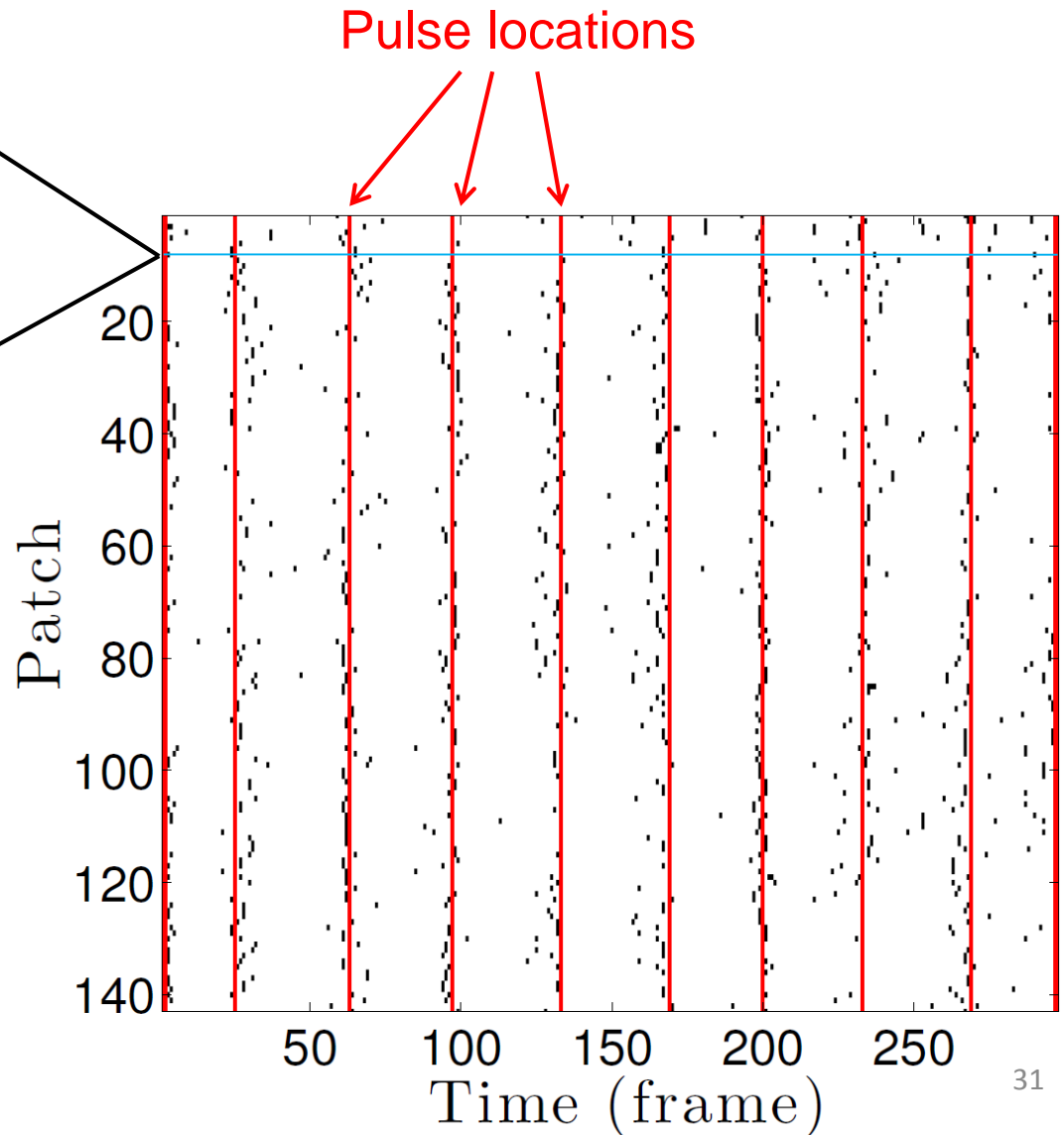
Color-amplified (x100)
0.83-1 Hz (50-60 bpm)

Heart Rate Extraction

Peak detection



Temporally bandpassed trace
(one pixel)



Heart Rate Extraction



Source

(Courtesy of Winchester Hospital. Do not copy)



← EKG

Hospital monitor

Bandpass signal +
peaks (pulse)



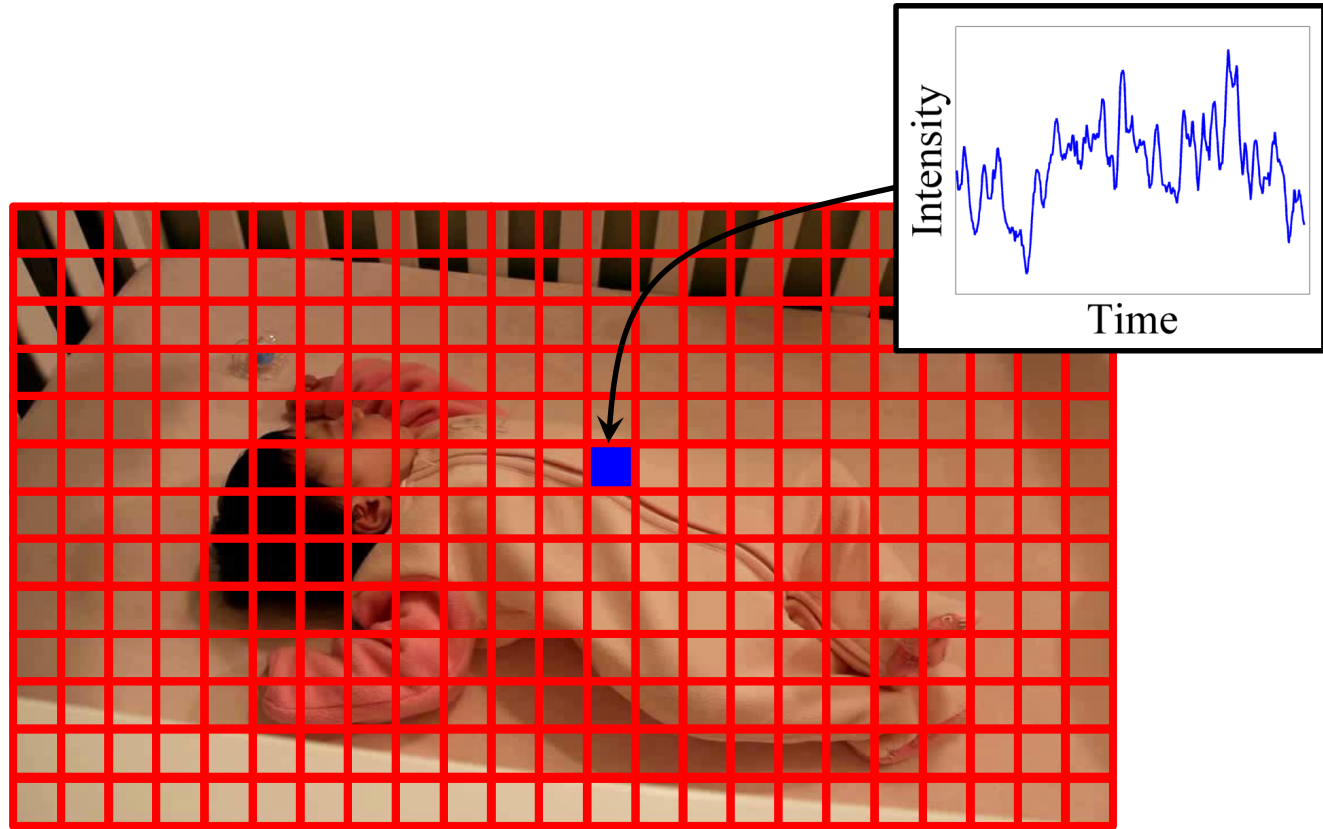
Estimated heart rate



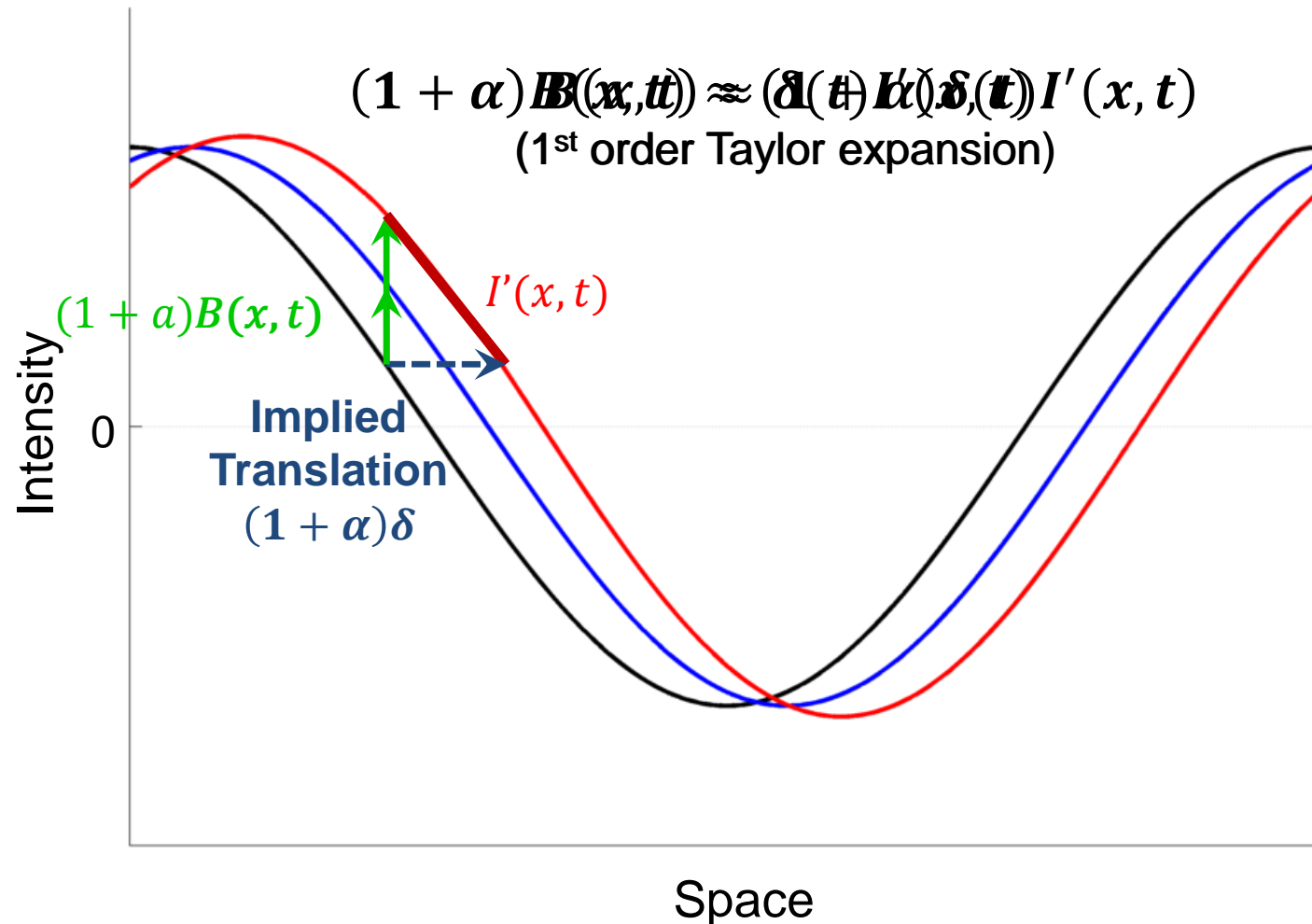
Color-amplified (x150)

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

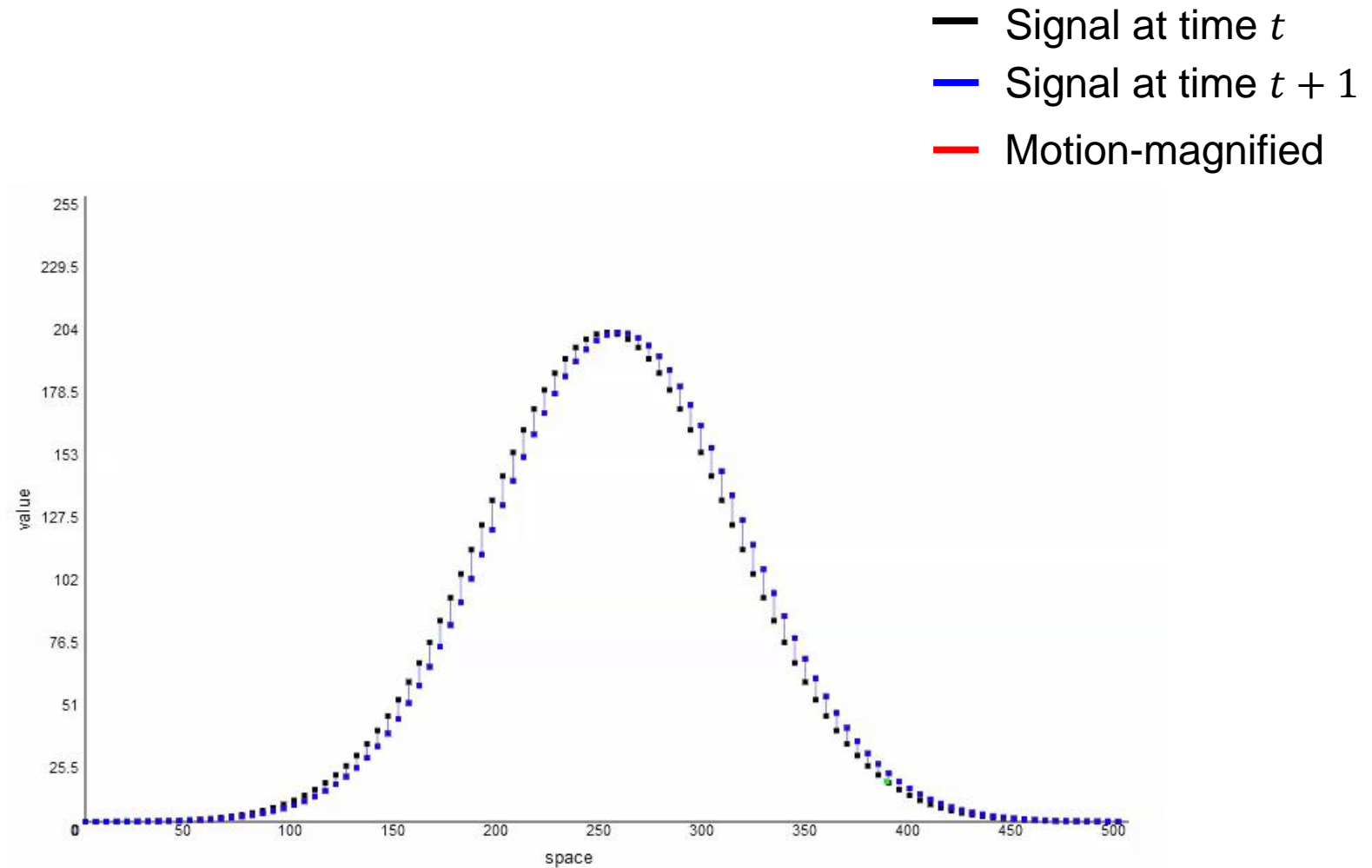
Why It Amplifies Motion



Relating Temporal and Spatial Changes

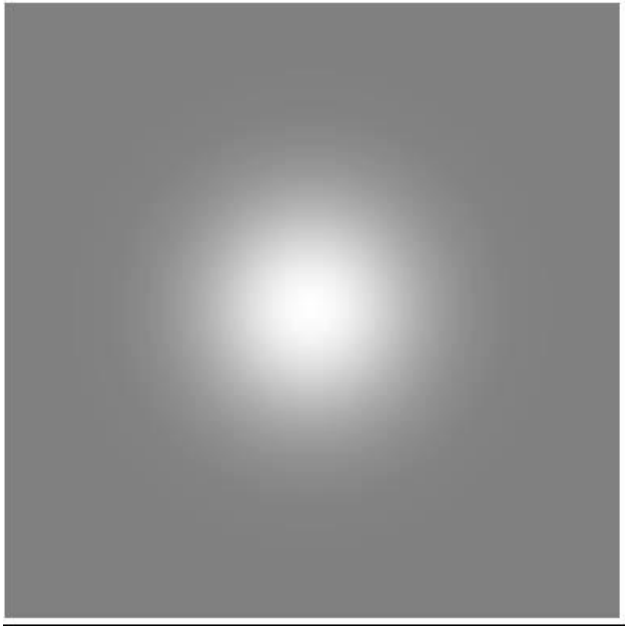


Relating Temporal and Spatial Changes



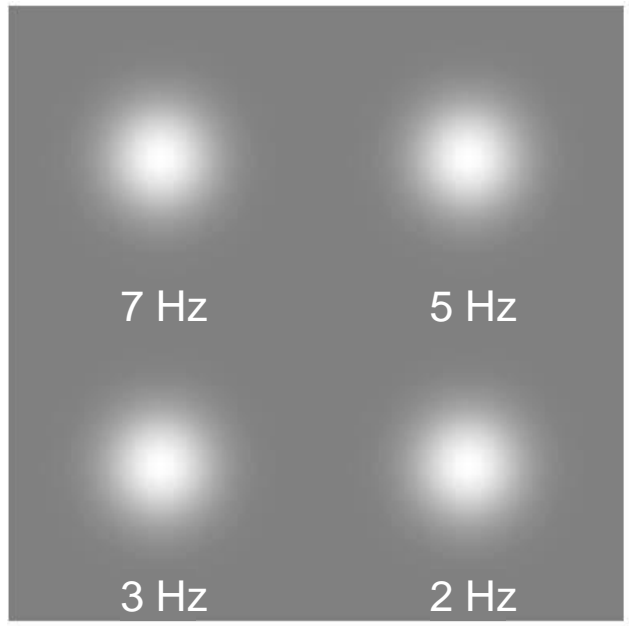
Courtesy of Lili Sun

Synthetic 2D Example



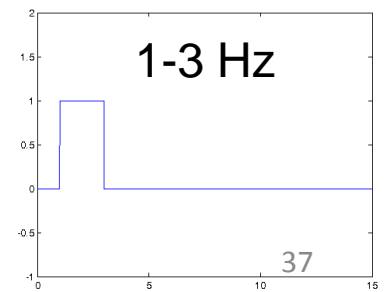
Source

Selective Motion Magnification

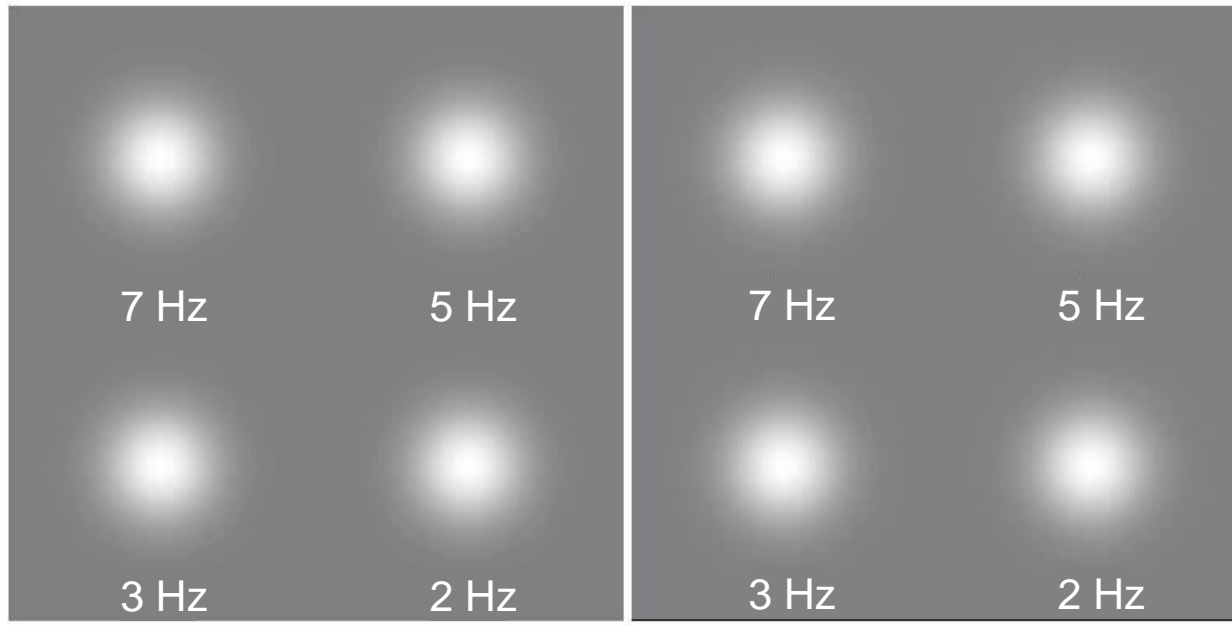


Source
(Single video with 4 blobs)

Temporal filter:



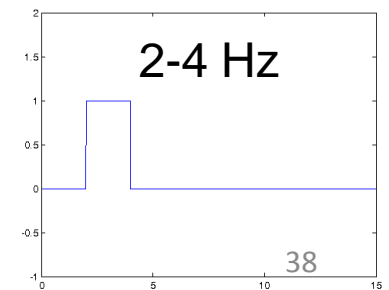
Selective Motion Magnification



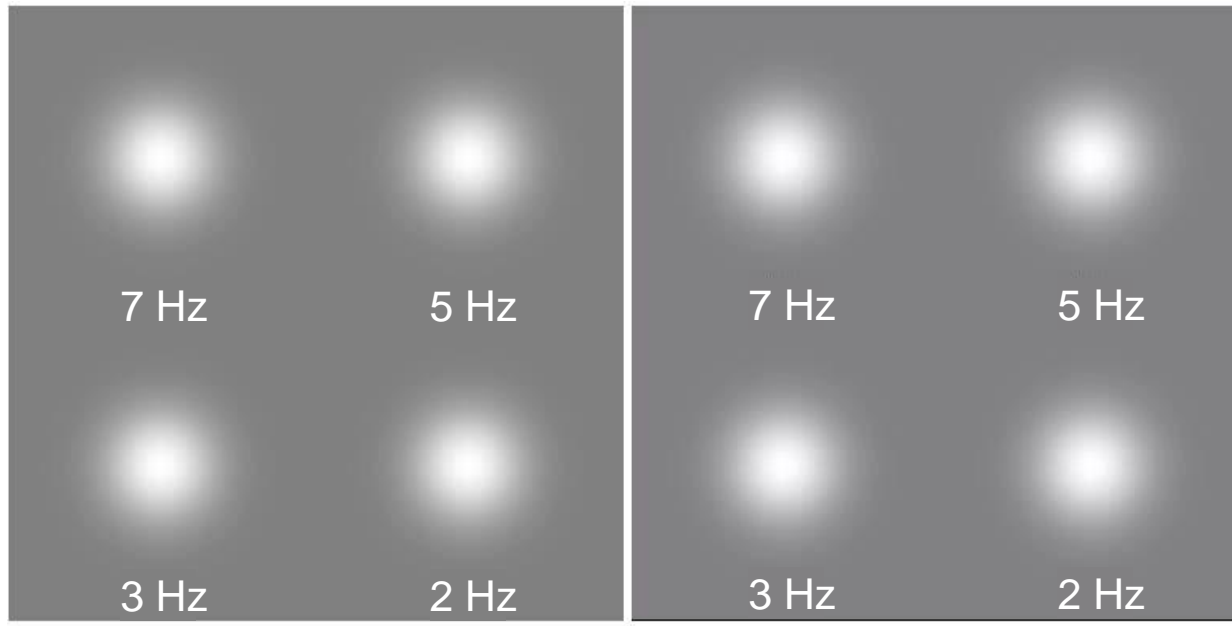
Source
(Single video with 4 blobs)

Motion-magnified (3 Hz)

Temporal filter:



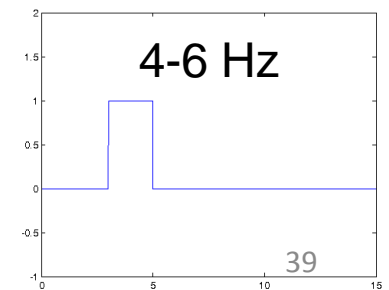
Selective Motion Magnification



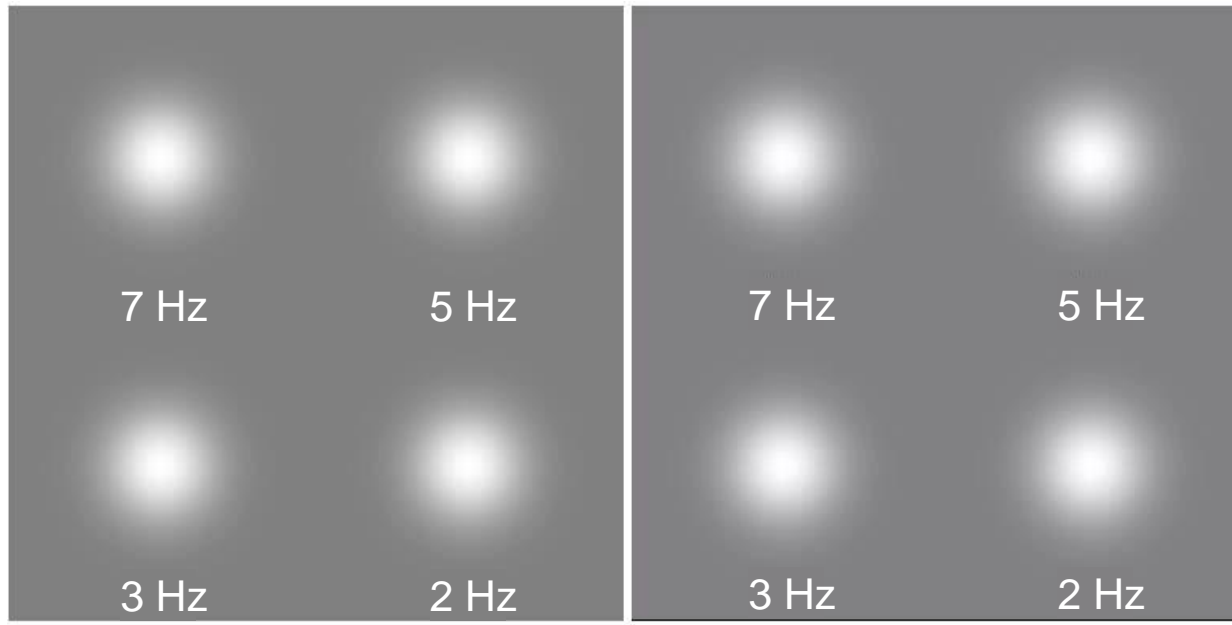
Source
(Single video with 4 blobs)

Motion-magnified (5 Hz)

Temporal filter:



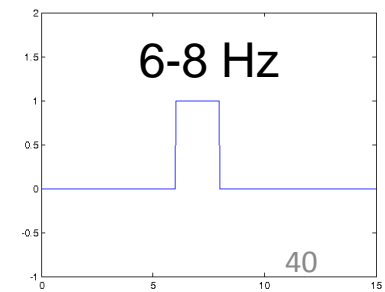
Selective Motion Magnification



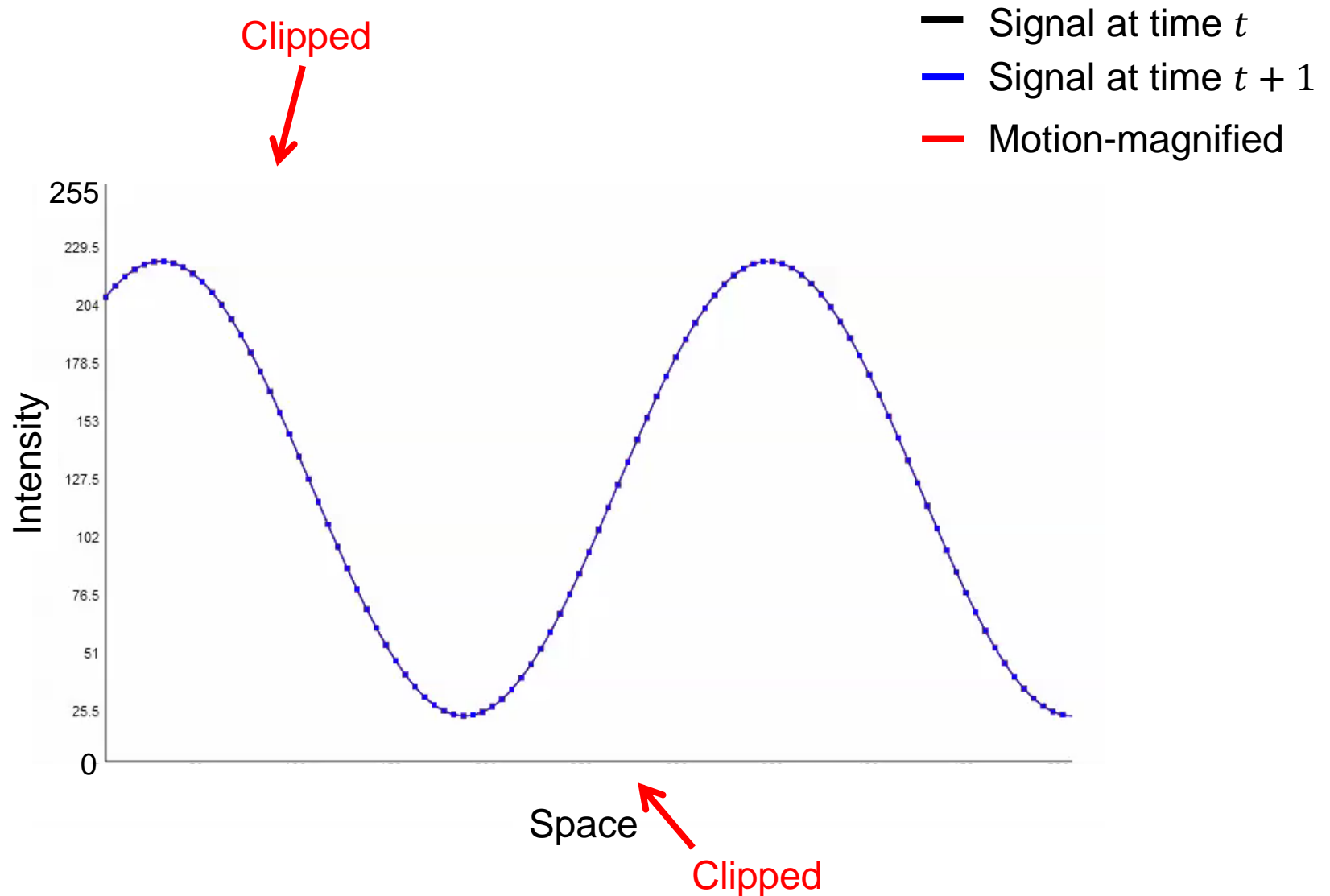
Source
(Single video with 4 blobs)

Motion-magnified (7 Hz)

Temporal filter:



When Does It Break?



Motion Magnification Artifacts

Artifact



Source

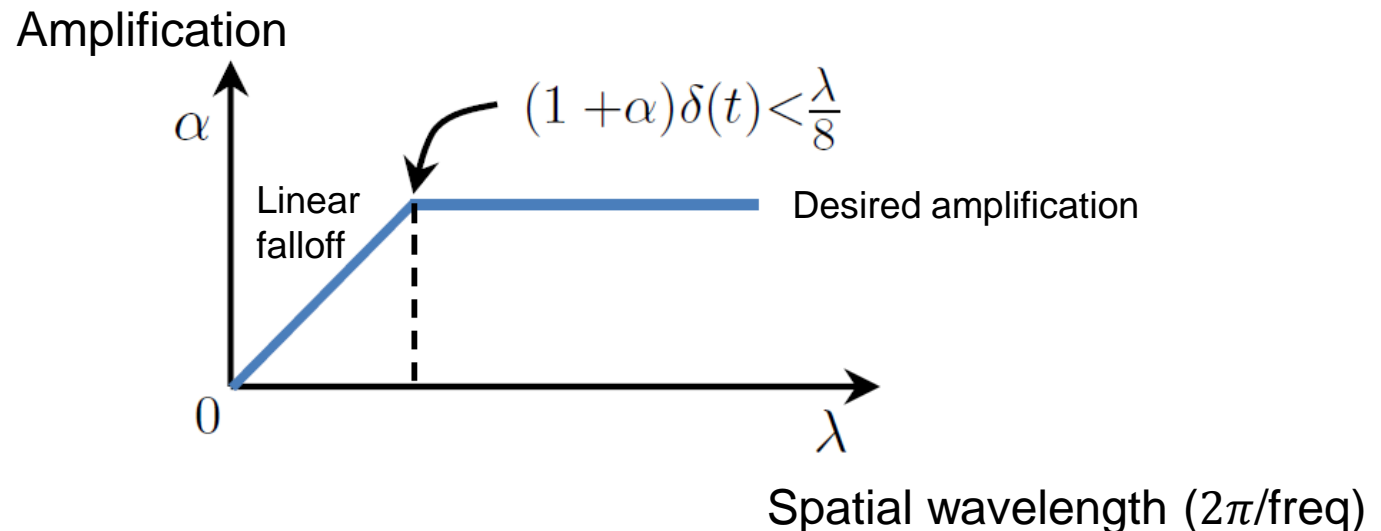


Motion-magnified (3.6-6.2 Hz, x60)

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



Motion Magnification Results



Source

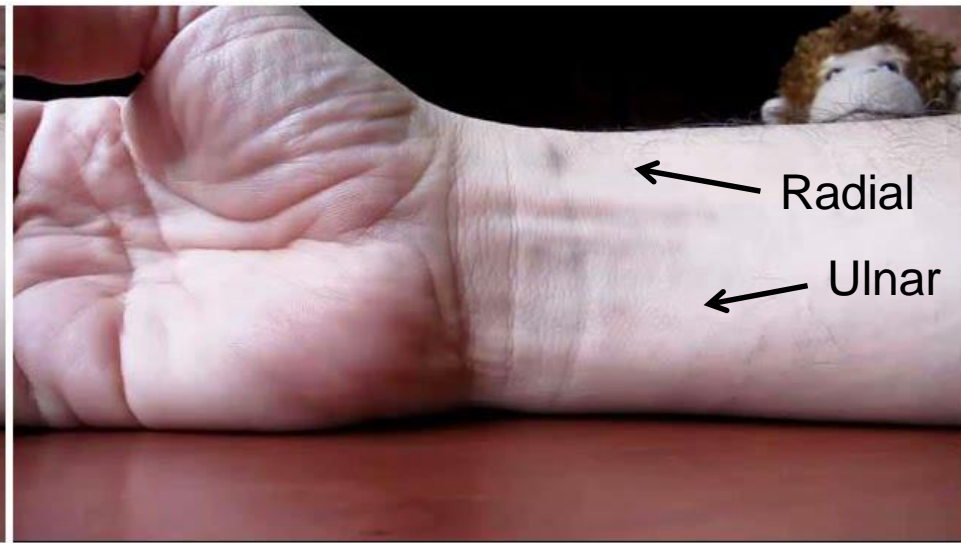


Motion-magnified (0.4-3 Hz, x10)

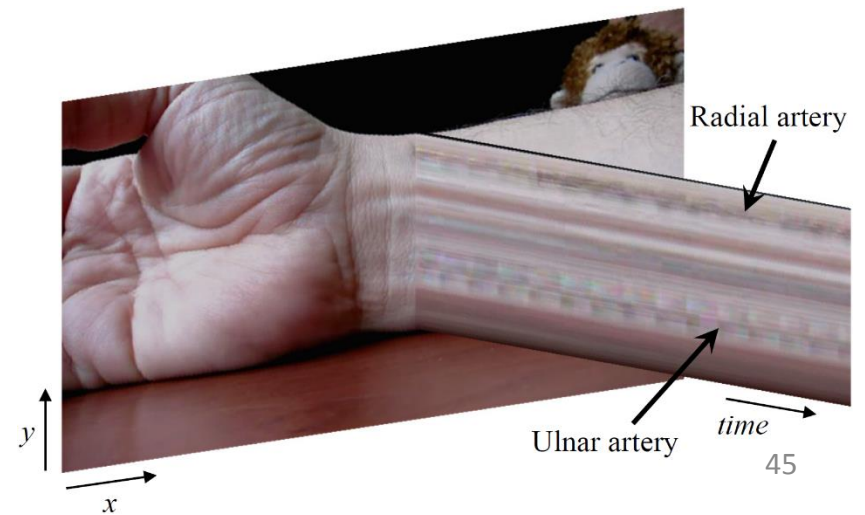
Motion Magnification



Source



Motion-magnified (0.4-3 Hz, x10)

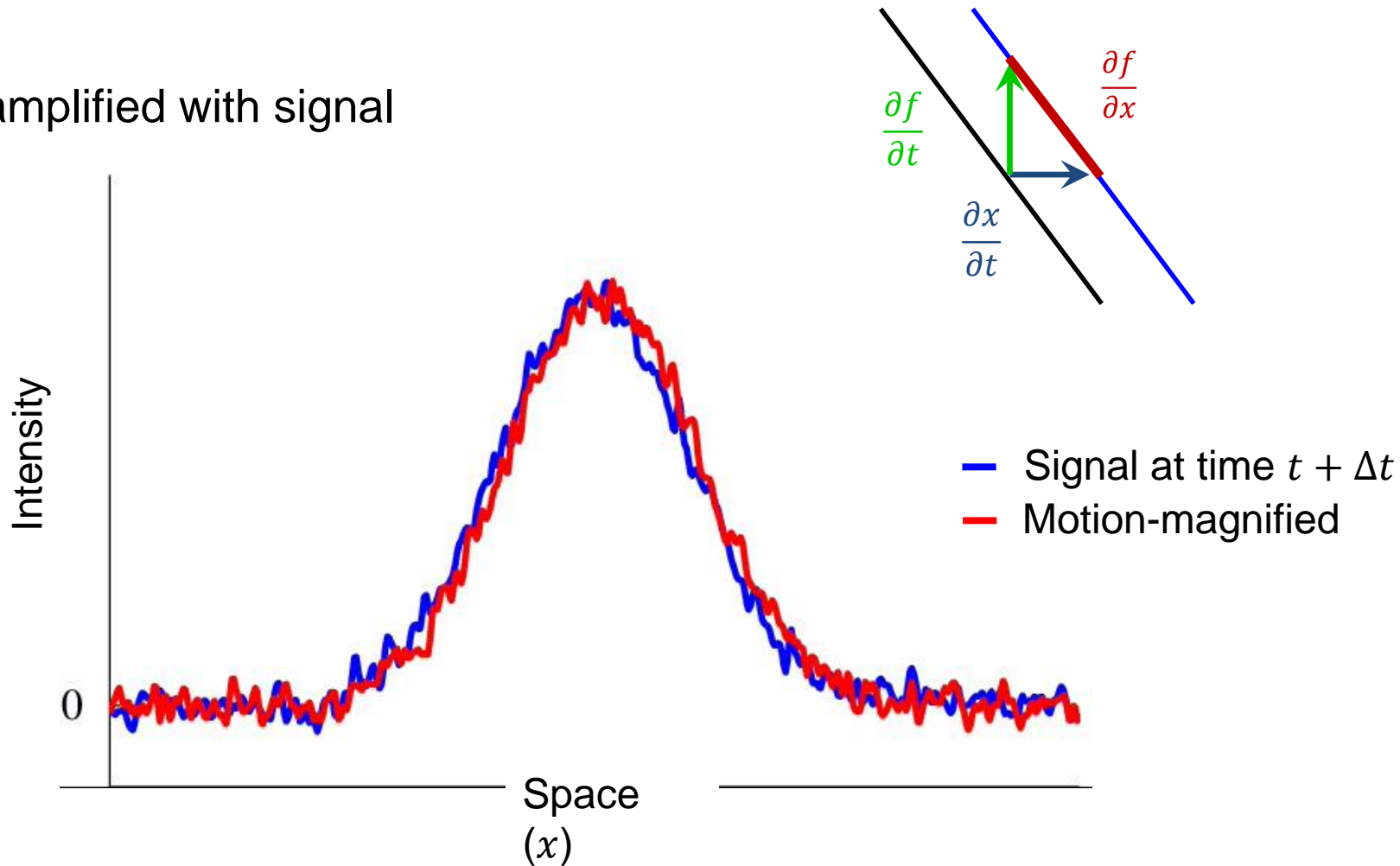


Discussion of pixelwise intensity amplification approach

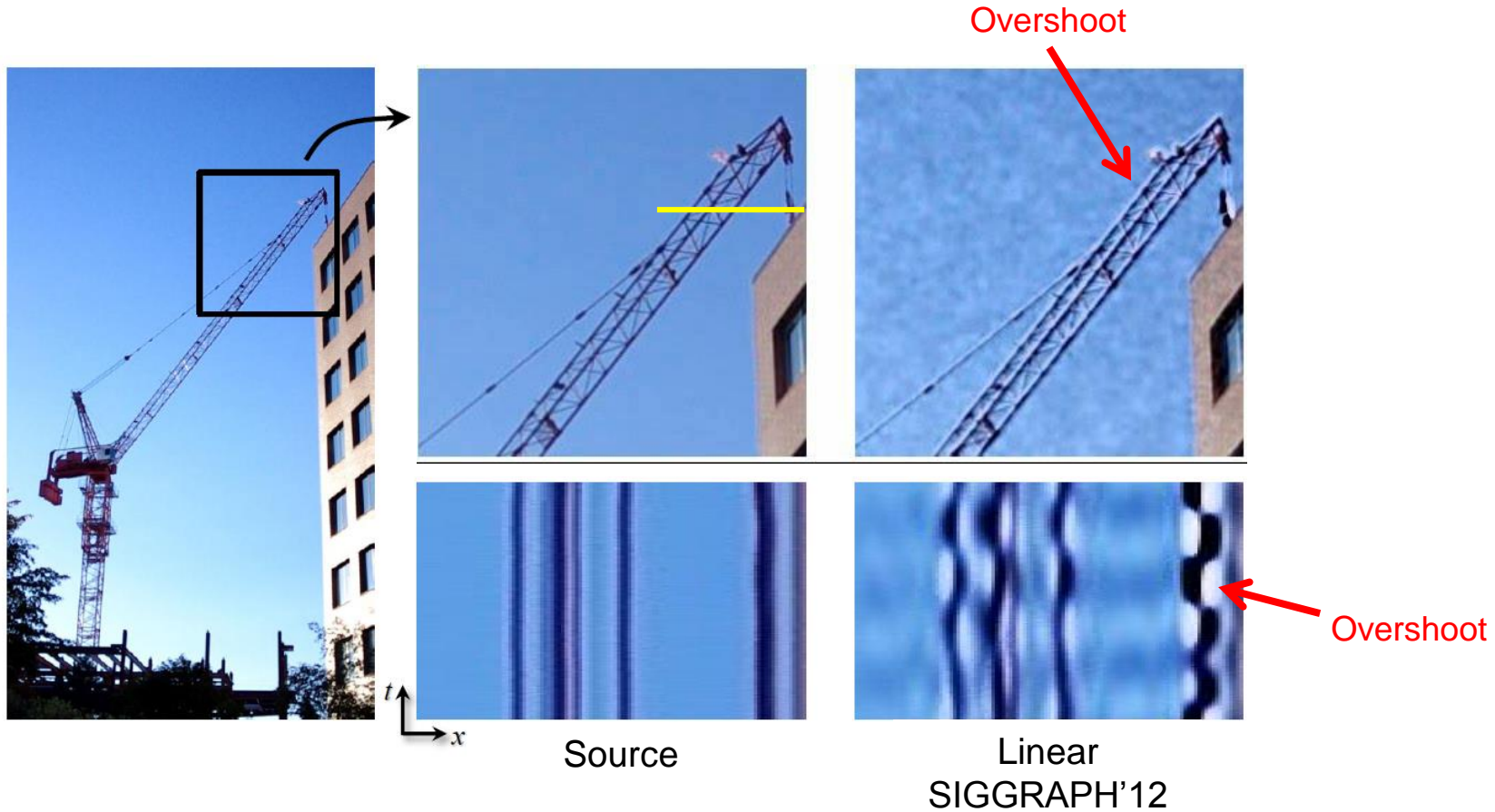
- Good:
 - Does not require explicit motion estimation or texture synthesis (robust)
 - Very fast (real time)
- Bad:
 - Can only handle very small motions
 - Amplifies noise

Limitations of Linear Motion Processing

- Noise amplified with signal



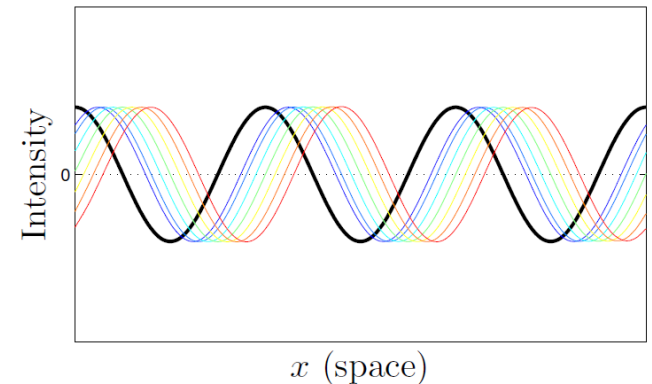
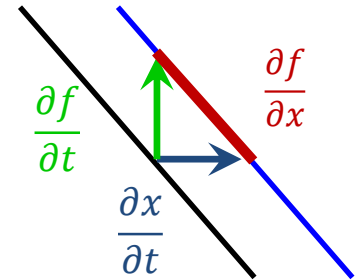
Limitations of Linear Motion Processing



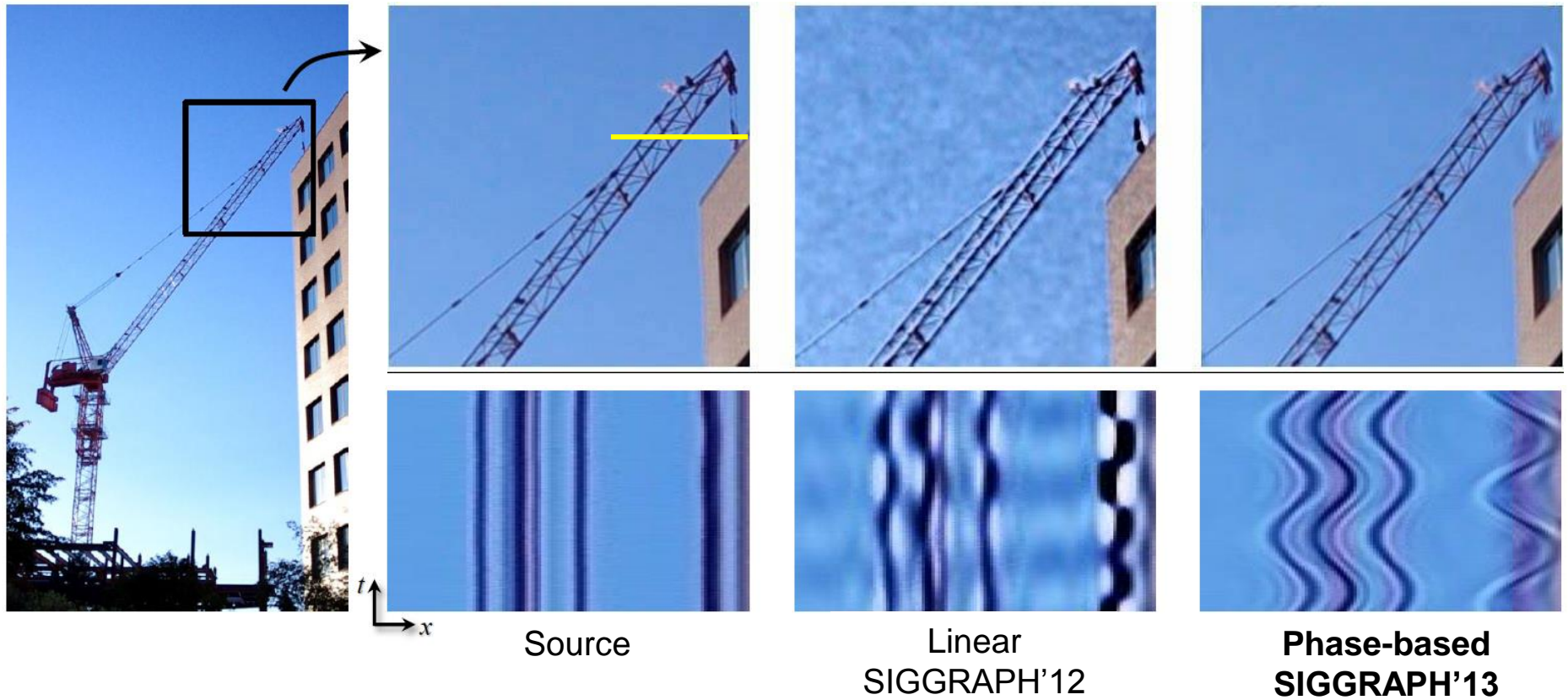
Eulerian approach part 2: shift phase instead of amplifying intensity

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**

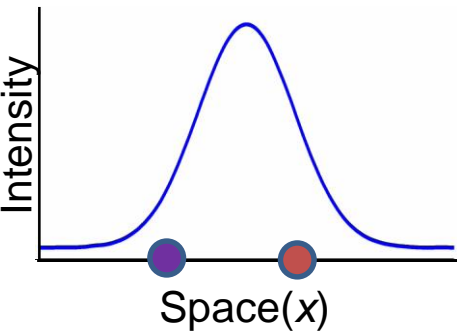


Linear vs. Phase-Based Motion Processing

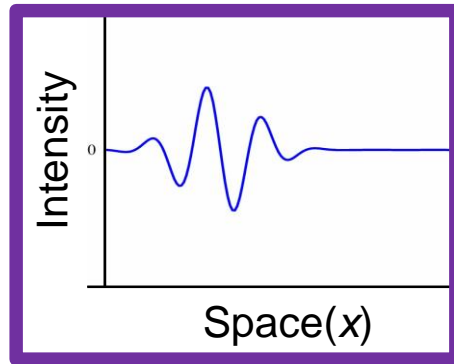


Phase over Time

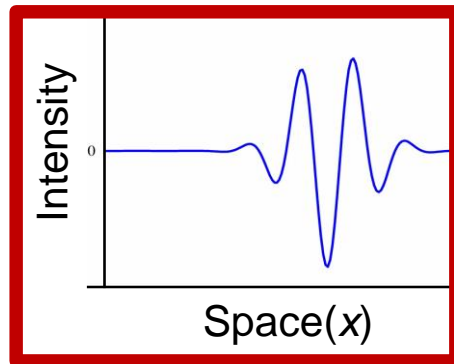
Input



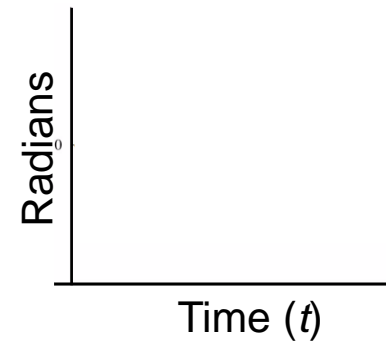
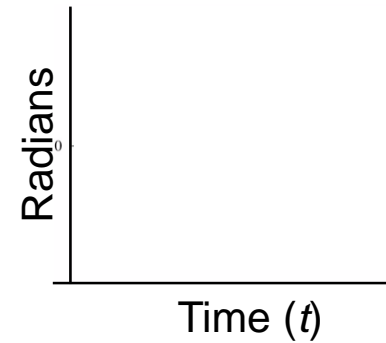
Wavelets



⋮

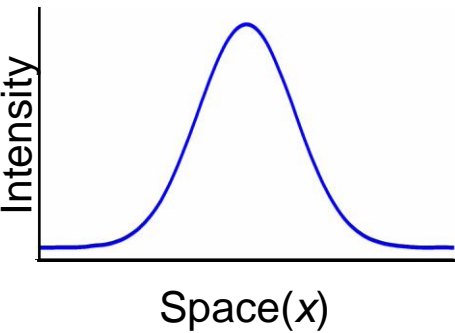


Phase over Time

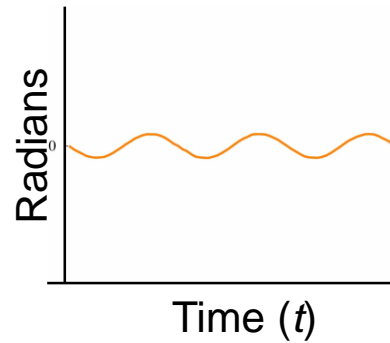


Phase over Time

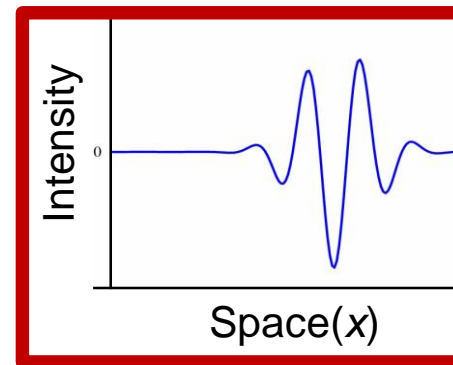
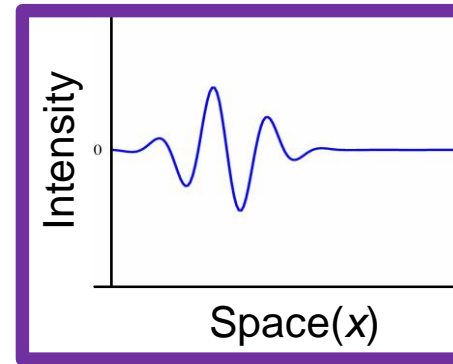
Input



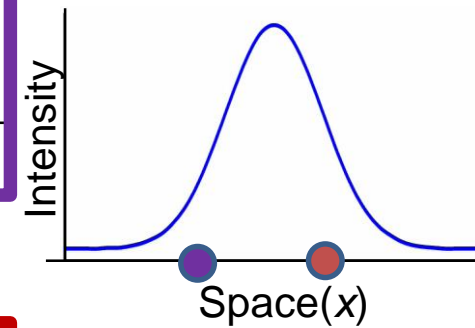
Phase over Time



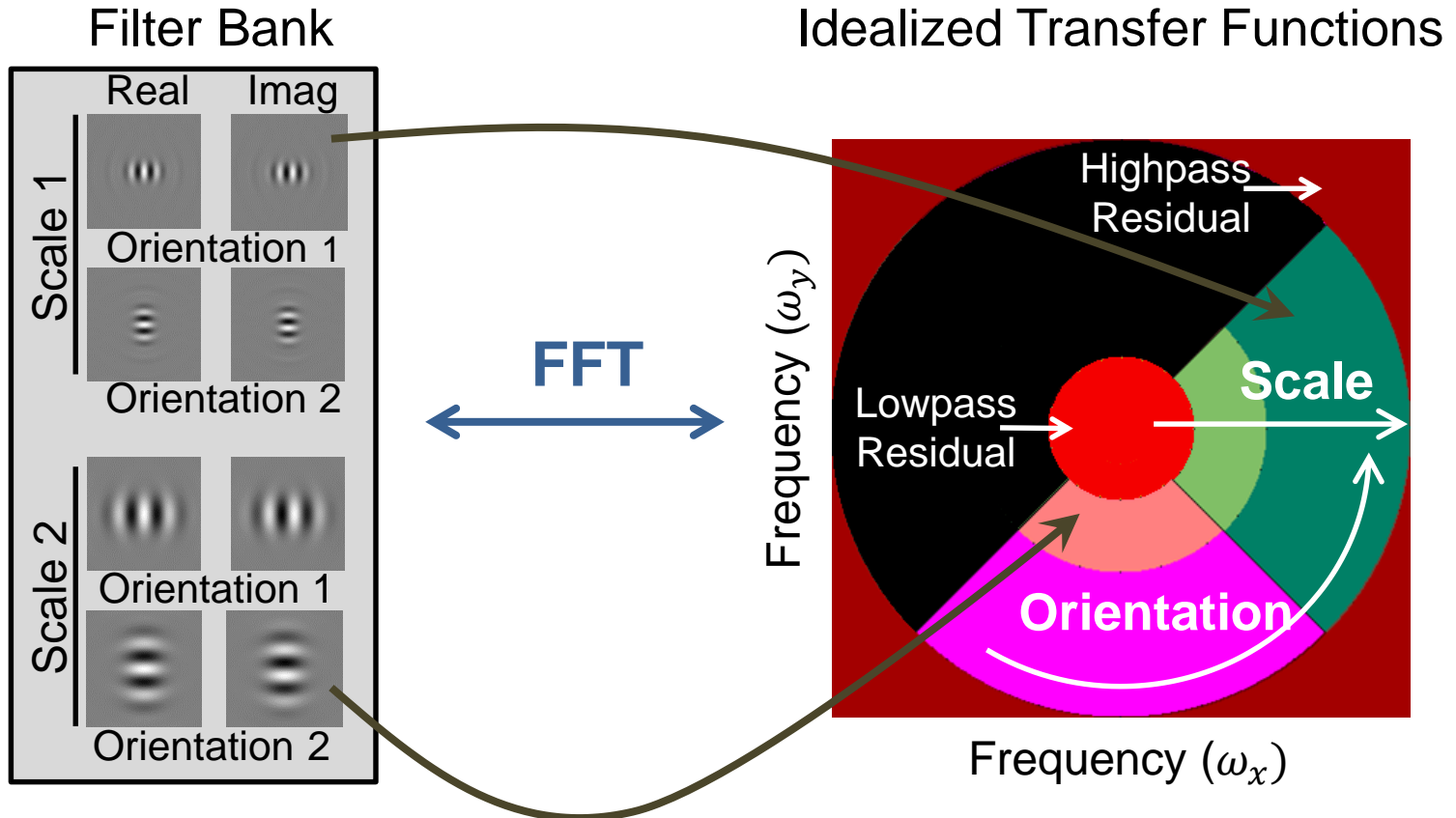
Wavelets



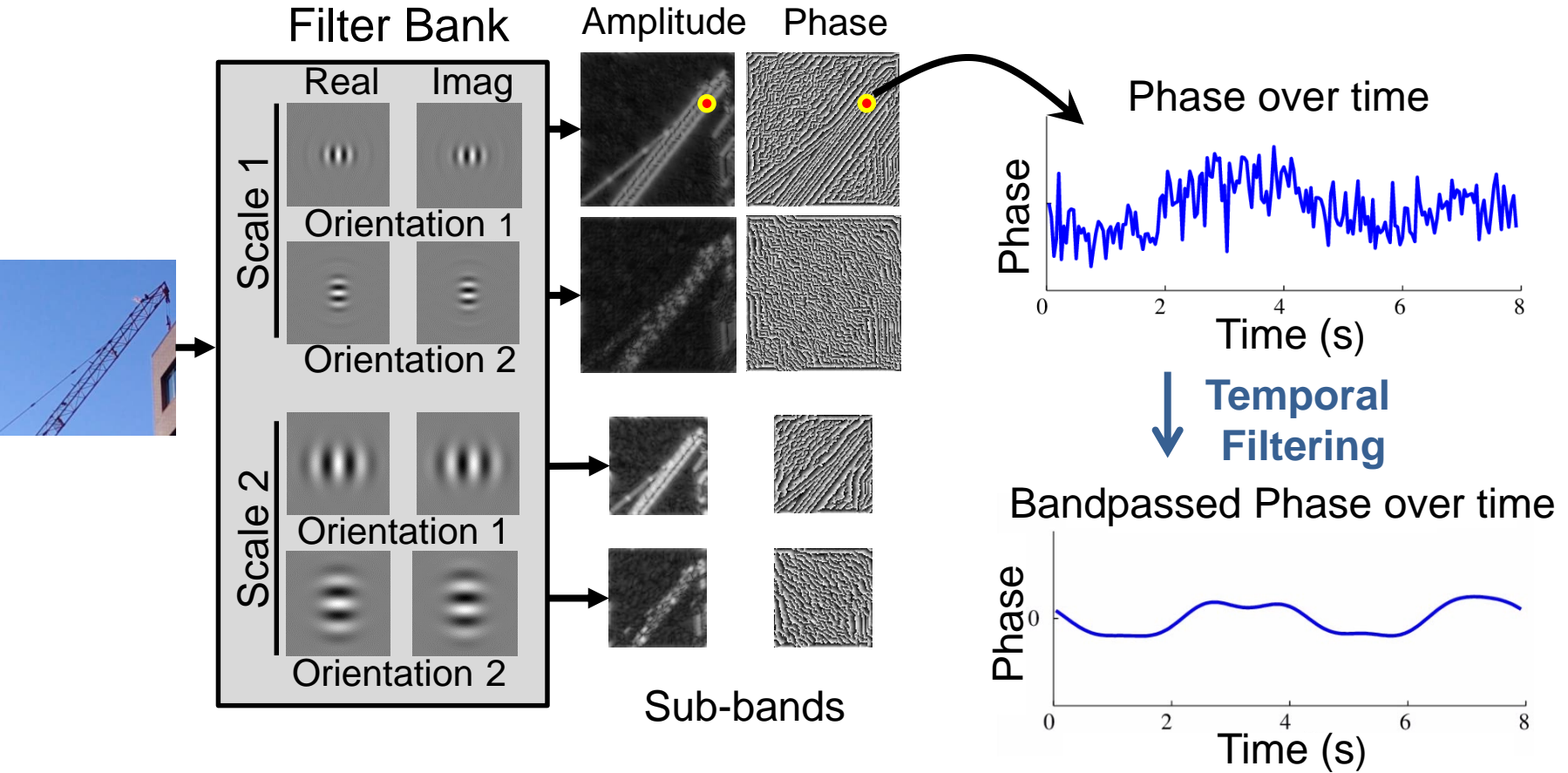
Motion-magnified



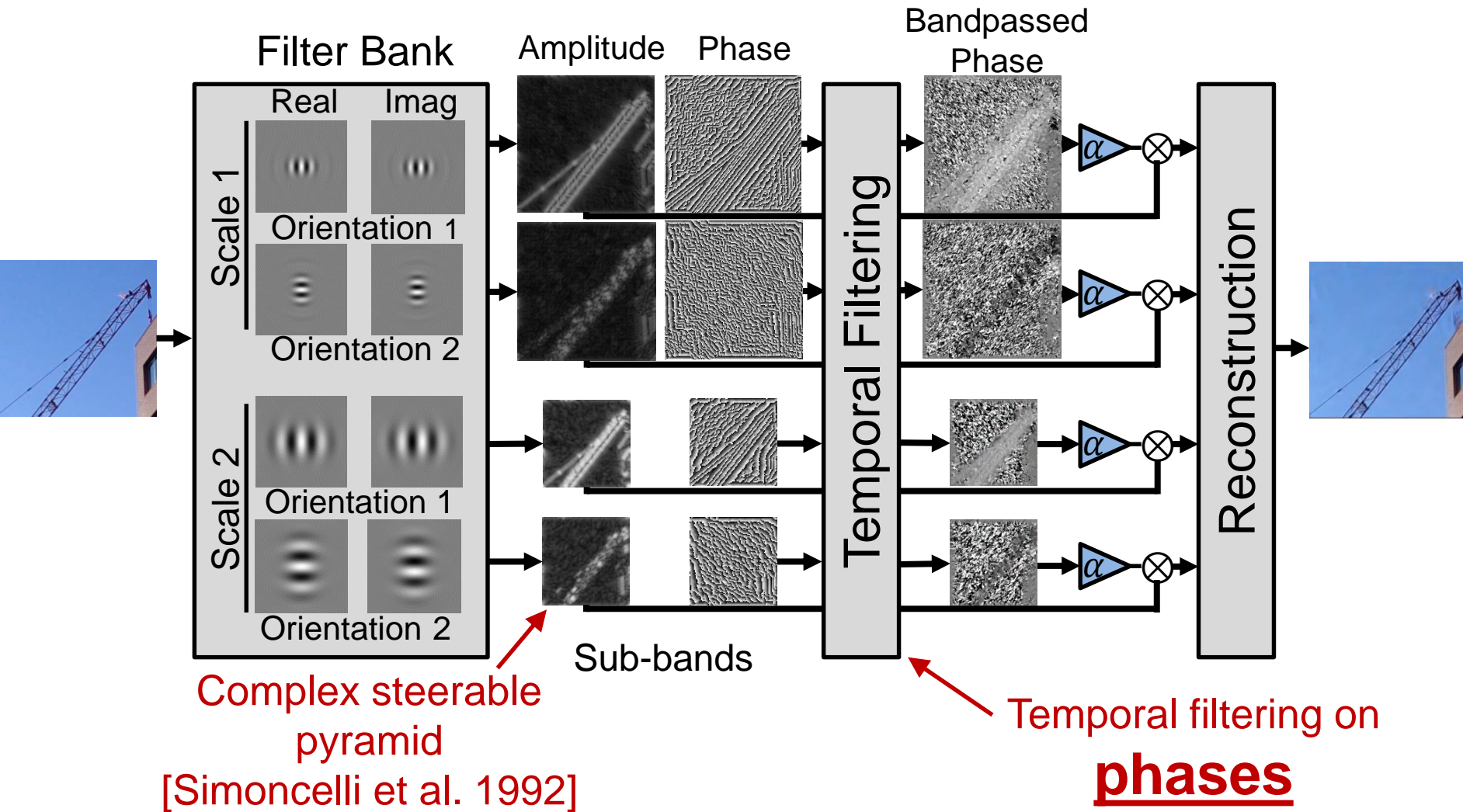
2D Complex Steerable Pyramid



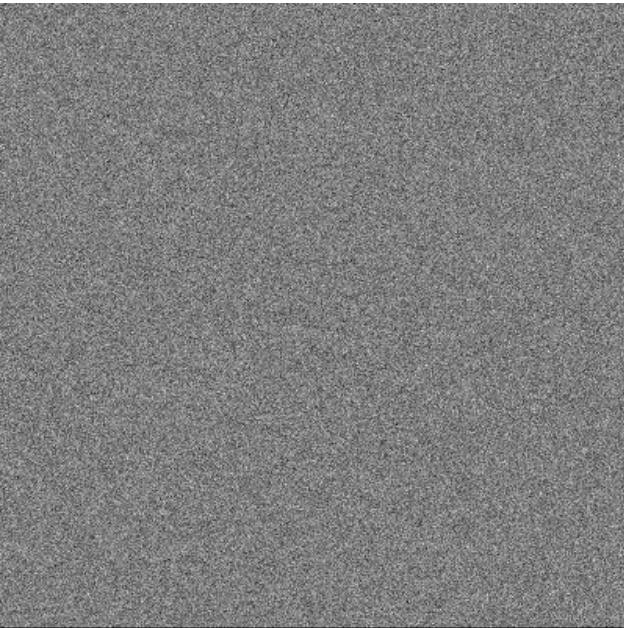
Phase over Time



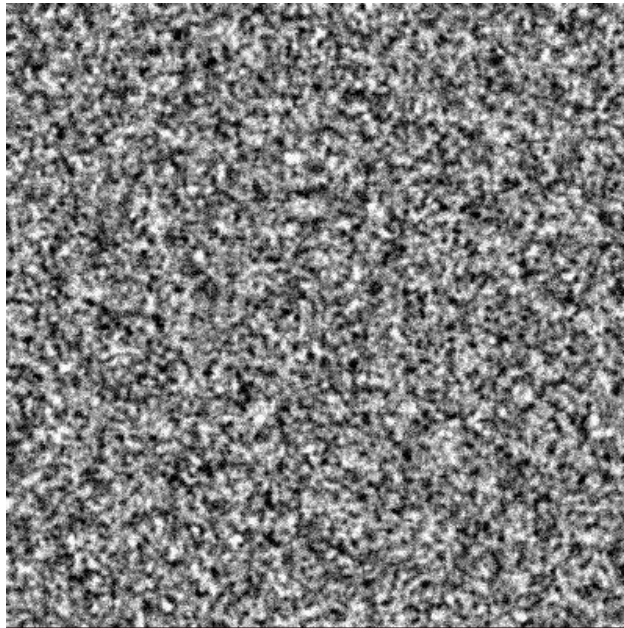
New Phase-Based Pipeline



Improvement #1: Less Noise

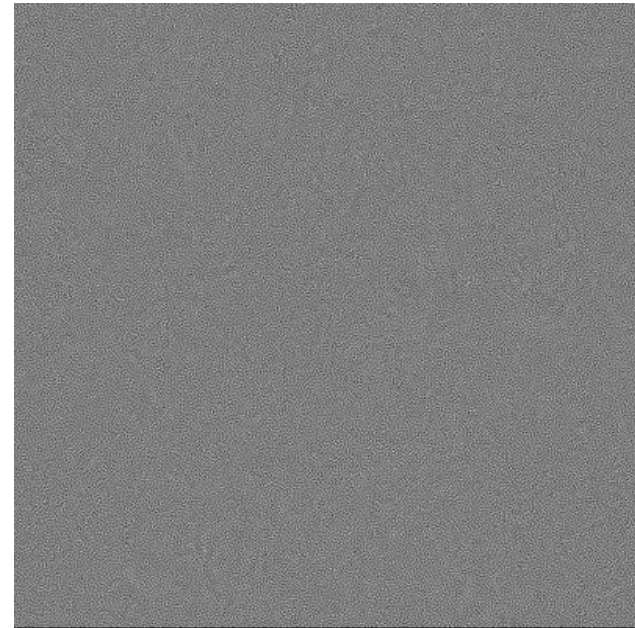


Source (IID
Noise, std=0.1)



Linear [Wu et al.
2012] (x50)

Noise amplified

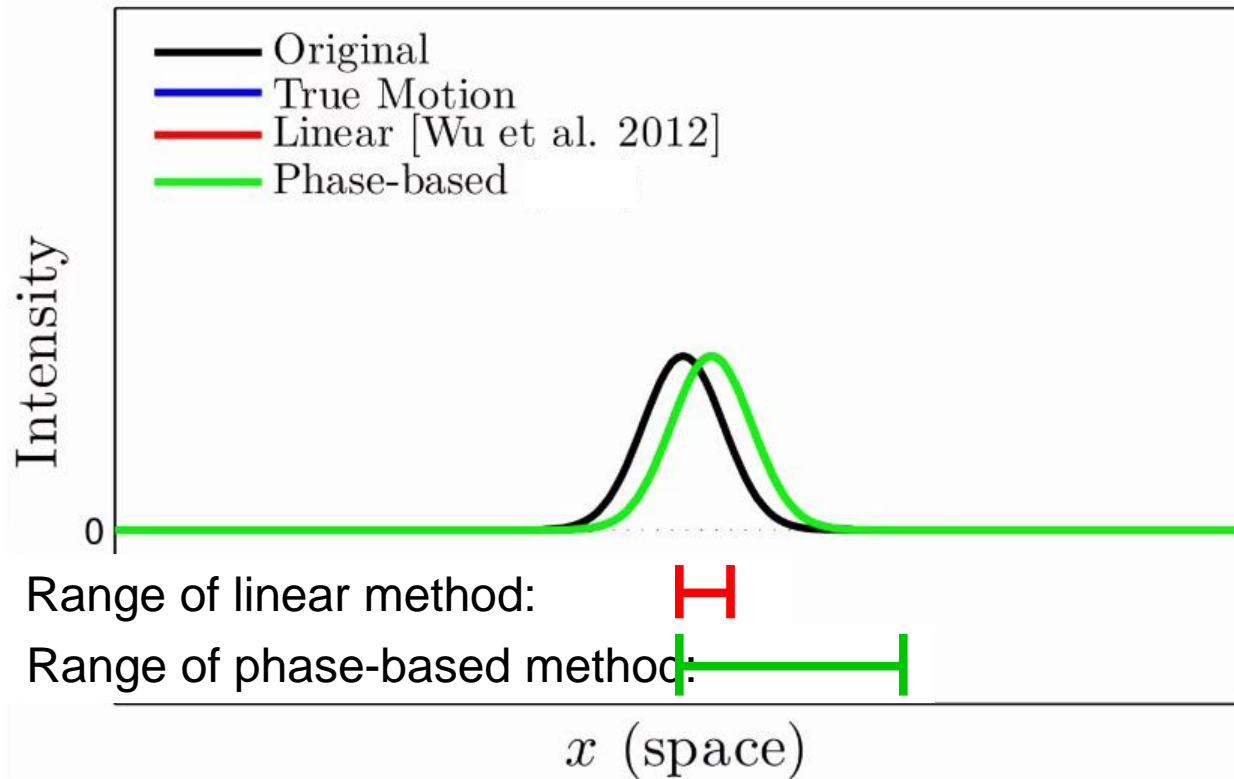


Phase-based
(x50)

Noise translated

Improvement #2: More Amplification

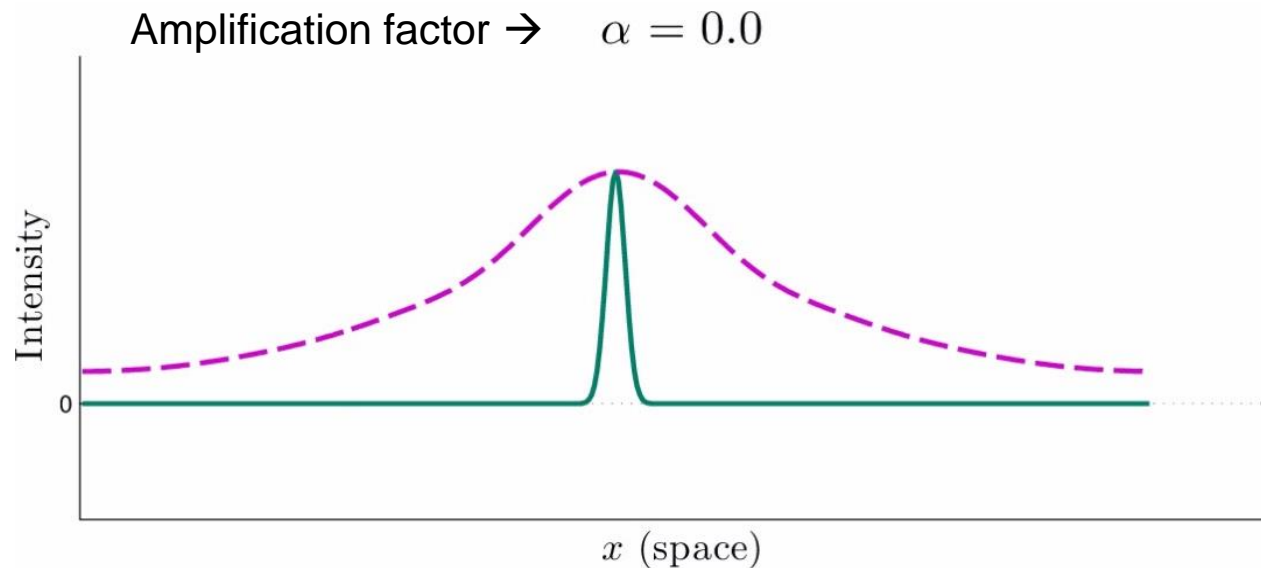
Amplification factor $\rightarrow \alpha = 0.0, \delta = 0.1 \leftarrow$ Motion in the sequence



**4 times the
amplification!**

Limits of Phase Based Magnification

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

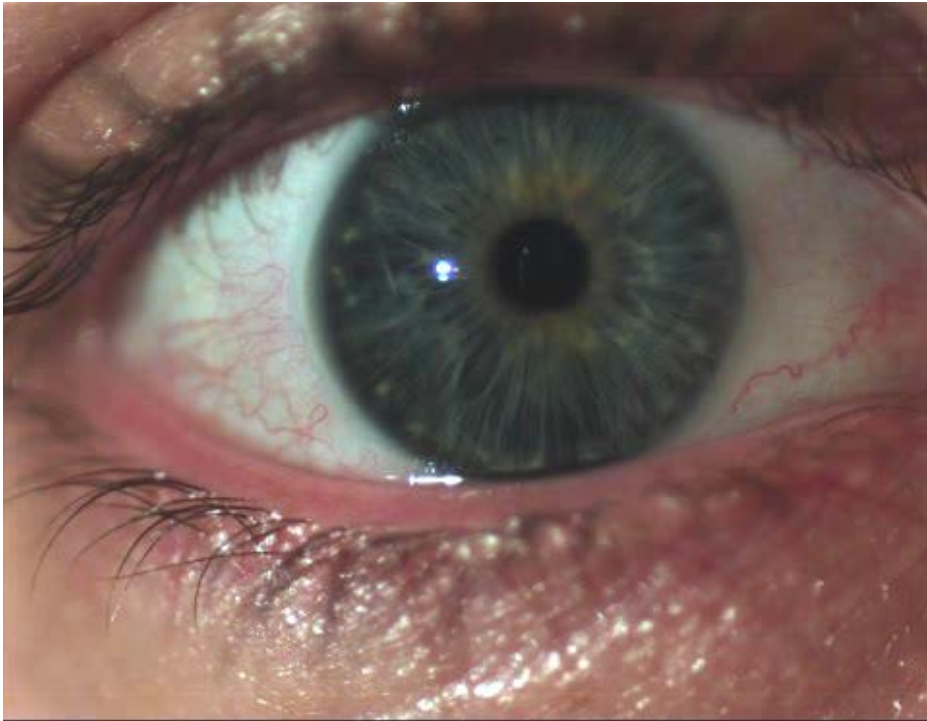


Comparison with [Wu et al. 2012]



Wu et al. 2012

Eye Movements



Source (500FPS)

Expressions



Source



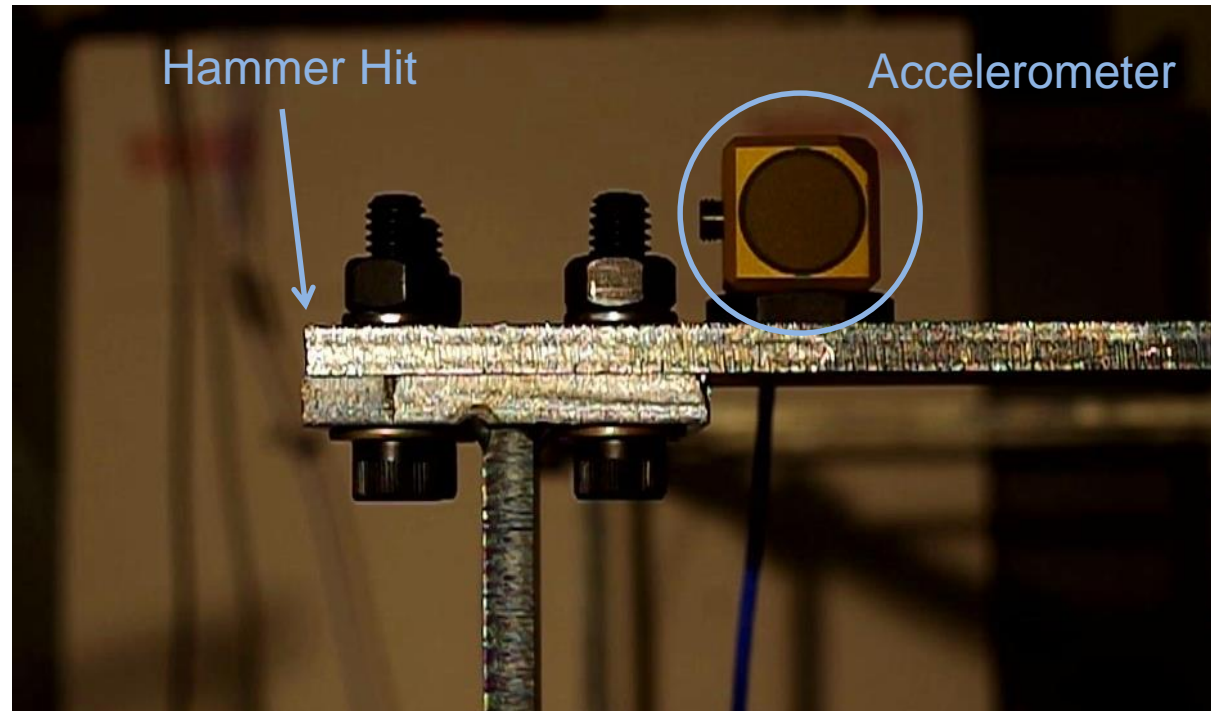
Low frequency motions



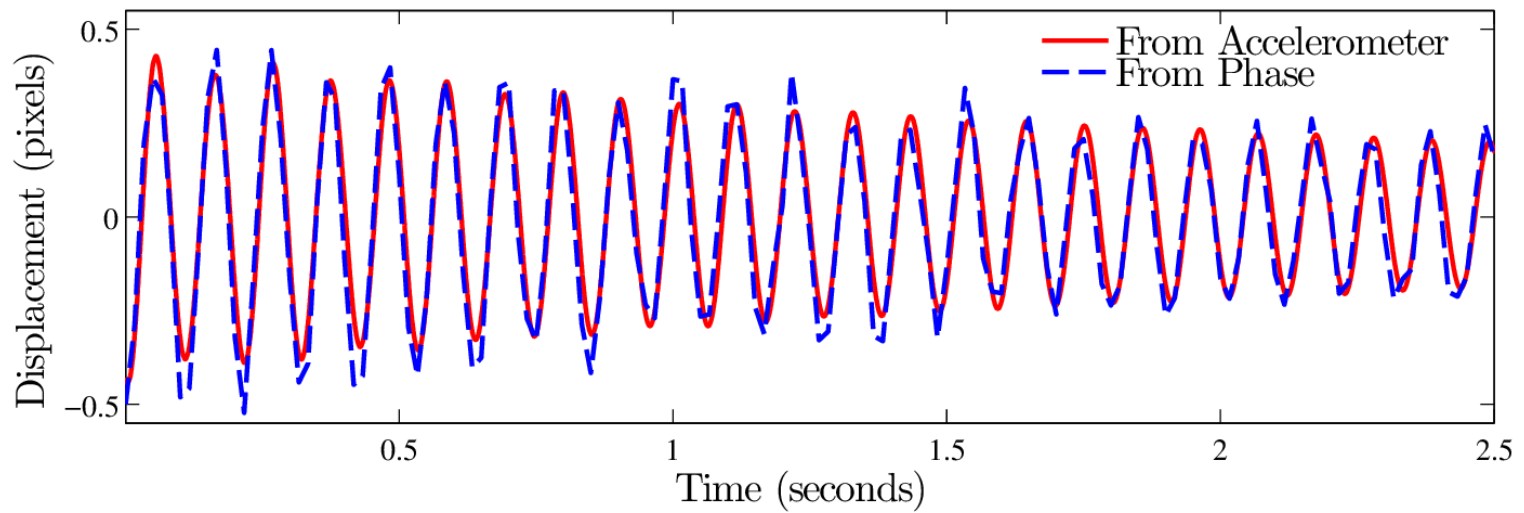
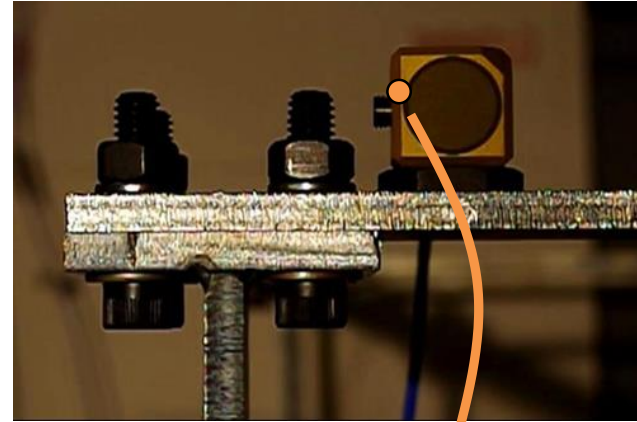
Mid-range frequency motions

Ground Truth Validation

- Induce motion (with hammer)
- Record with accelerometer



Ground Truth Validation



Motion Attenuation



Source

Sequence courtesy Vimeo user Vincent Laforet

Car Engine

Source



Car Engine

22Hz Magnified



Car Engine

Source

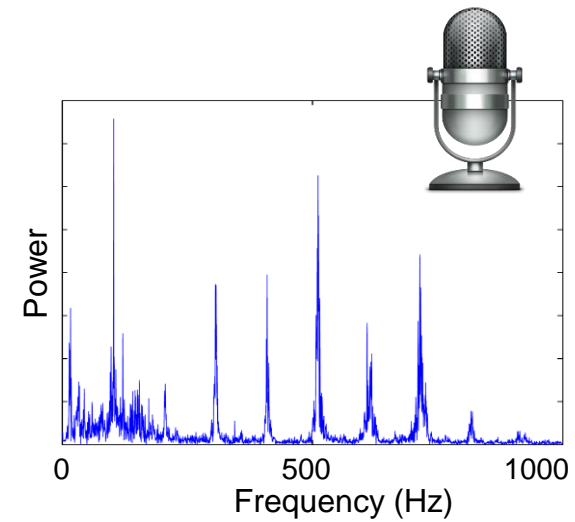
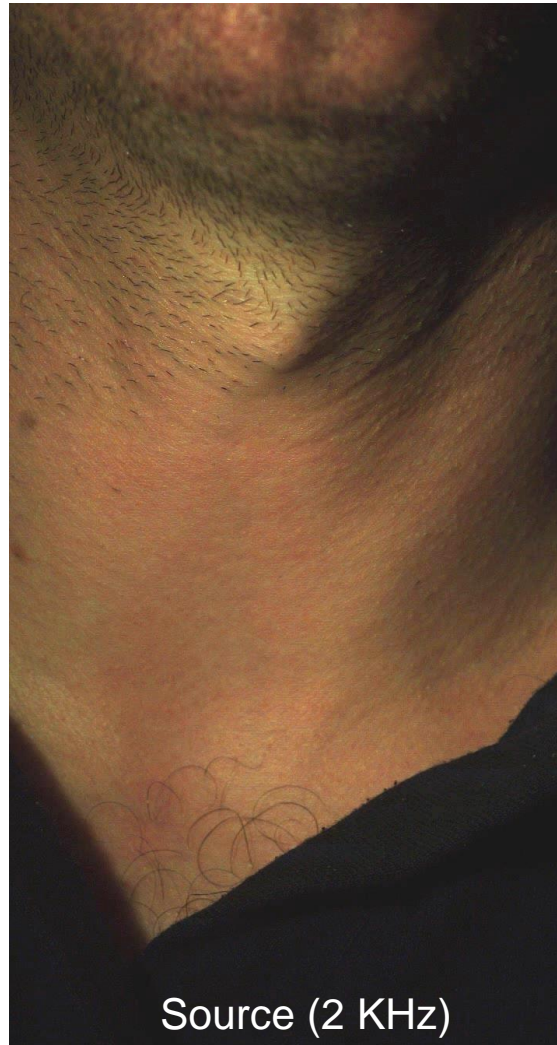
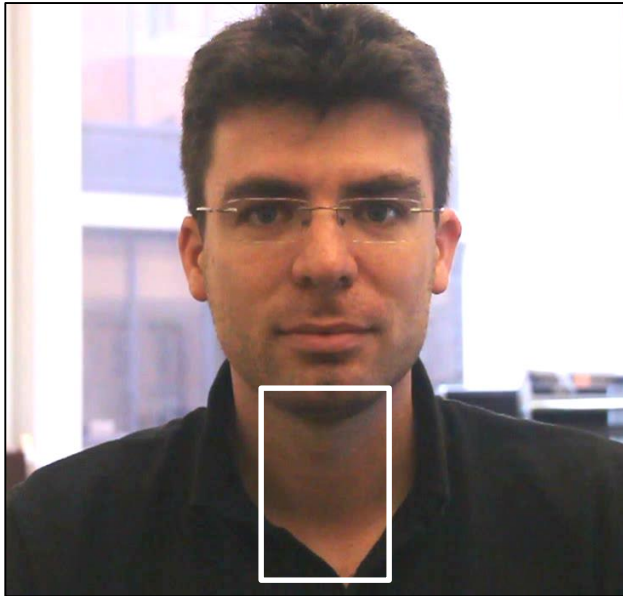


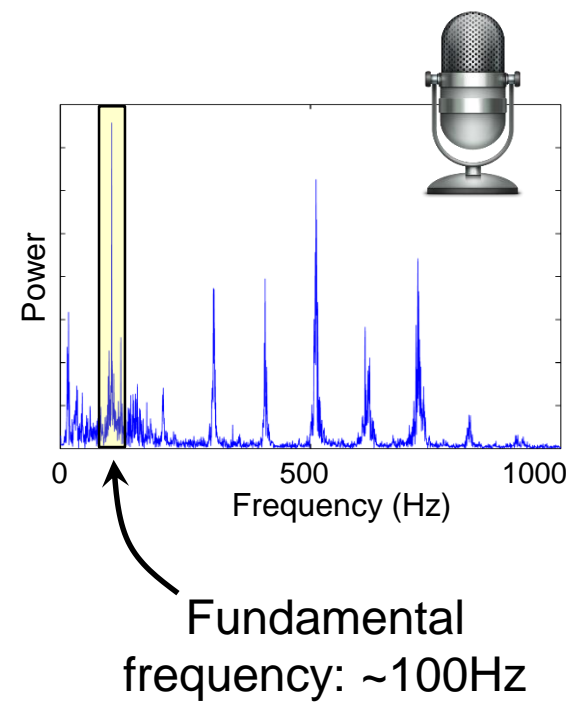
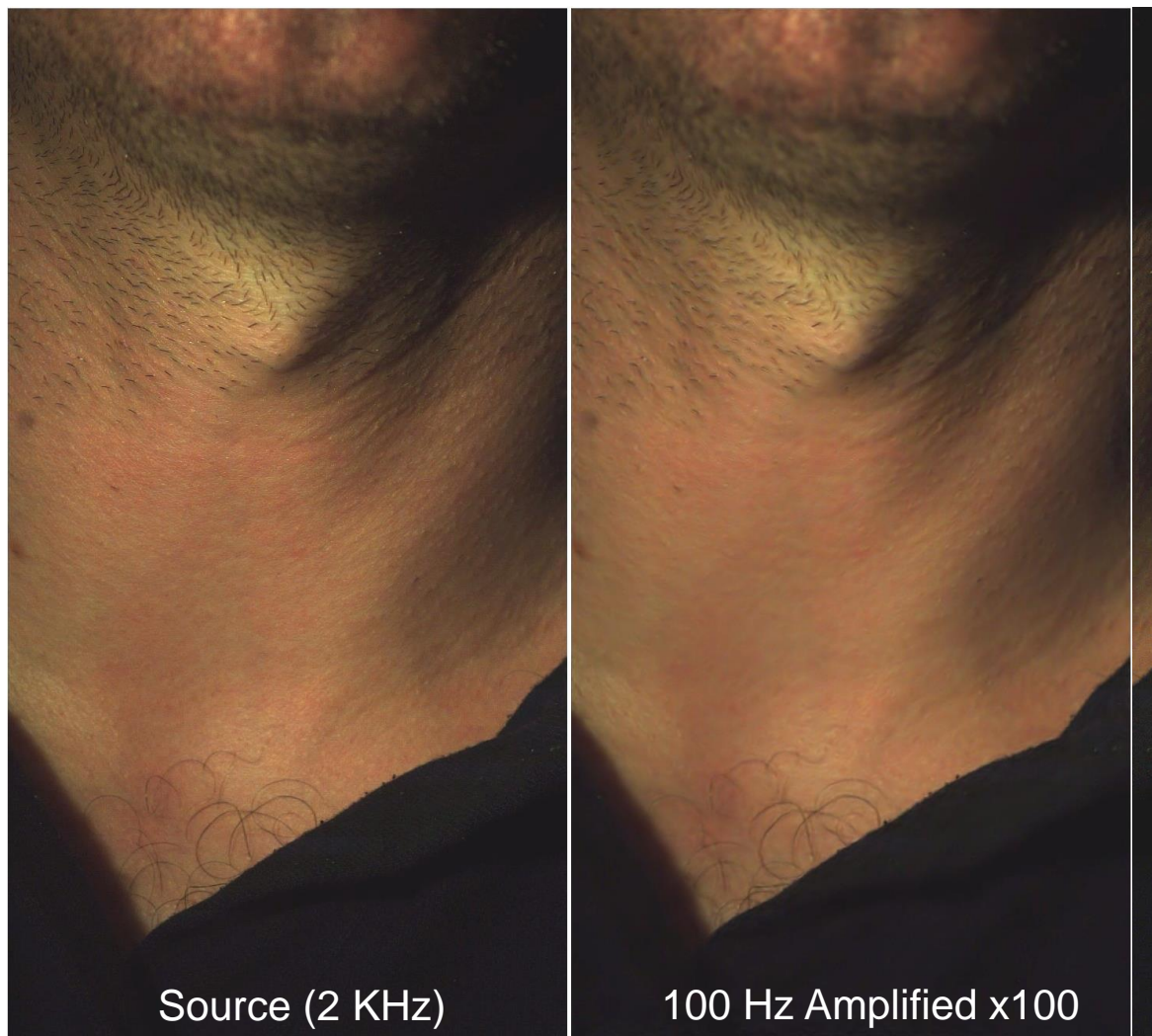
Car Engine

22Hz Magnified



Neck Skin Vibrations







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 amplification (compared to pointwise approach)
 - May have difficulty with non-periodic motion and large motions

Non-periodic Motions and Large Motions

Non-periodic motion



Source (300 FPS)



Motion Magnification x50



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



(slides adopted from Siggraph presentation)

Remote Sound Recovery



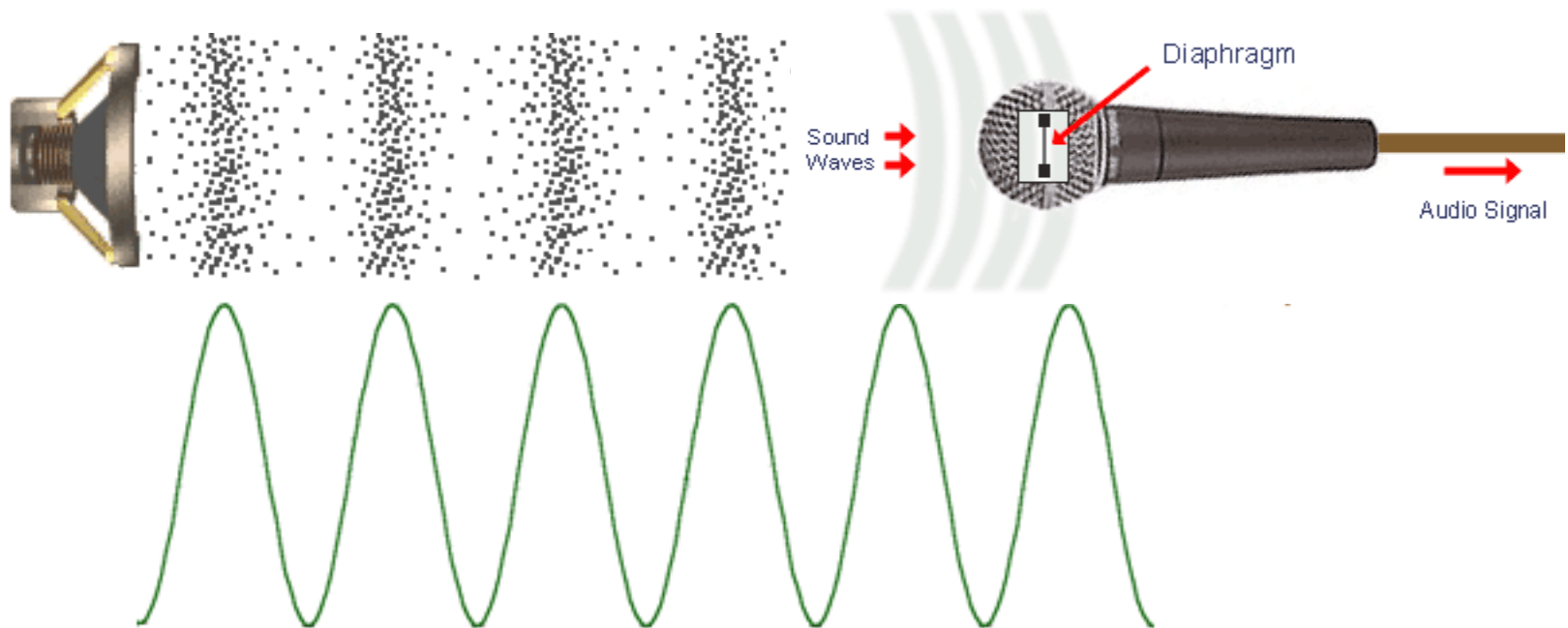
SIGGRAPH2014



Sound and Motion



SIGGRAPH2014

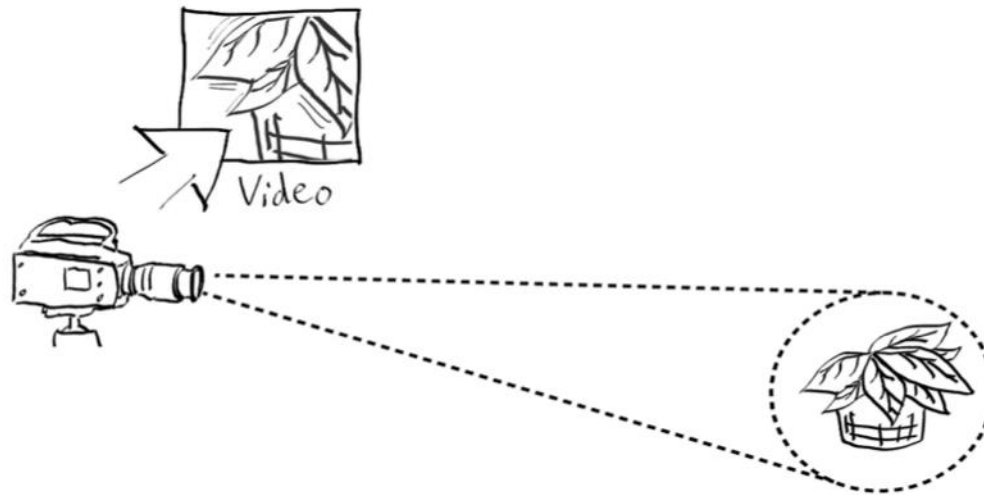
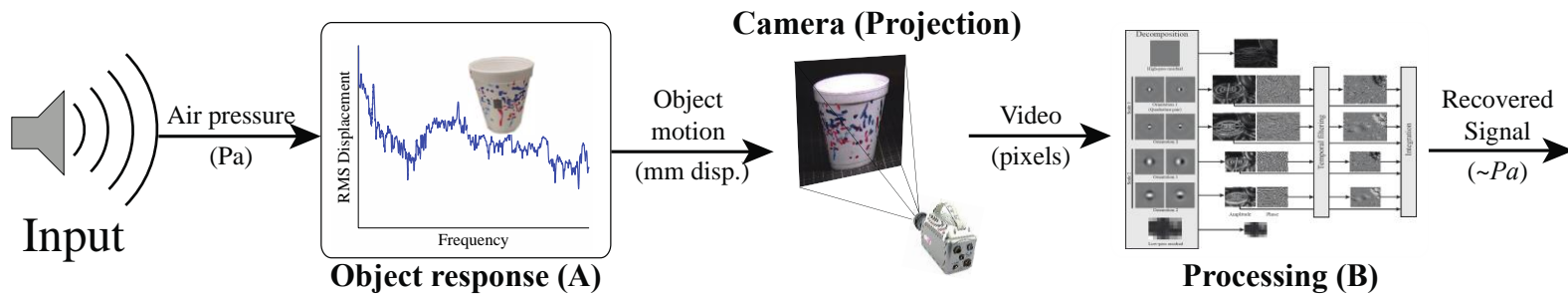


Source: mediacollege.com

The Visual Microphone



SIGGRAPH2014

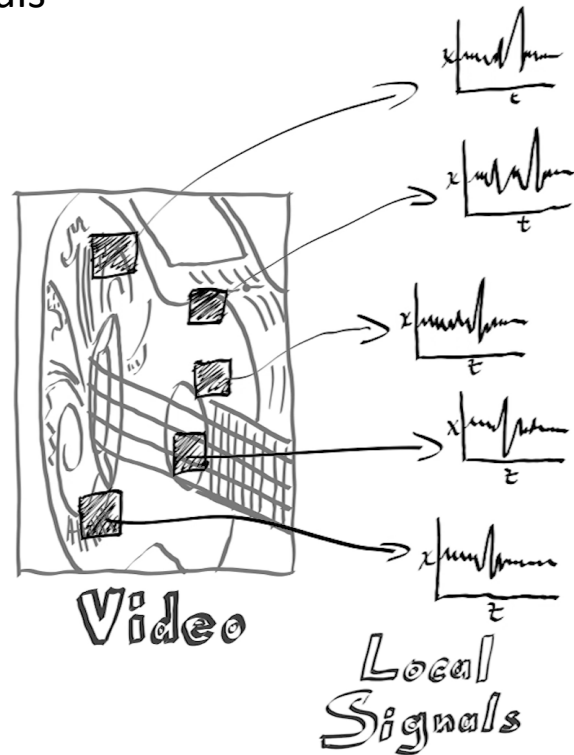


Processing



SIGGRAPH2014

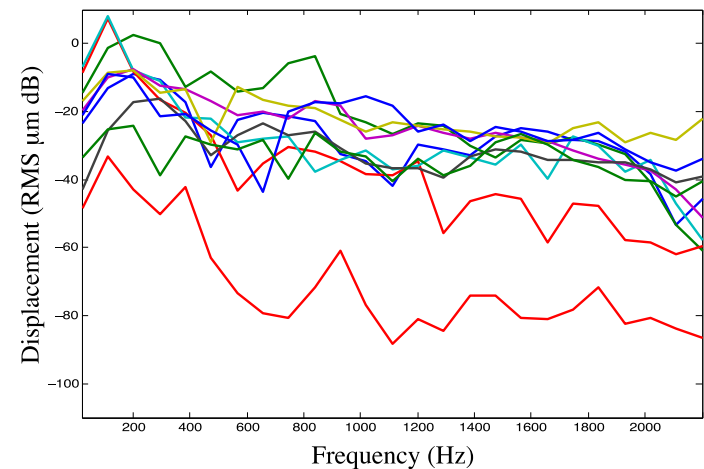
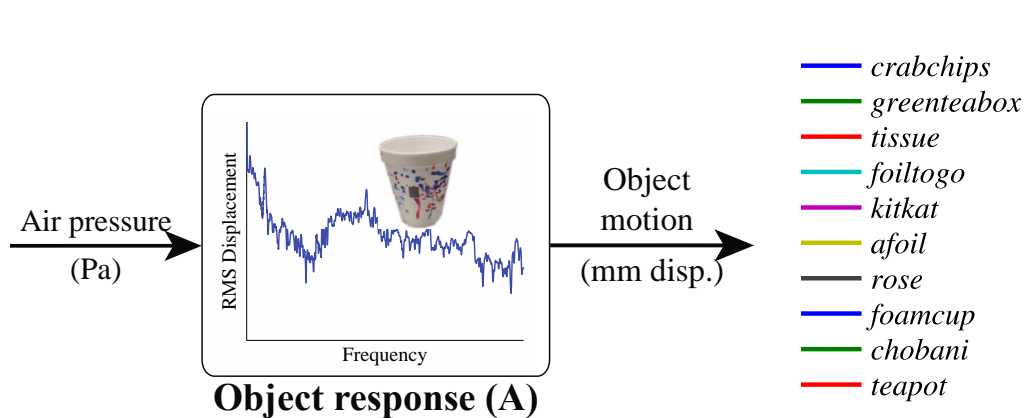
- Extract local motion signals
- Average and Align
- Post-process



Some materials are better microphones than others



SIGGRAPH2014

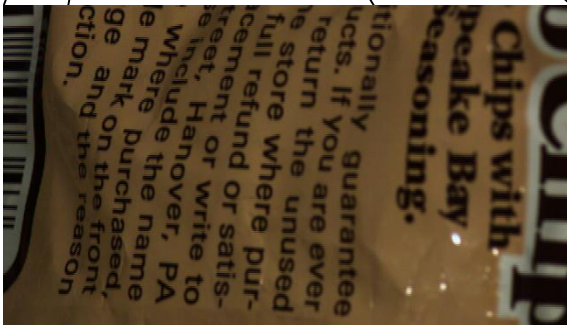


(c) Frequency responses

Sound Recovered from Video



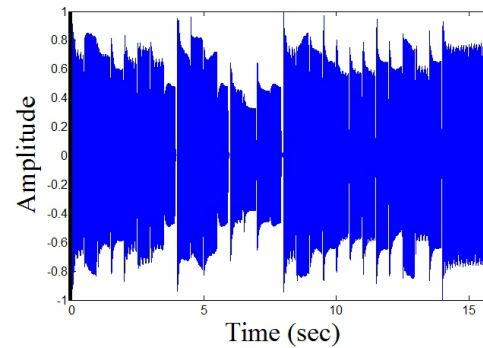
SIGGRAPH2014



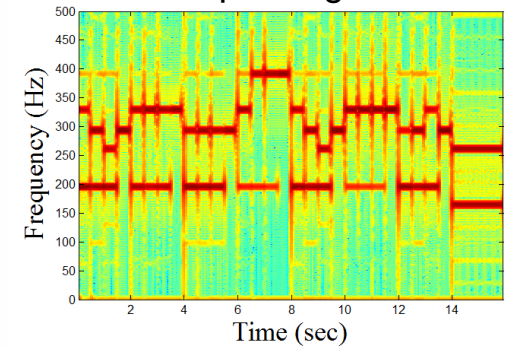
2200Hz video

Source sound in the room

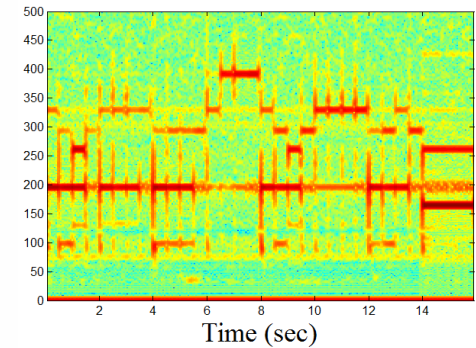
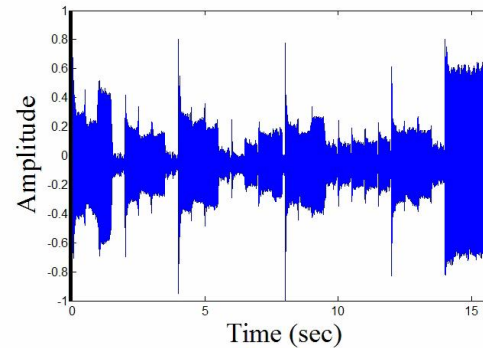
Waveform



Spectrogram



Recovered sound



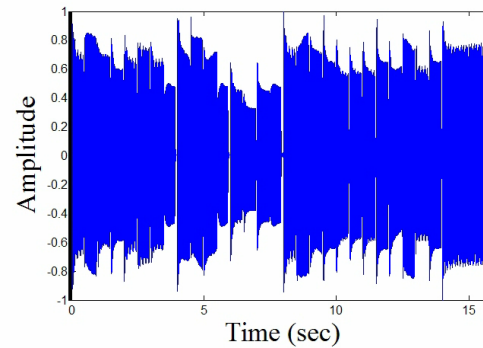
Sound Recovered from Video



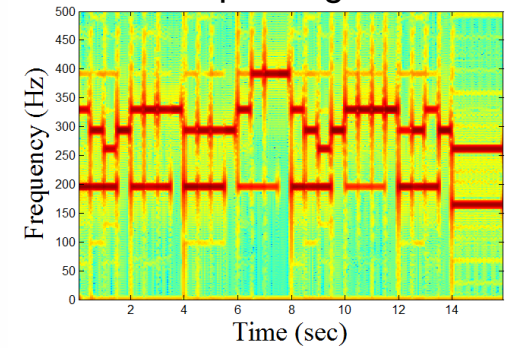
SIGGRAPH2014

Source sound in the room

Waveform



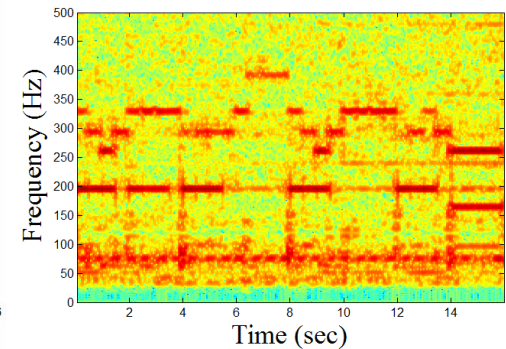
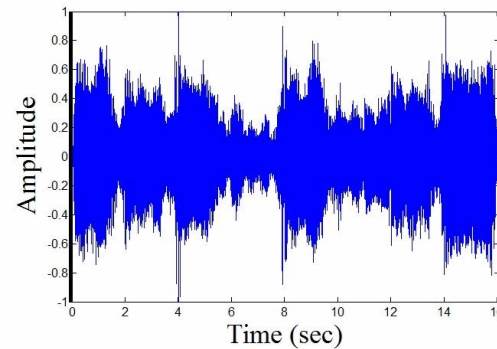
Spectrogram



Recovered sound



2200Hz video



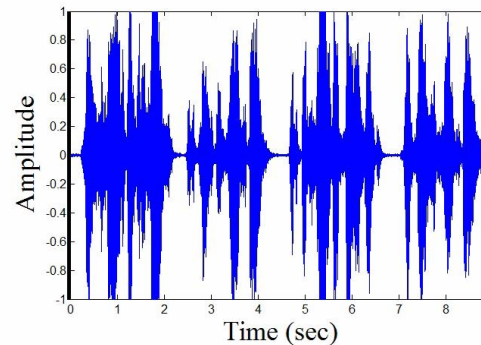
Sound Recovered from Video



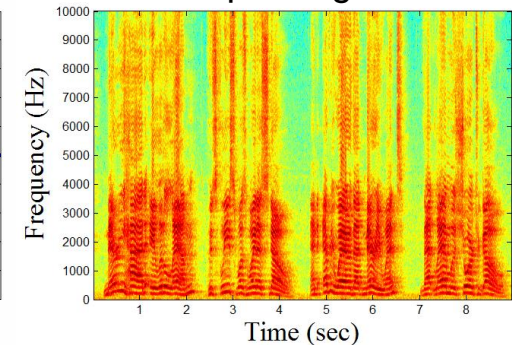
SIGGRAPH2014

Source sound in the room

Waveform



Spectrogram



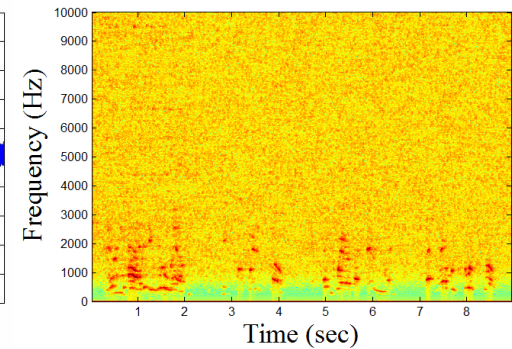
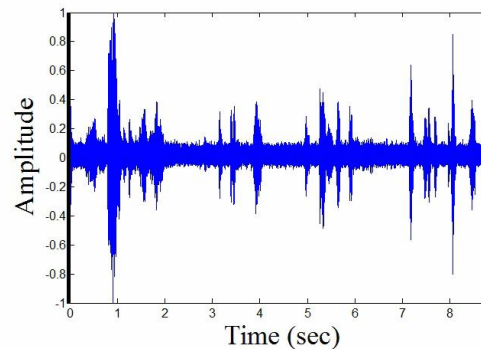
(small patch on the chip bag)



20 kHz video



Recovered sound





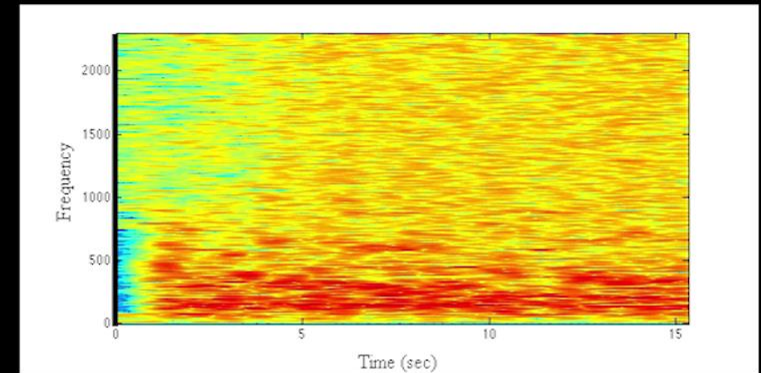
High speed video
(actual video playing here)



Object



Automatic Identification of Recovered Audio



Sound Recovered From Video of Earbuds

Rolling Shutter



SIGGRAPH2014



<https://www.flickr.com/photos/sorenragdale/3904937619/>
<http://www.flickr.com/photos/boo66/5730668979/>

Rolling Shutter



SIGGRAPH2014

Rolling Shutter

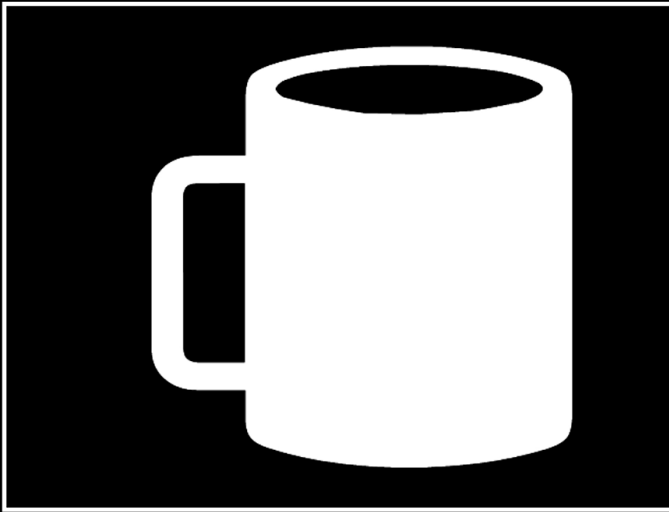


Image of vibrating object
projected on sensor



Image read from
rolling shutter sensor

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

Rolling Shutter



SIGGRAPH2014

Rolling Shutter

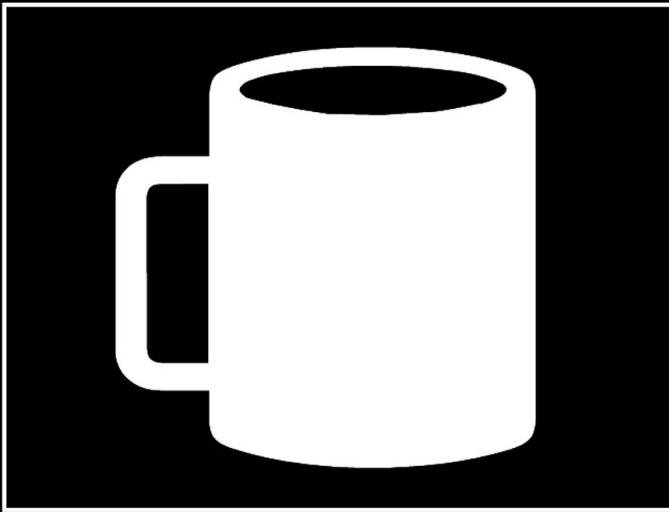


Image of vibrating object
projected on sensor



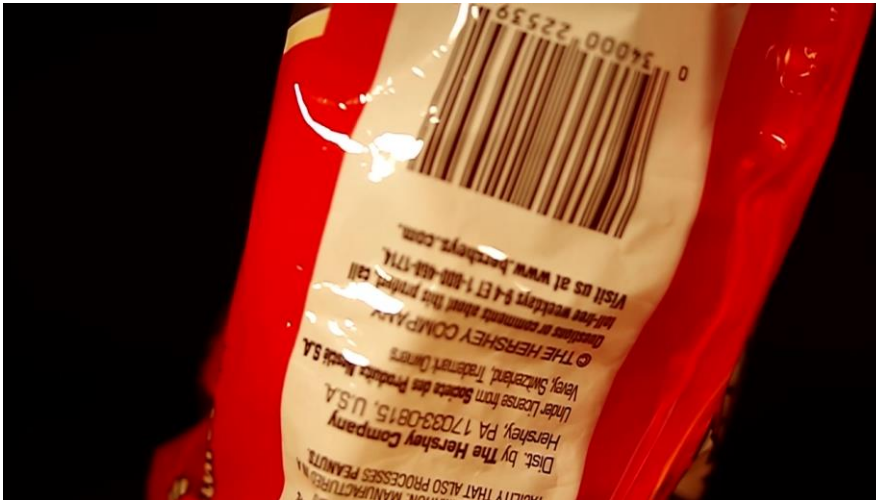
Image read from
rolling shutter sensor

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

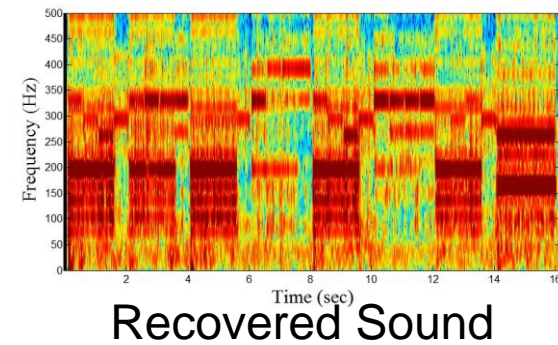
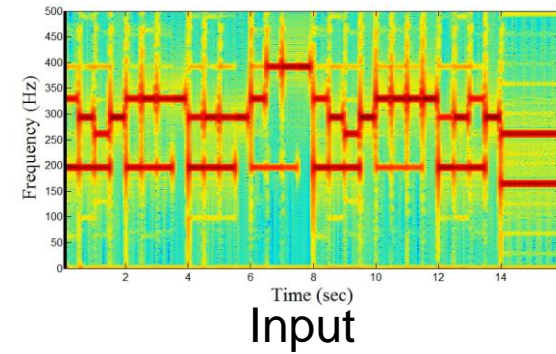
Rolling Shutter



SIGGRAPH2014



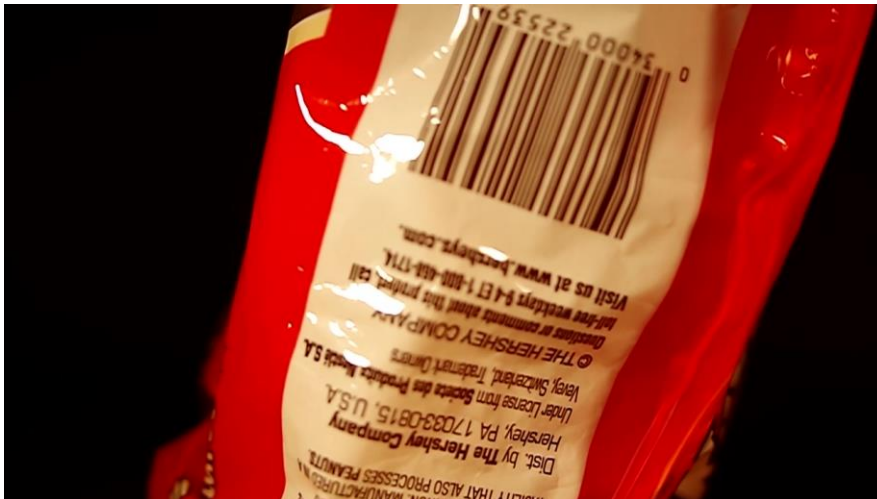
Input video (60 fps)



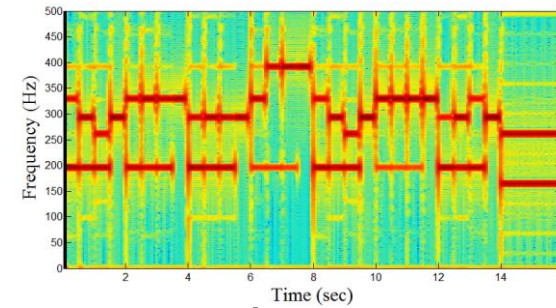
Rolling Shutter



SIGGRAPH2014

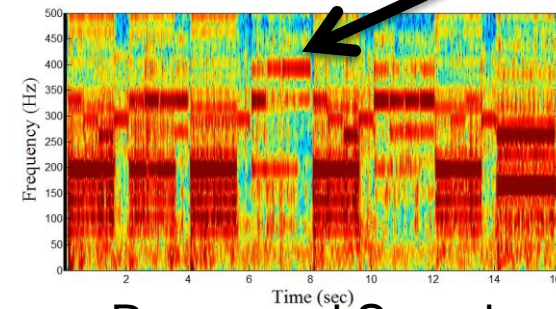


Input video (60 fps)



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)