Massachusetts Institute of Technology

CSAIL

Thesis Defense:

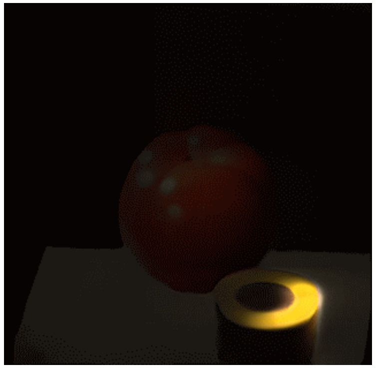
Analysis and Visualization of Temporal Variations in Video

Nov 25 2013

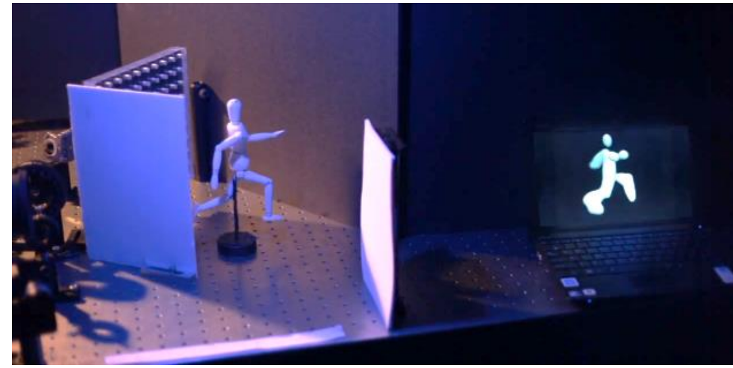
Michael Rubinstein

MIT CSAIL

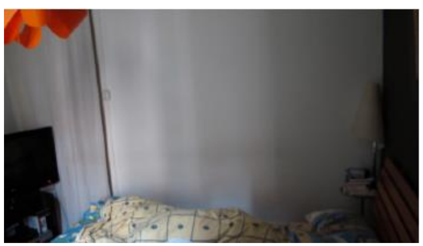
Seeing the Unseen in Images/Videos



[Velten et al., ***Femto-Photography,*** 2011]



[Velten et al., ***CORNAR,*** 2012]

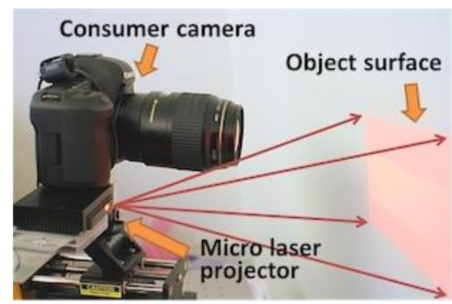


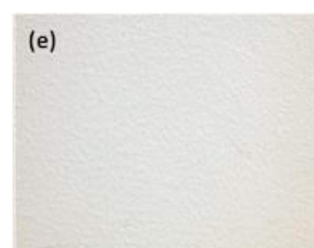
Input

Result

True image

[Torralba and Freeman, ***Accidental Images,*** 2012]



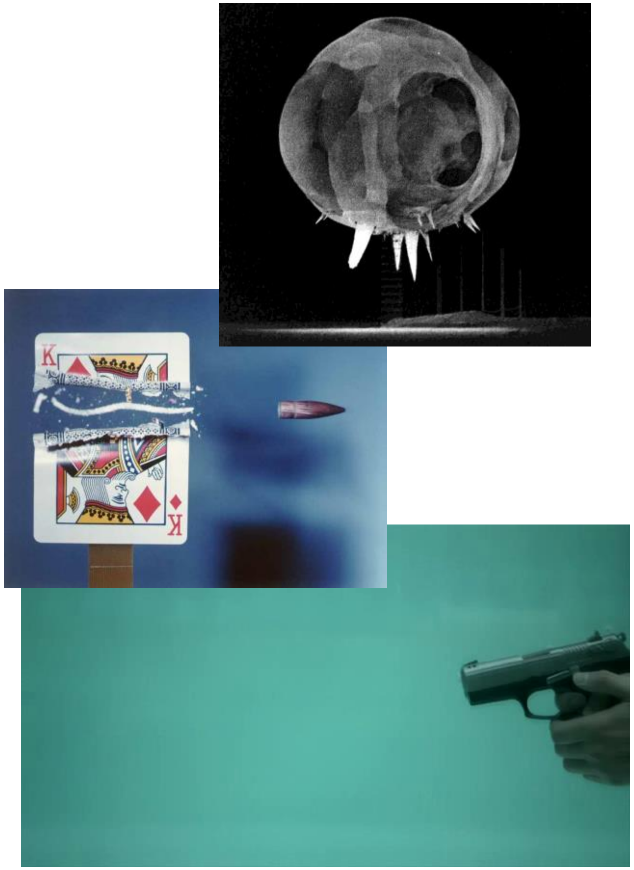


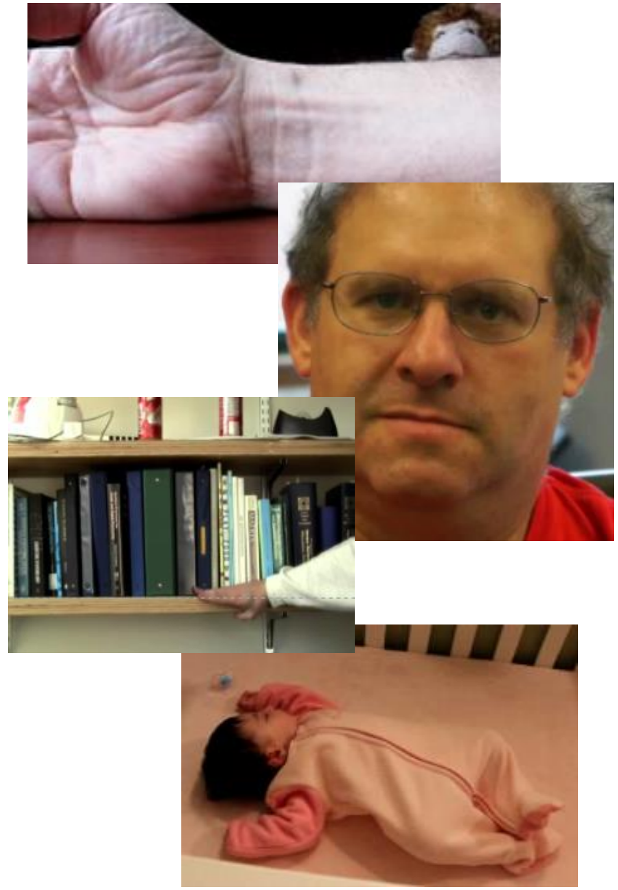
Input

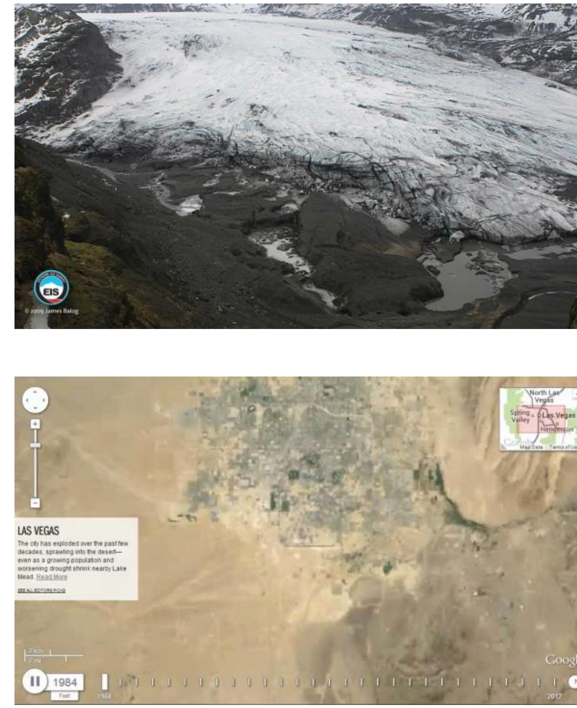
Result

[Shih et al., ***Laser Speckle Photography,*** 2012]

Timescales in Imagery







Milliseconds
10^4 fps (high-speed)

Seconds, Minutes
10^1 fps (standard videos)

Months, years
10^{-4} fps (time-lapse)

Timescale

Distracting Temporal Variation

- Too many confounding changes
  - Mixed changes at different timescales
  - Lighting changes, objects appearing/disappearing, …

Source

Short-term changes attenuated

Remove changes to make long-term variation more visible!

Imperceptible Temporal Variation

***Magnify** the variation to make it visible!*

- Changes are too small

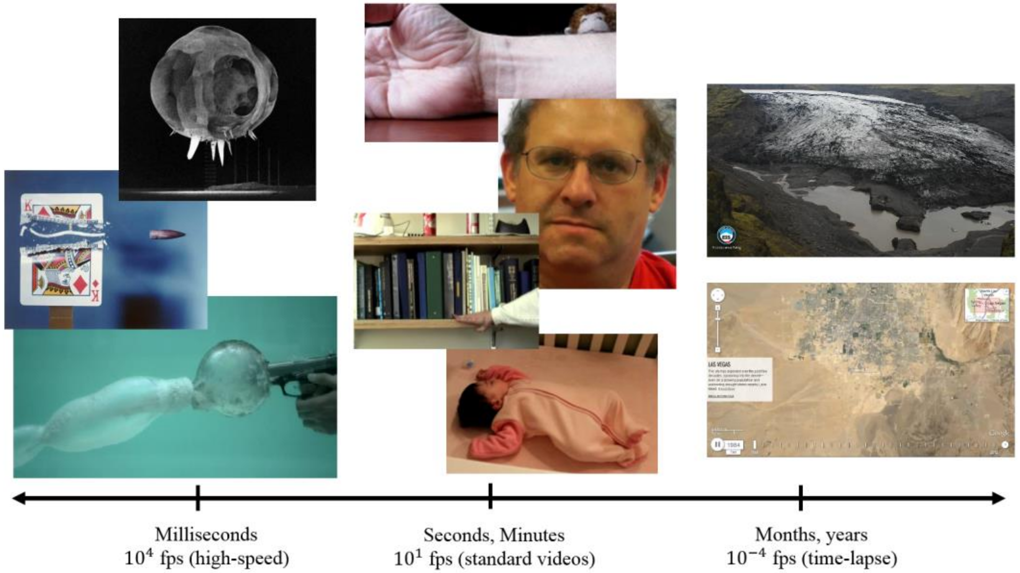
[Liu et al. 2005]

This Thesis

- Assist the analysis of **temporal phenomena** captured by imagery

- **Reveal interesting temporal signals** that may not be easily visible in the original data

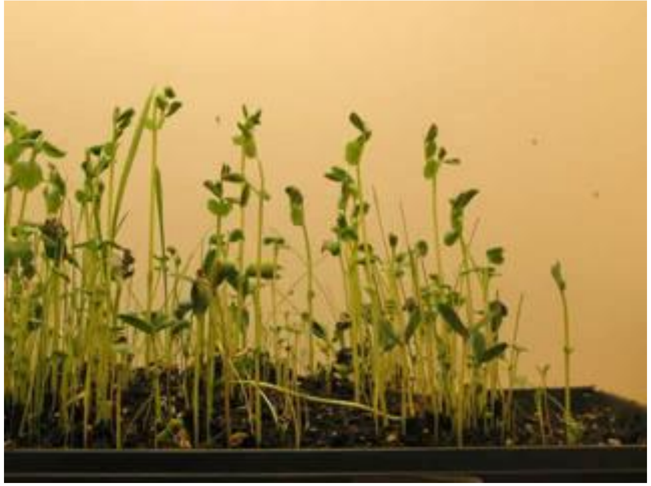
- Leverage available imagery
  - Regular video, natural setting



Our approach: analyzing images/videos and re-rendering changes in them such that the interesting temporal signals are more apparent

Talk Outline

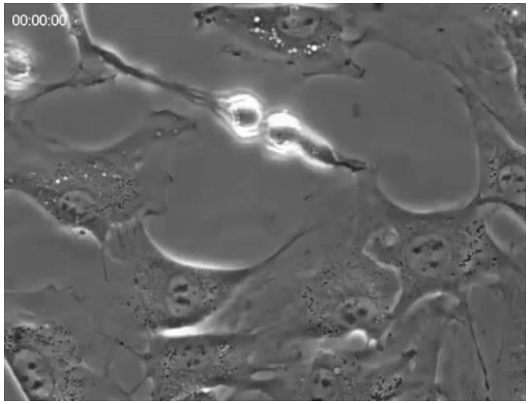
- Removing distracting variations
  - Motion Denoising



- Magnifying imperceptible variations
  - Eulerian Video Magnification
  - Phase-based Video Motion Processing

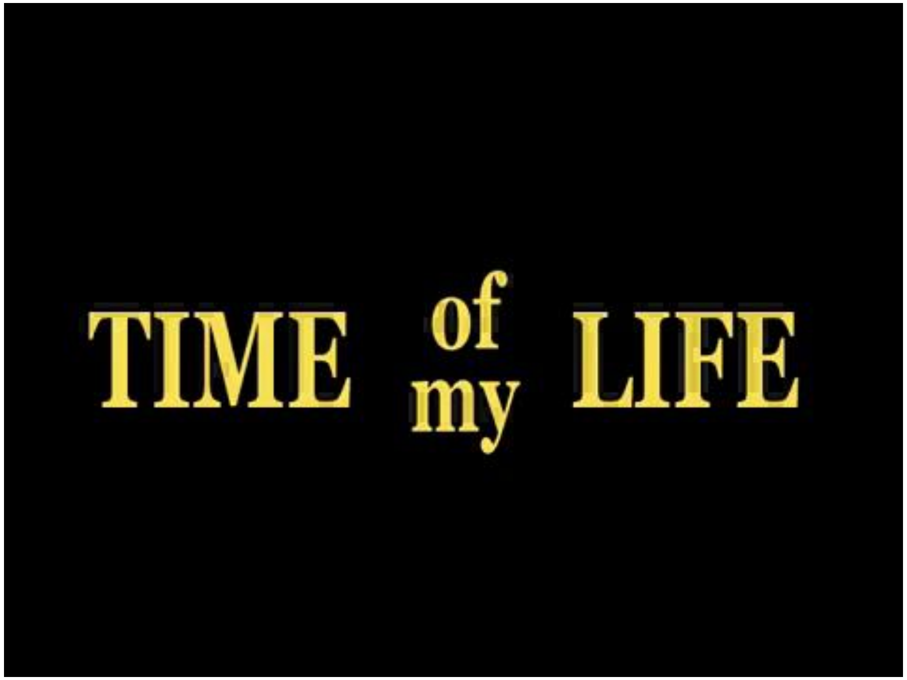
- Ongoing research and future work

Time-lapse Videos



For Personal Use Too

9 months

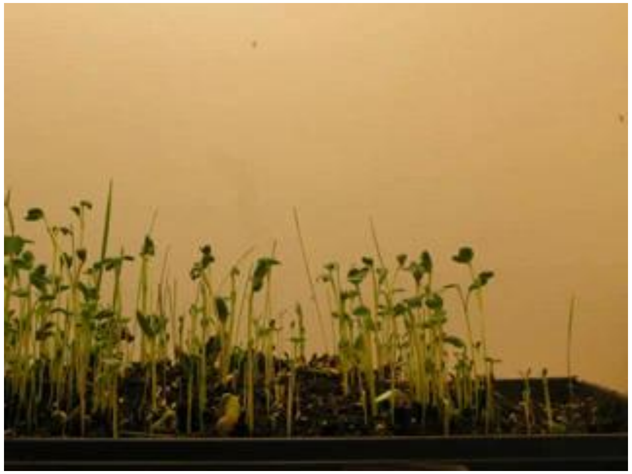


16 years

10 years

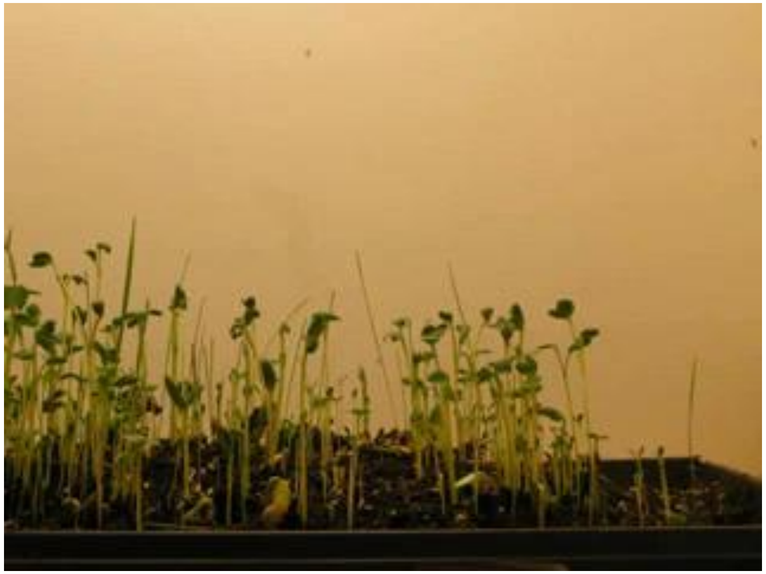
1 year

Stylized Jerkiness



Timescale Separation

- Decompose the video into long-term and short-term changes



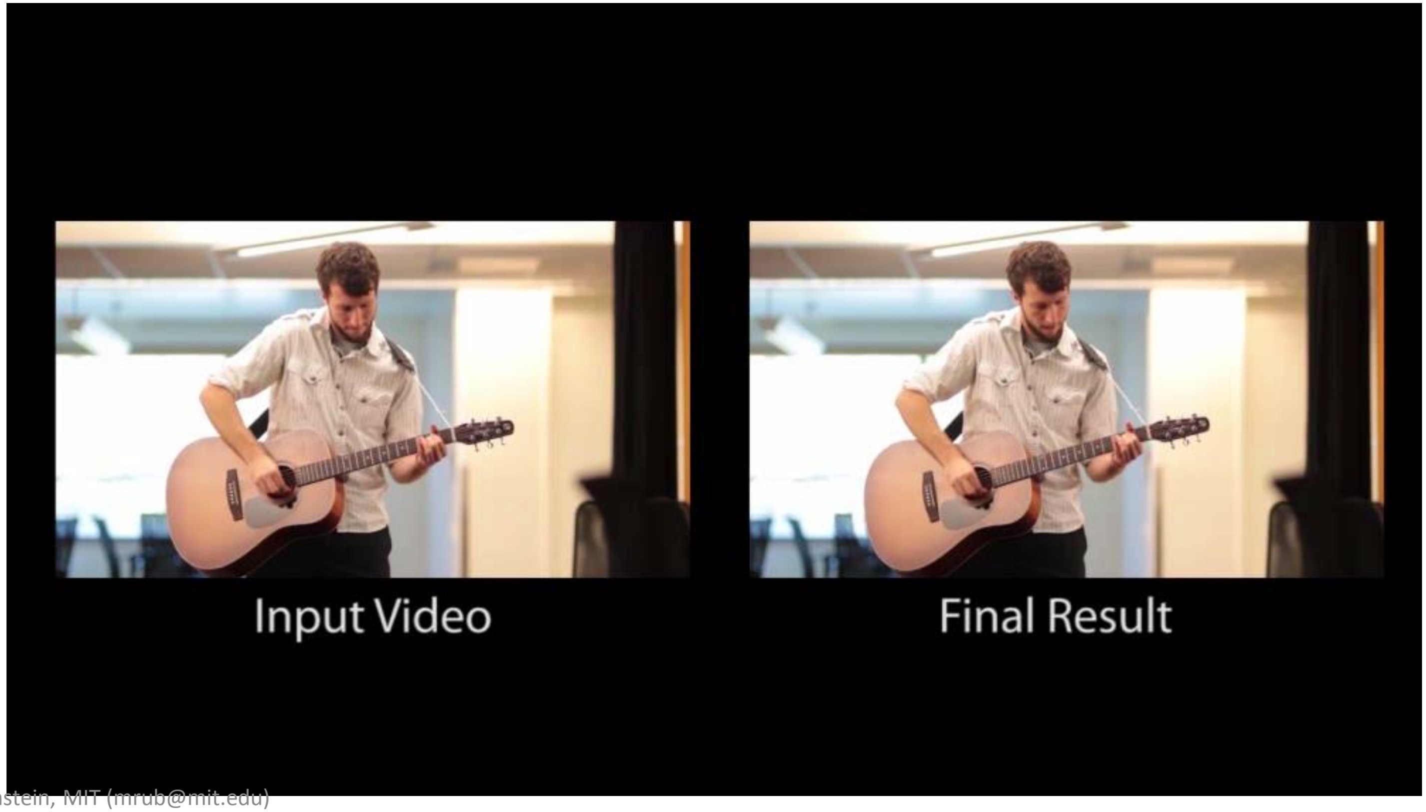
Motion Denoising

Motion Denoising with Application to Time-lapse Photography, CVPR 2011
With Ce Liu, Peter Sand, Fredo Durand, William T. Freeman

Related Work

- **Video stabilization** [Matsushita et al. 2006], [Liu et al. 2011], [Grundmann et al. 2011]
  - Can denoise camera motions, but we need pixel-wise stabilization

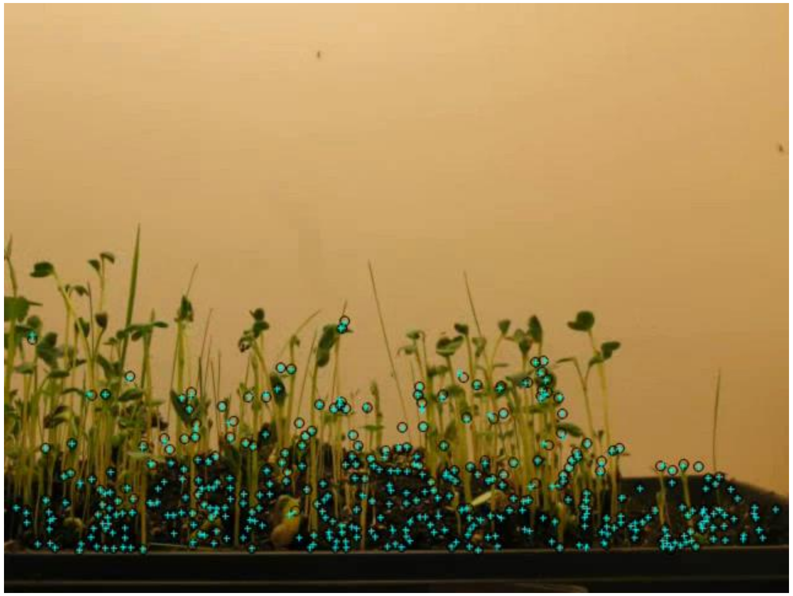
- **Selectively De-Animating Video** [Bai et al. 2012]



How to Denoise Motion?

- Pixel-wise temporal low-pass filtering
  - Pixels of different objects are averaged

- Smoothing motion trajectories
  - Motion estimation in time-lapse videos is challenging:
    - Brightness ***in***consistencies
    - Motion ***dis***continuities

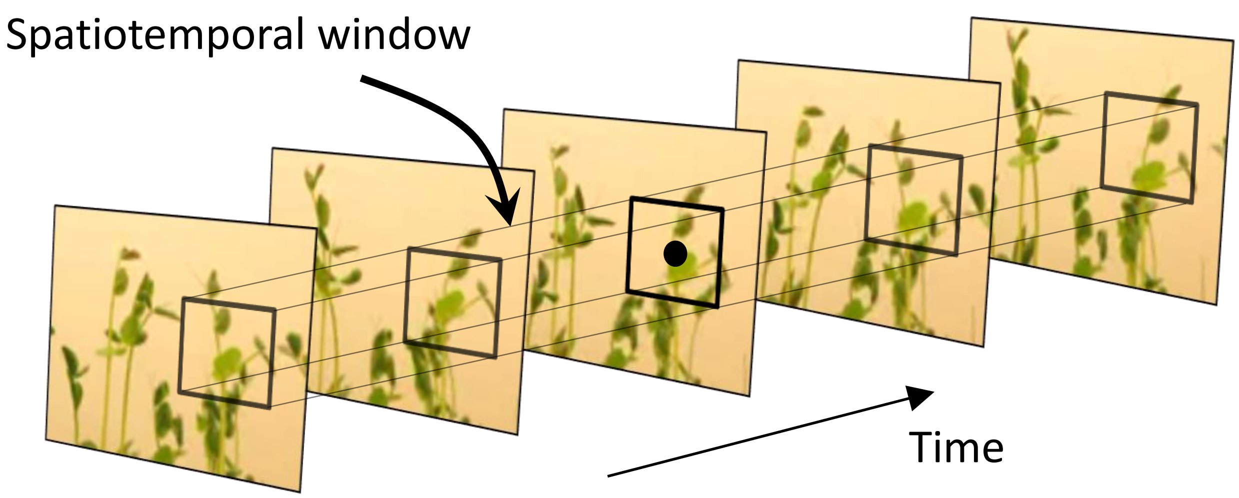


KLT tracks

Basic Idea

- **<u>Assumption:</u>** scene is changing slowly and perturbed by random motions (and color changes)

- **<u>Approach:</u>** reshuffle the pixels in space and time to reconstruct the smooth scene
  - Allow the filter to "look around" within local spatiotemporal windows



Formulation

- Solve for a spatiotemporal displacement (offset) field, $\boldsymbol{w}$:

$$E(\boldsymbol{w}) = \sum_{p} |I(p + \boldsymbol{w}(p)) - I(p)| +$$

Fidelity (to input video)

$$\alpha \sum_{p,r \in N_t(p)} \|I(p + \boldsymbol{w}(p)) - I(r + \boldsymbol{w}(r))\|^2 +$$

Temporal coherence (of the output video)

$$\gamma \sum_{p,q \in N(p)} \lambda_{pq} |\boldsymbol{w}(p) - \boldsymbol{w}(q)|$$

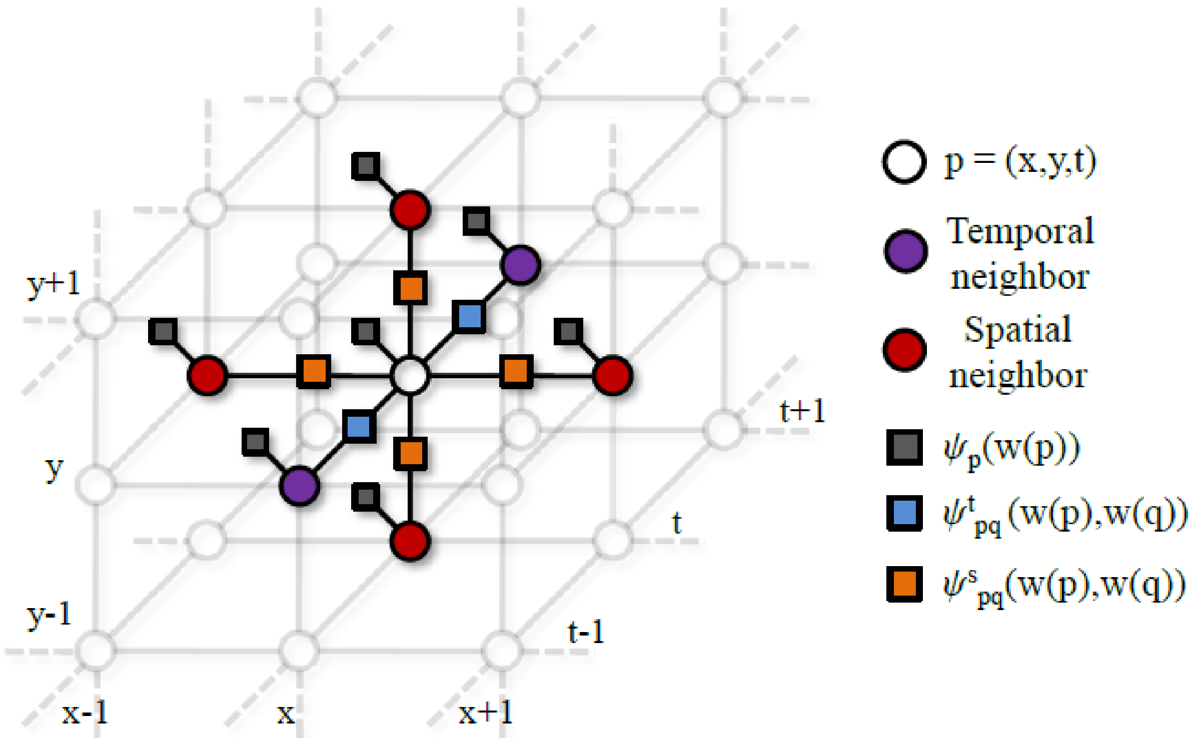
Regularization (of the warp)

$I(p)$ - the input video

$I(p + \boldsymbol{w}(p))$ - the output video

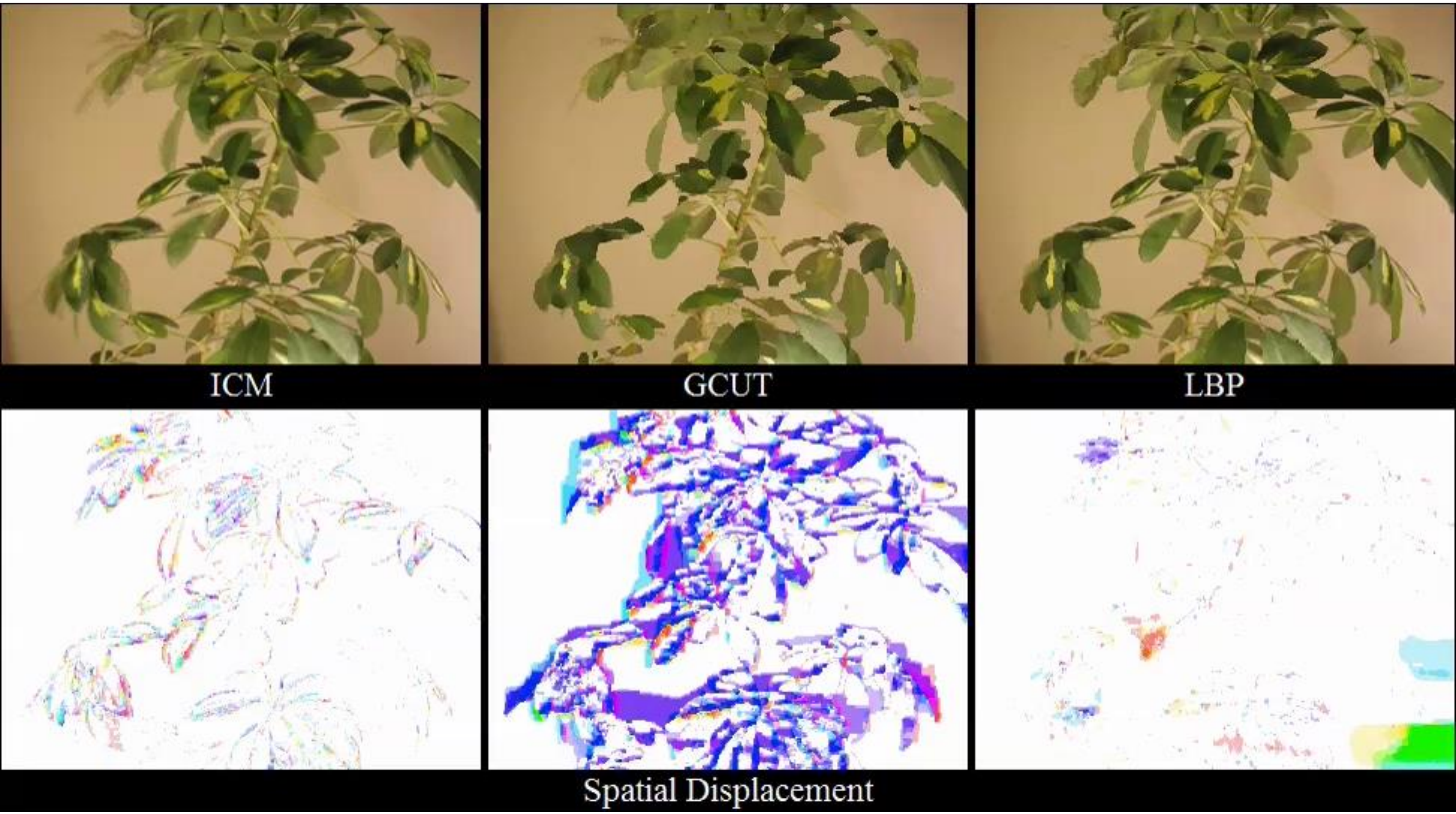
Optimization

- Optimized discretely on a 3D MRF
  - Nodes represent pixels
  - state space of each pixel = volume of possible spatiotemporal offsets



Results

Comparison with Other Optimization Techniques



Iterated conditional modes

Graph Cut (α-expansion)

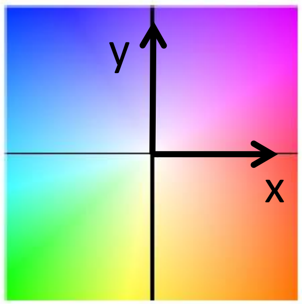
[Boykov et al. 2002]

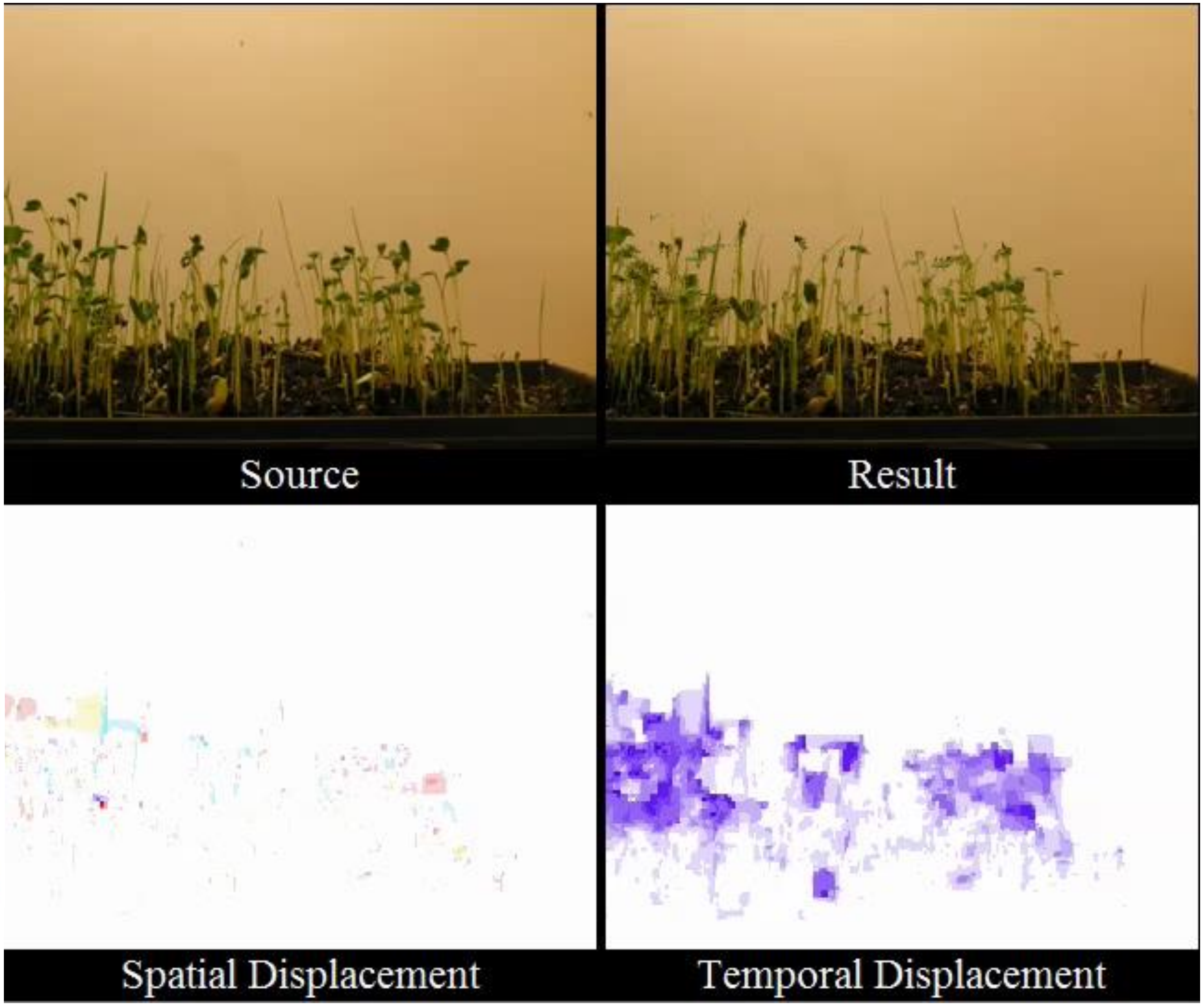
Belief Propagation

Comparison with Pixel-wise Temporal Filtering

Source | Sliding mean | Sliding median | **Motion denoising**

Results


y
x


Source
Result
Spatial Displacement
Temporal Displacement


future
past

Support Size

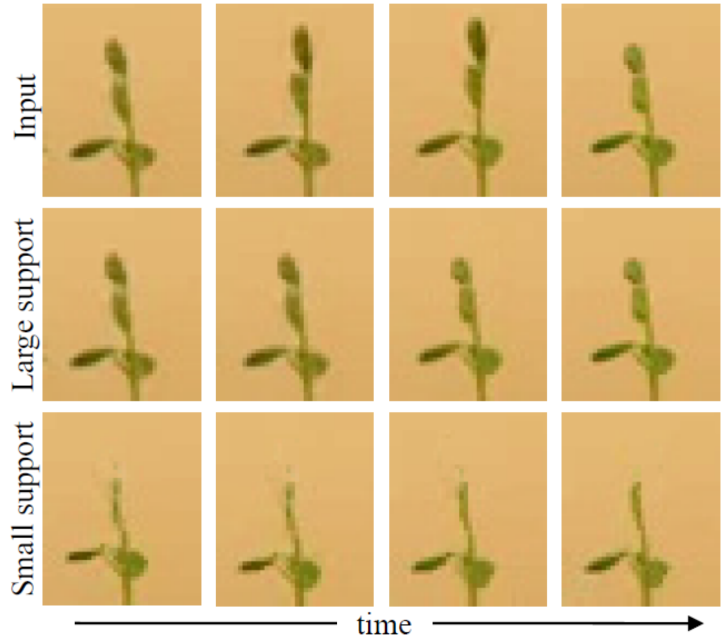
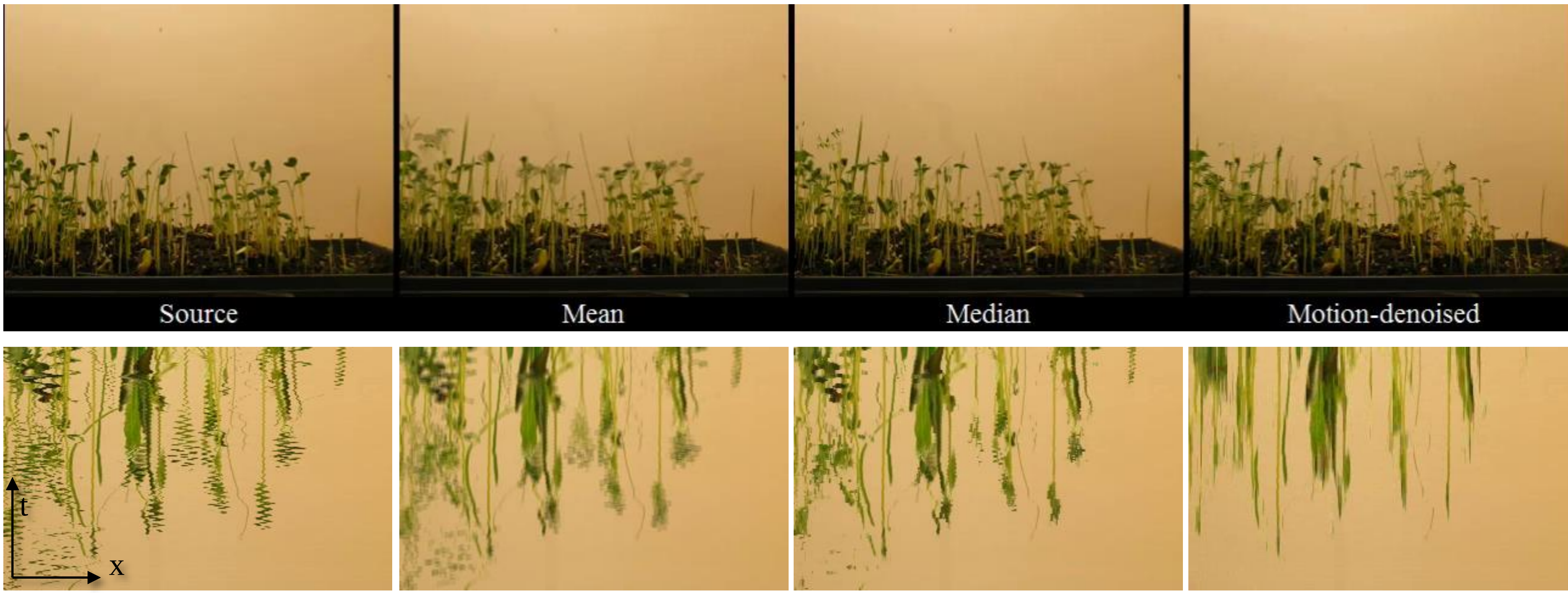


Figure 7. Zoom-in on the rightmost plant in the `sprouts` sequence in four consecutive frames shows that enlarging the search volume used by the algorithm can greatly improve the results. "Large support" corresponds to a $31 \times 31 \times 5$ search volume, while "small support" is the $7 \times 7 \times 5$ volume we used in our experiments.

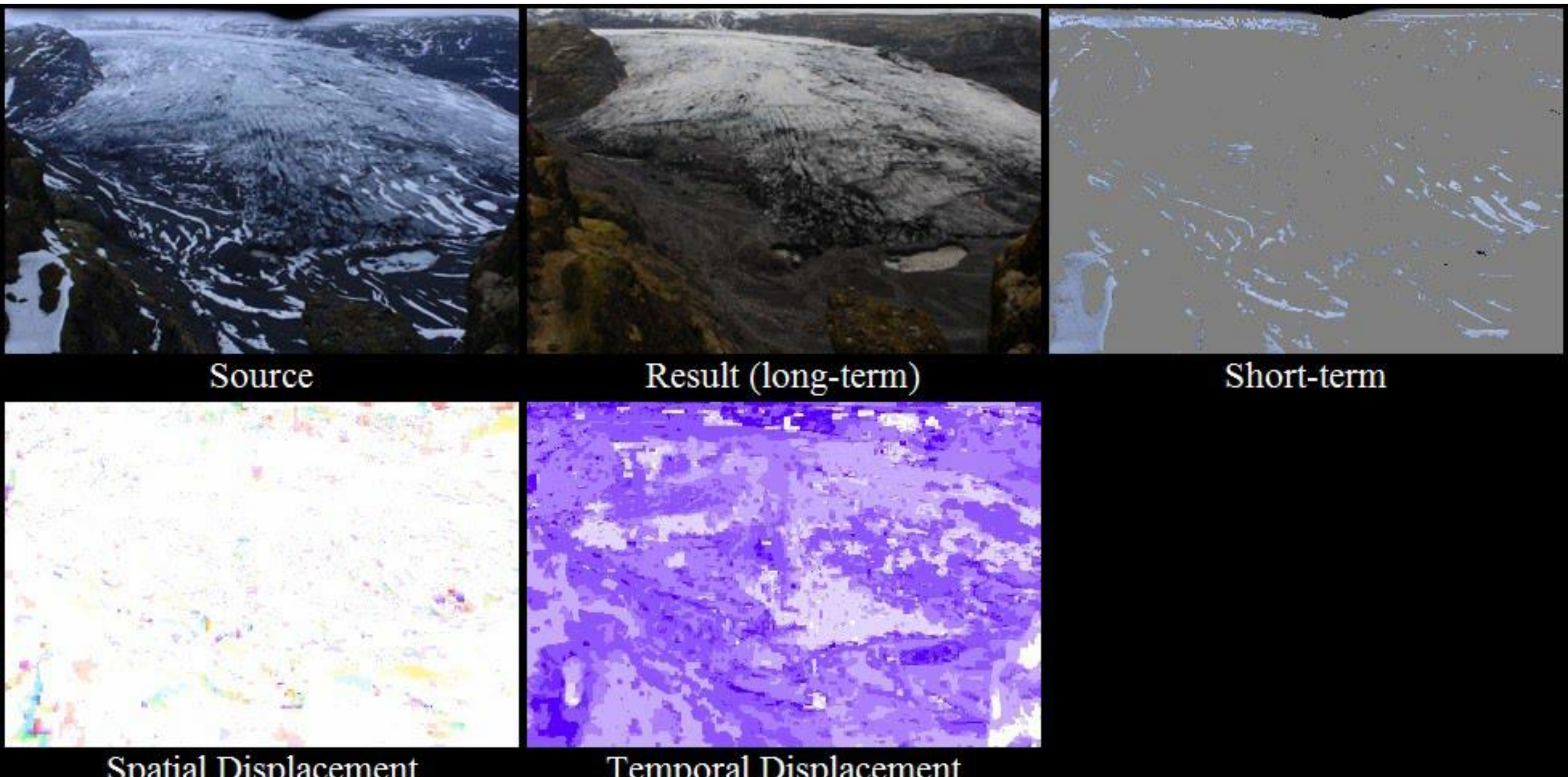
Comparison with Pixel-wise Temporal Filtering



Source | Sliding mean | Sliding median | **Motion denoising**

Timescale Decomposition


t
x


Source
Result (long-term)
Short-term
Spatial Displacement
Temporal Displacement

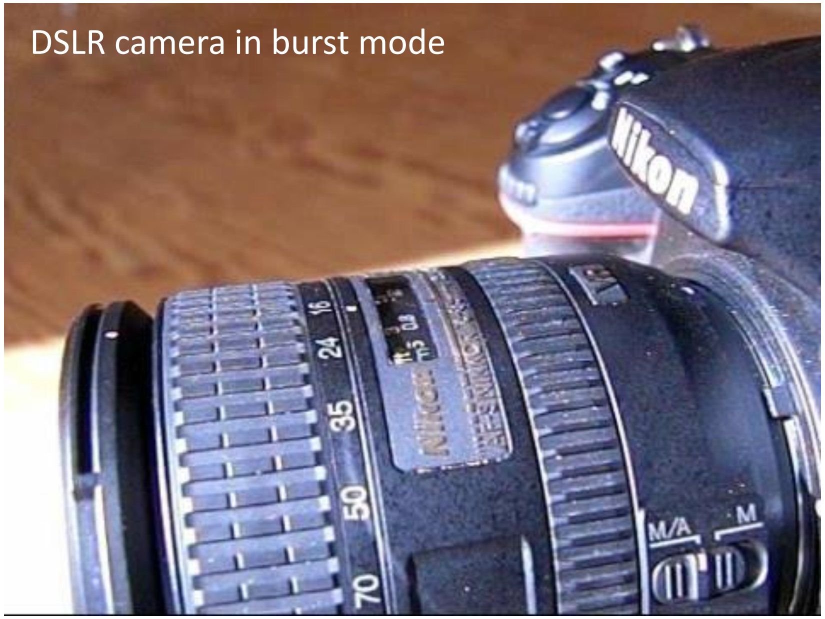
Talk Outline

- Removing distracting variations
  - Motion Denoising

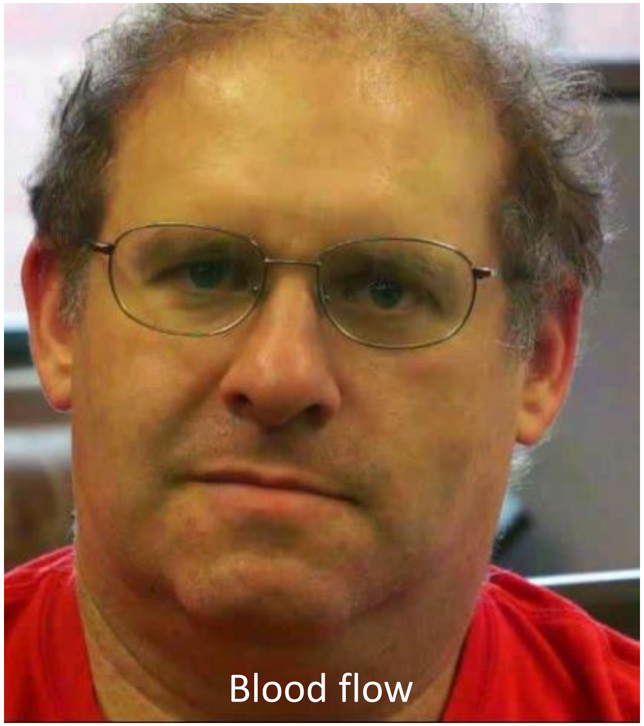
- Magnifying imperceptible variations
  - Eulerian Video Magnification
  - Phase-based Video Motion Processing

- Ongoing research and future work

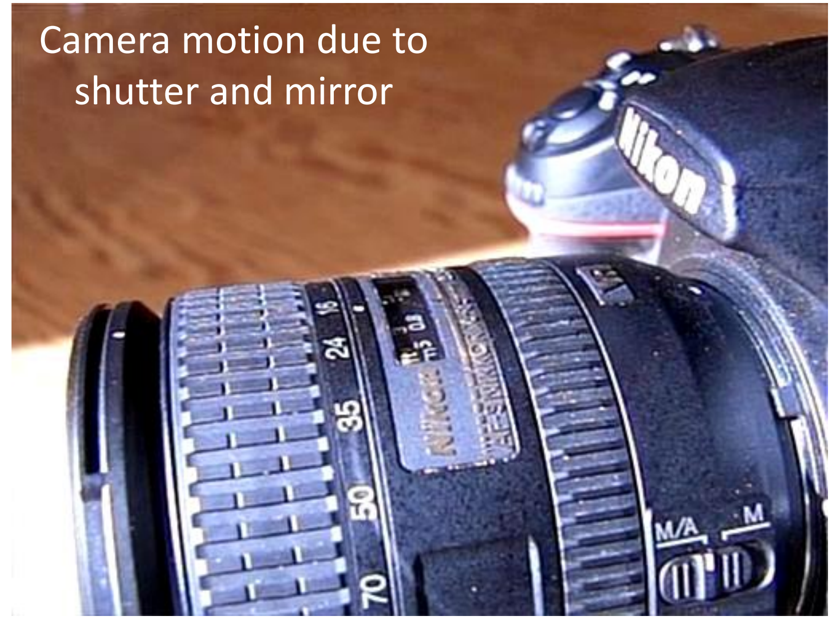
Imperceptible Changes in the World



Imperceptible Changes in the World



Blood flow



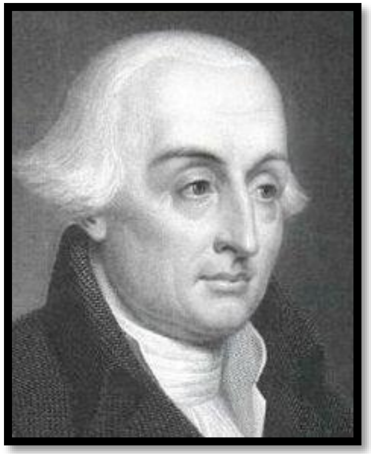
Camera motion due to shutter and mirror

Breathing

Micro-expressions

Lagrangian and Eulerian Perspectives (Fluid Dynamics)

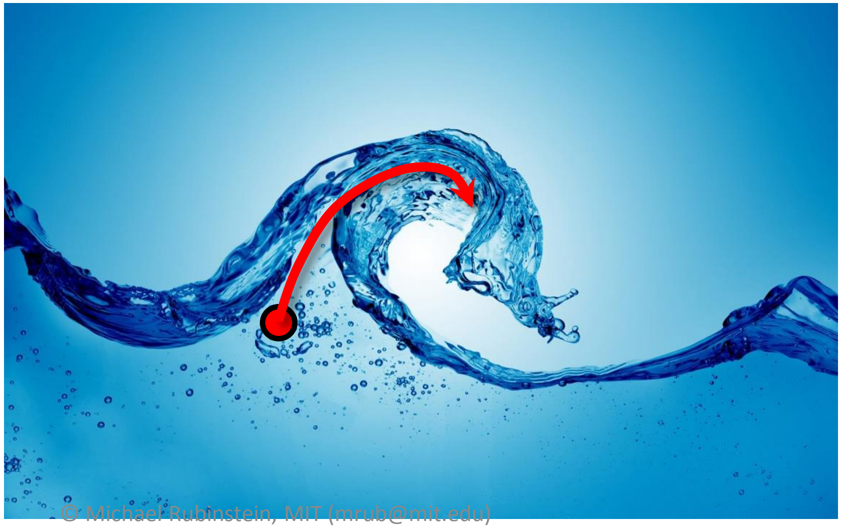
- Specifications of physical measurements through space and time:



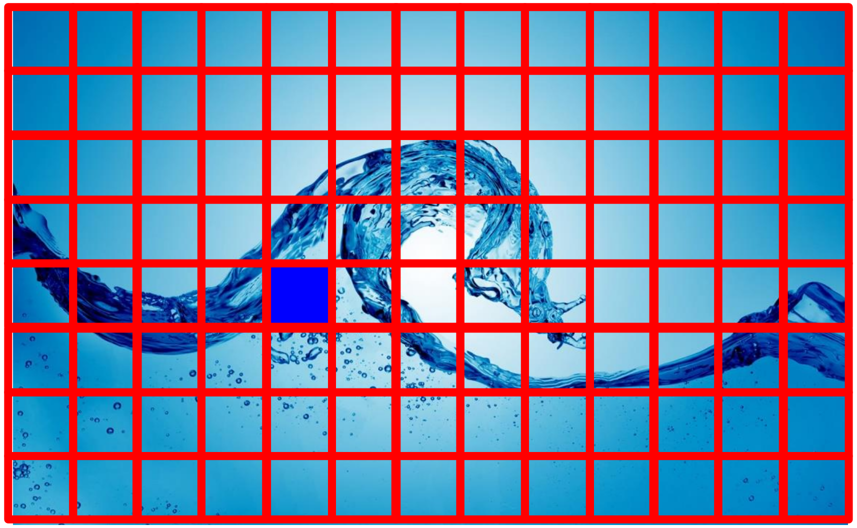
Lagrangian

Eulerian

Track particles



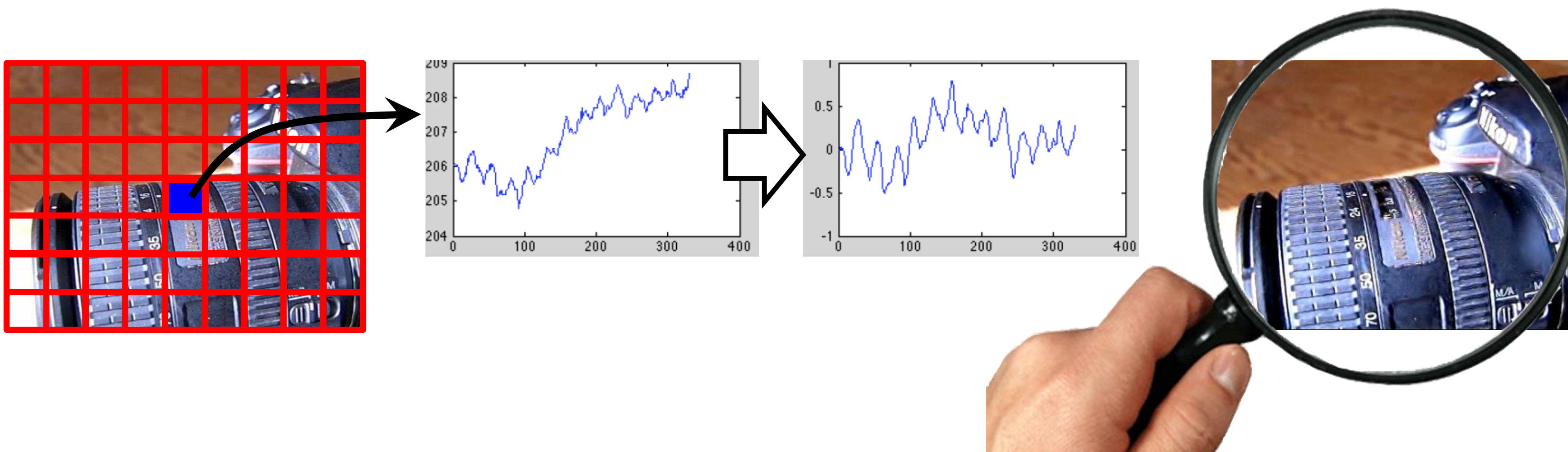
Measure changes within fixed voxels in space



Basic Idea

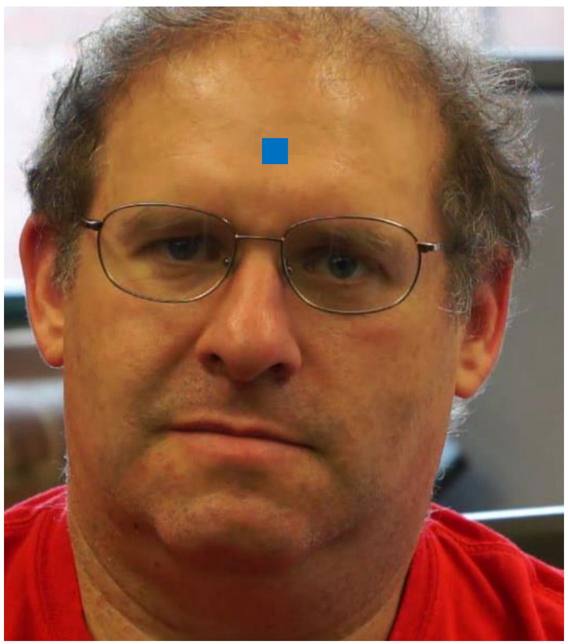
- **Amplify temporal pixel color variations**
  - Each pixel processed <u>independently</u>
  - Treat each pixel as a time series
  - Apply standard 1D signal processing to it
  - Amplify particular <u>temporal</u> frequencies

Eulerian Video Magnification (SIGGRAPH 2012)
With Hao-Yu Wu, Eugene Shih, John Guttag, Fredo Durand, Bill Freeman

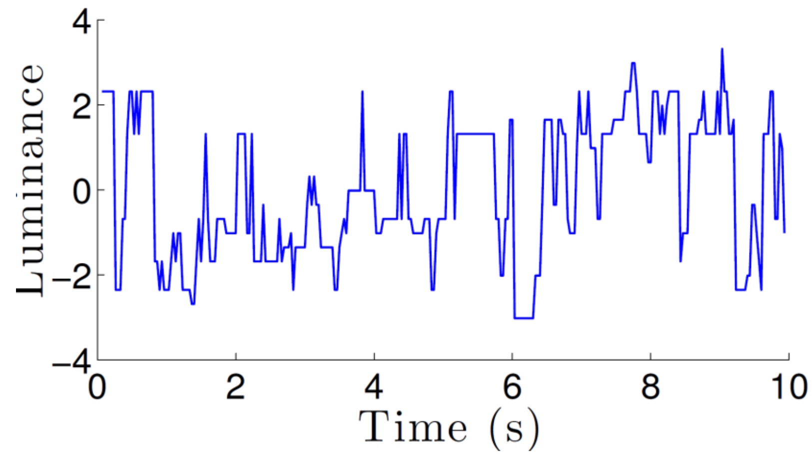


Subtle Color Variations

- The face gets slightly redder when blood flows
  - Very low amplitude: 0.5 intensity level in an 8-bit scale (0-255)



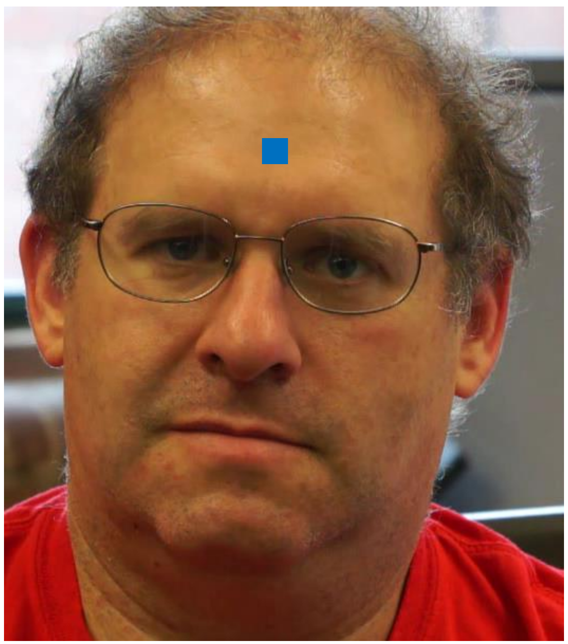
Input frame



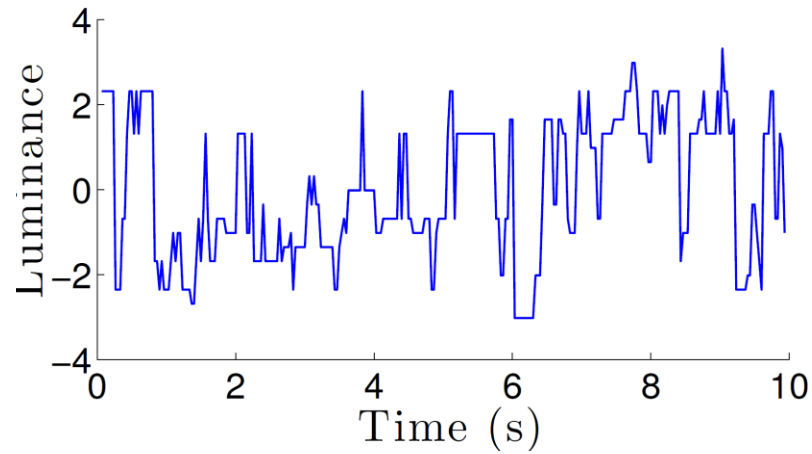
Luminance trace (zero mean)

Subtle Color Variations

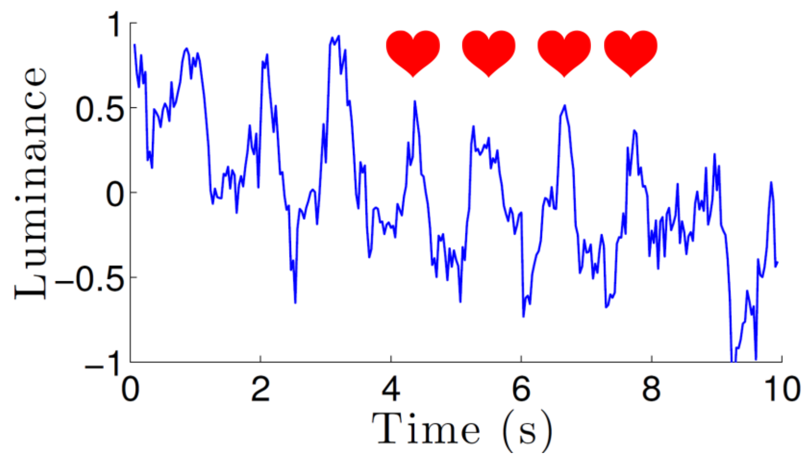
1. Average spatially to overcome sensor and quantization noise



Input frame



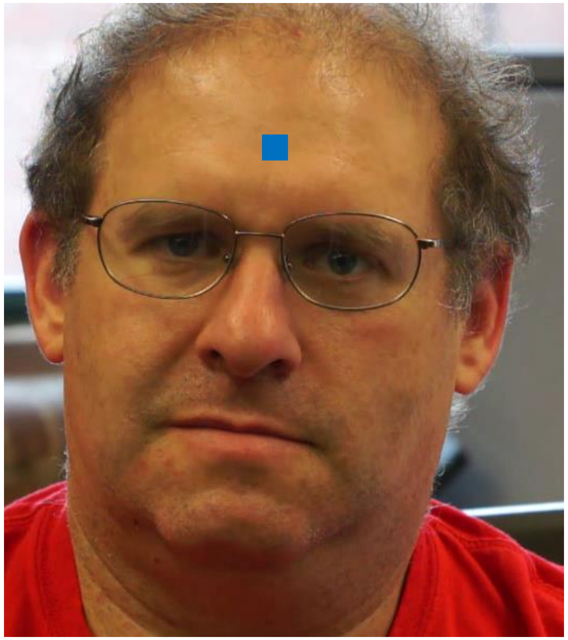
Luminance trace (zero mean)



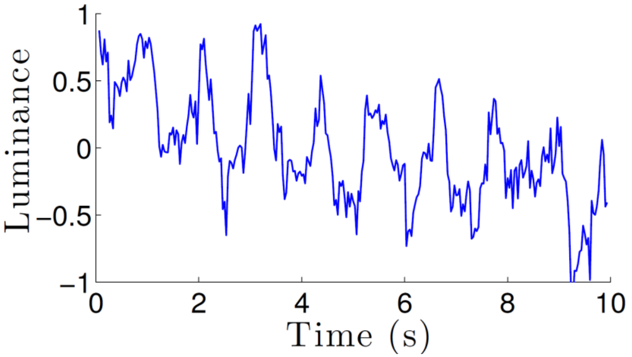
Spatially averaged luminance trace

Amplifying Subtle Color Variations

2. Filter temporally to extract the signal of interest

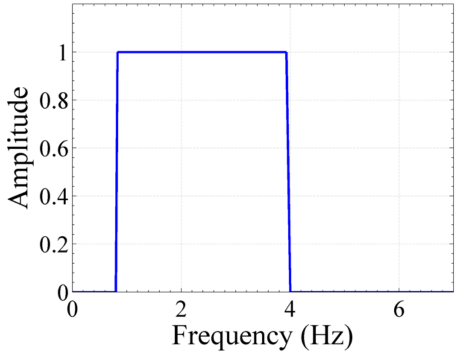


Input frame



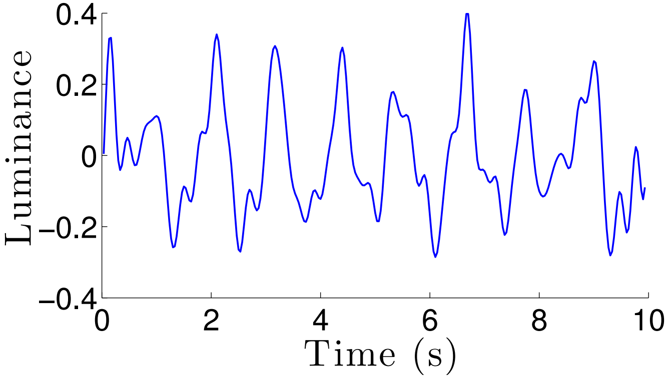
Spatially averaged luminance trace

⊗



Temporal filter

=



Temporally bandpassed trace

Color Amplification Results

Source

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

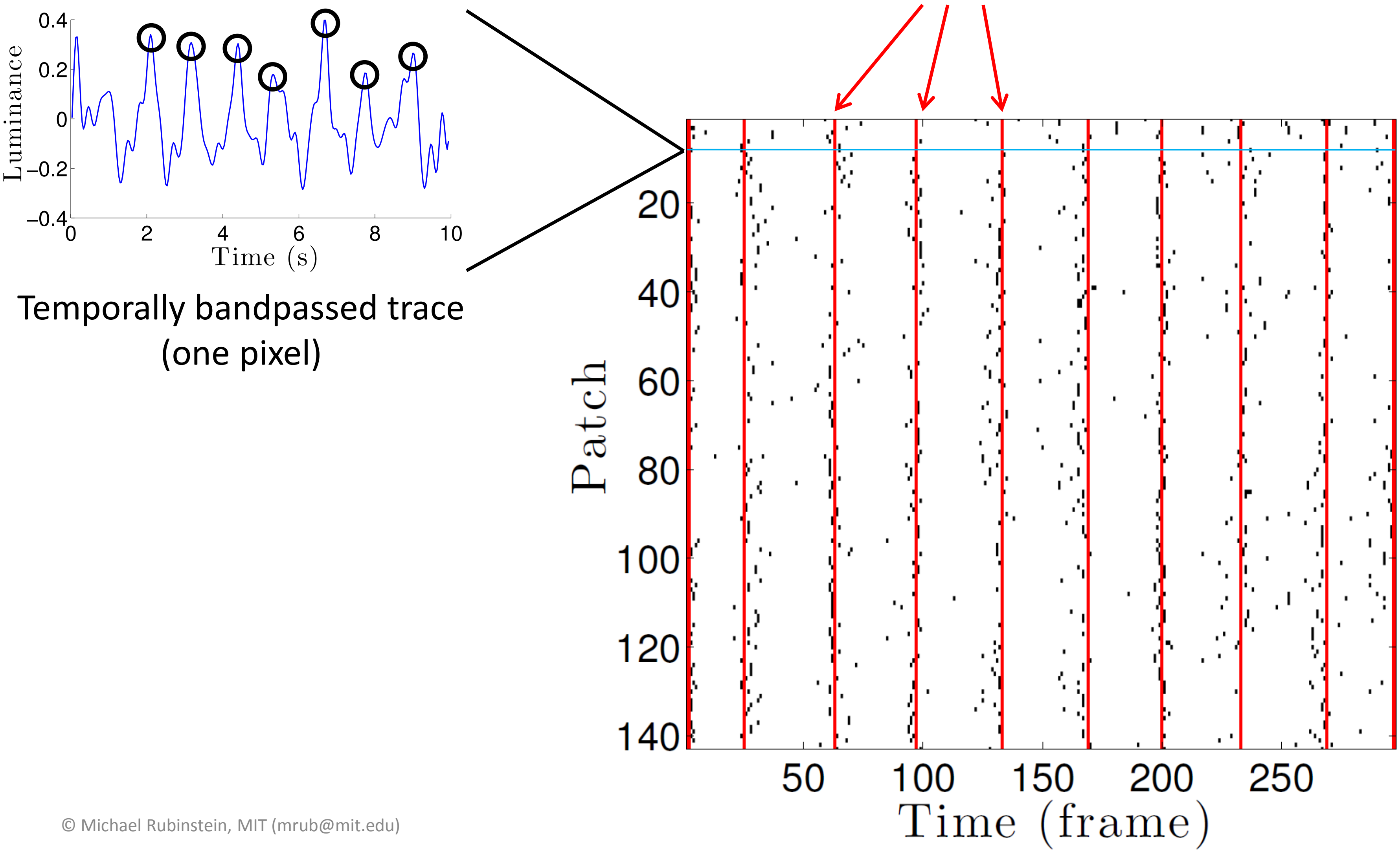
Bruce Wayne's Pulse

Christian Bale, Batman Begins (2005)

Courtesy of Warner Bros. Pictures

Heart Rate Extraction

Peak detection



Temporally bandpassed trace (one pixel)

Extracting Heart Rate

Source
(Courtesy of Winchester Hospital. Do not copy)

EKG

Hospital monitor

Bandpass signal + peaks (pulse)

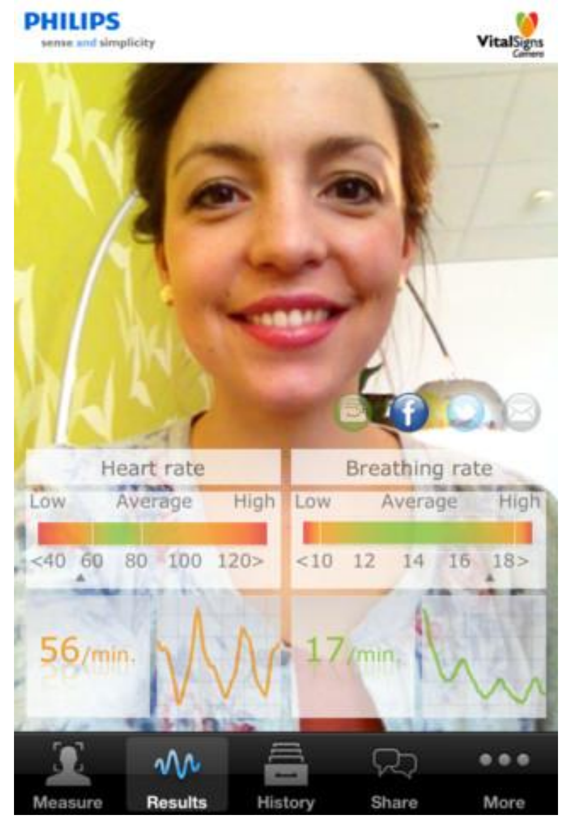
Estimated heart rate

Color-amplified (x150) 2.33-2.67 Hz (140-160 bpm)

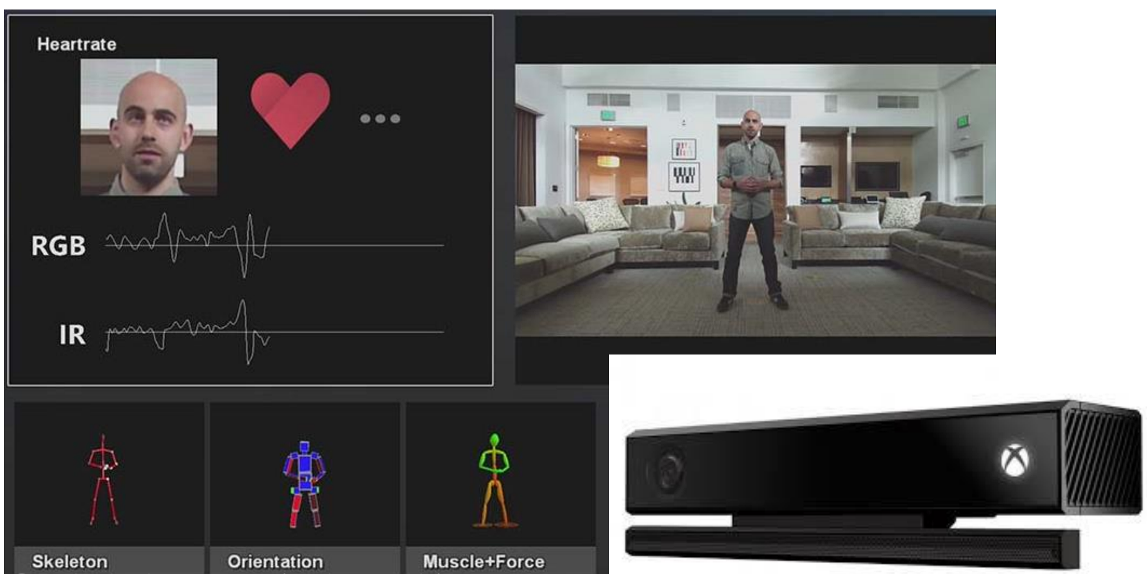
With Dr. Donna Brezinski and the Winchester Hospital staff

Related Work: Pulse Detection from Video

"Cardiocam" [Pho, Picard, McDuff 2010]



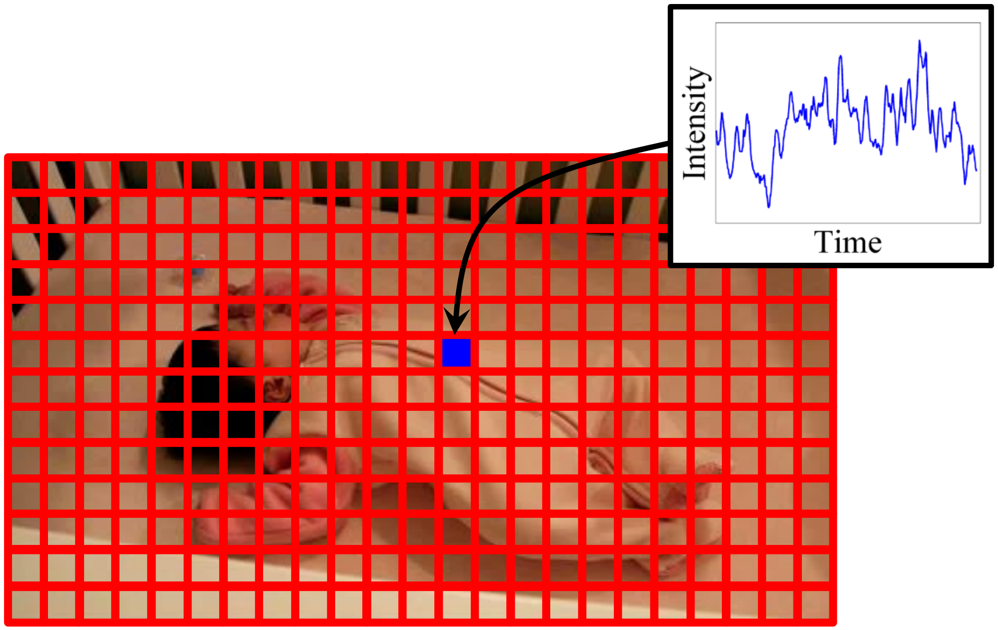
"Vital Signs Camera" – Philips
proprietary



Kinect (Xbox One)
proprietary

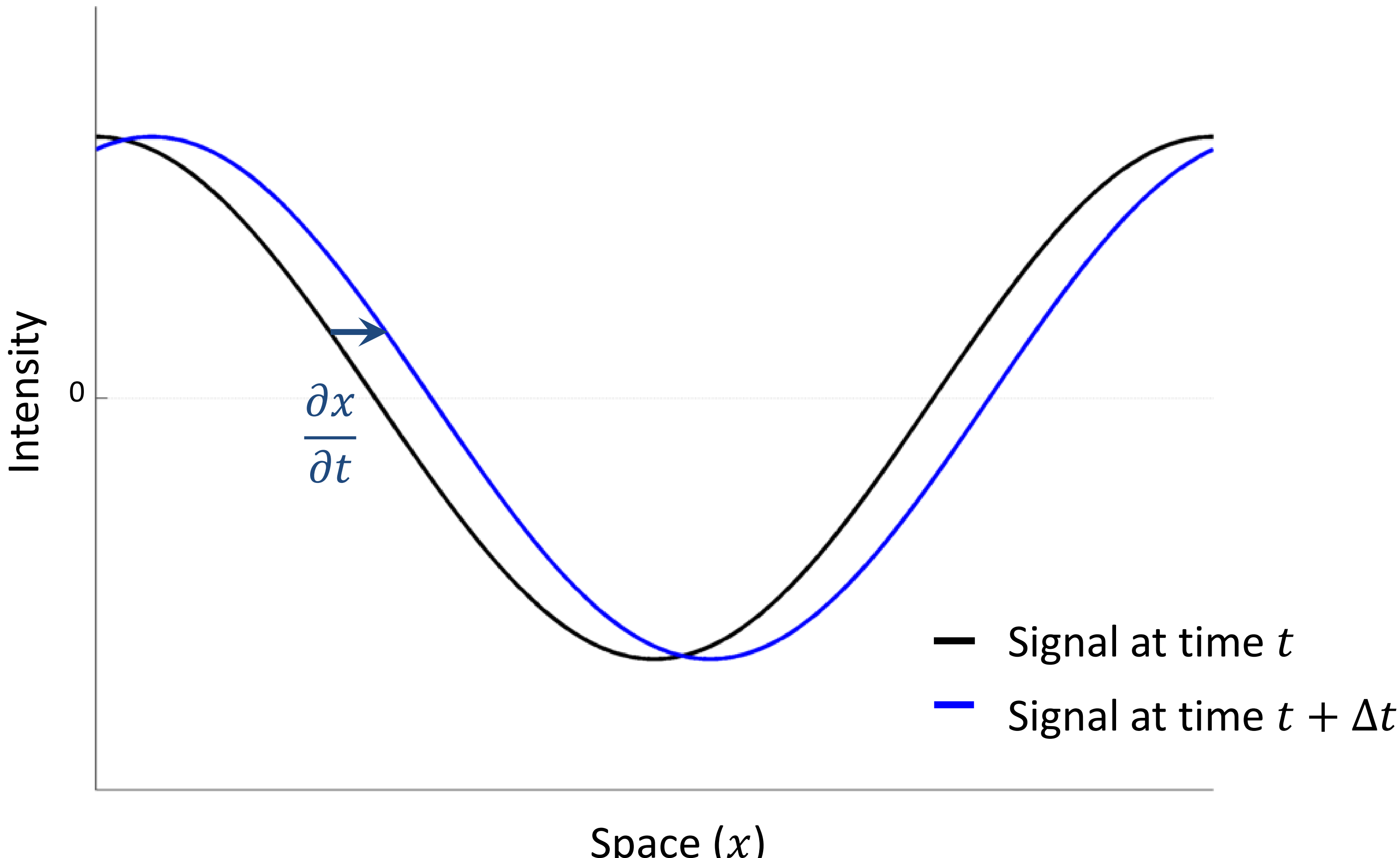
Why Does it Amplify Motion?

- By increasing temporal variation – we can increase spatial motion!



Differential Brightness Constancy

- Scenario: a 1D translating image profile

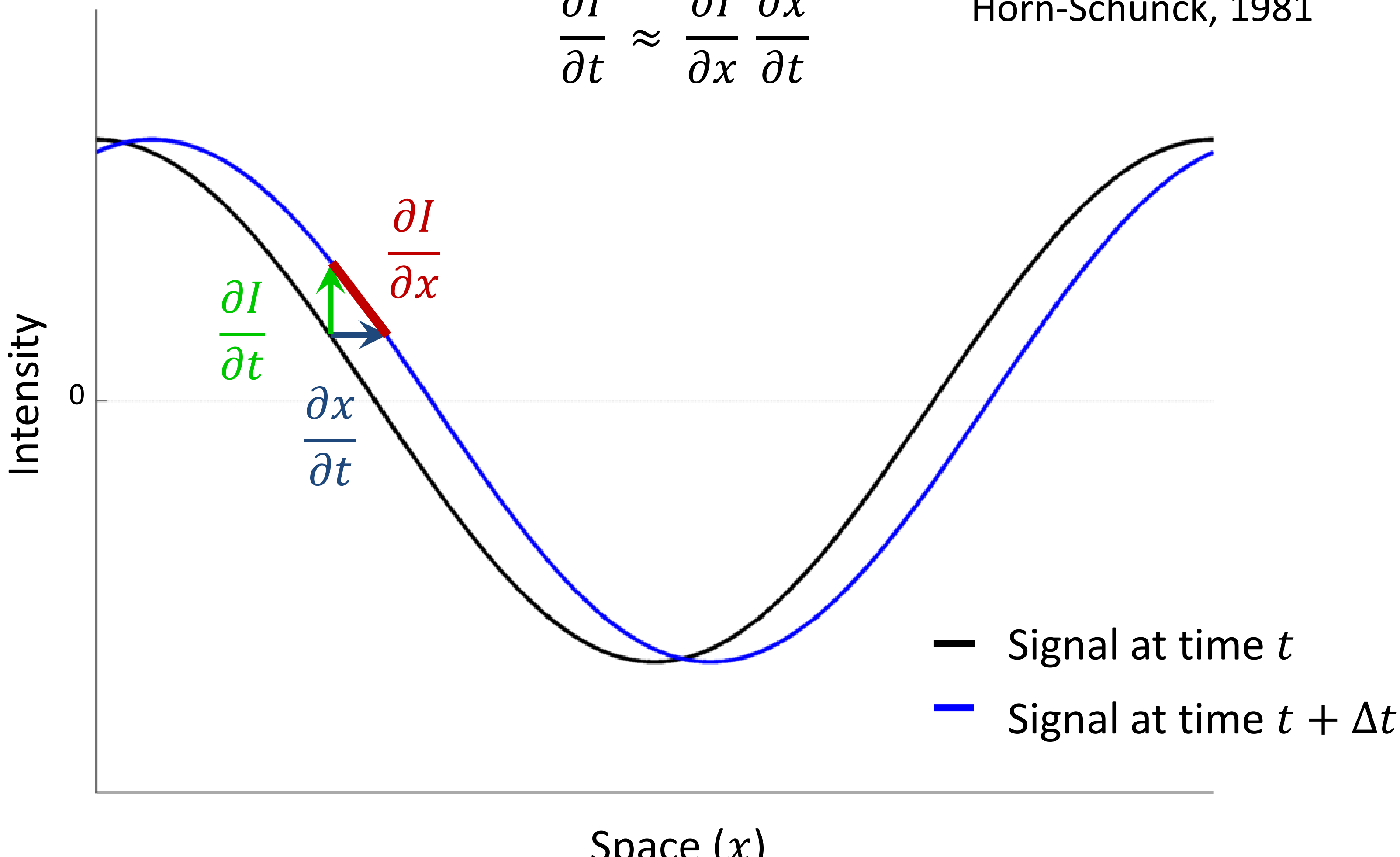


Differential Brightness Constancy

- Measure temporal variation $\partial I / \partial t$ (at each pixel)

Lucas-Kanade, 1981
Horn-Schunck, 1981

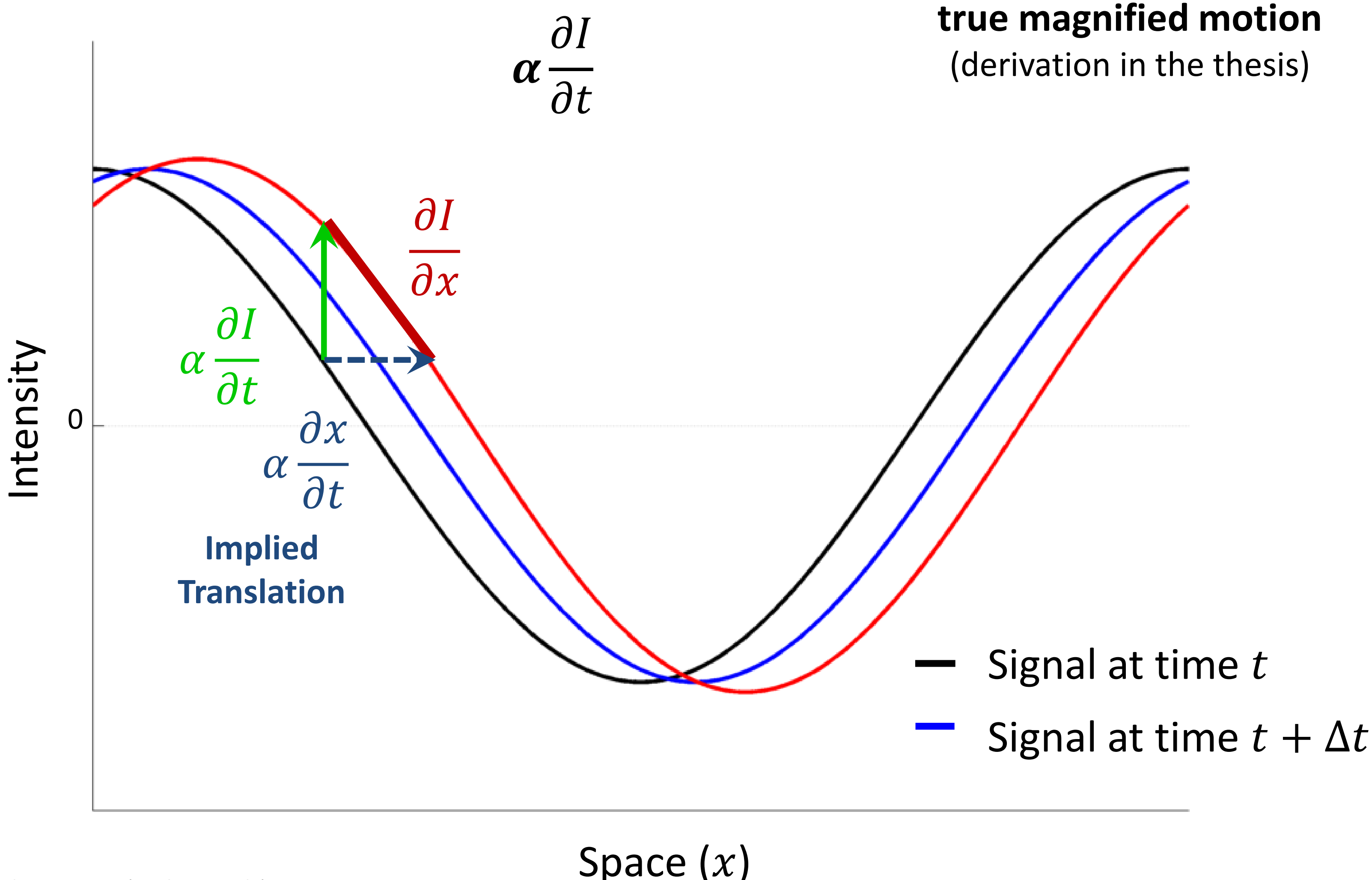
$$\frac{\partial I}{\partial t} \approx \frac{\partial I}{\partial x} \frac{\partial x}{\partial t}$$



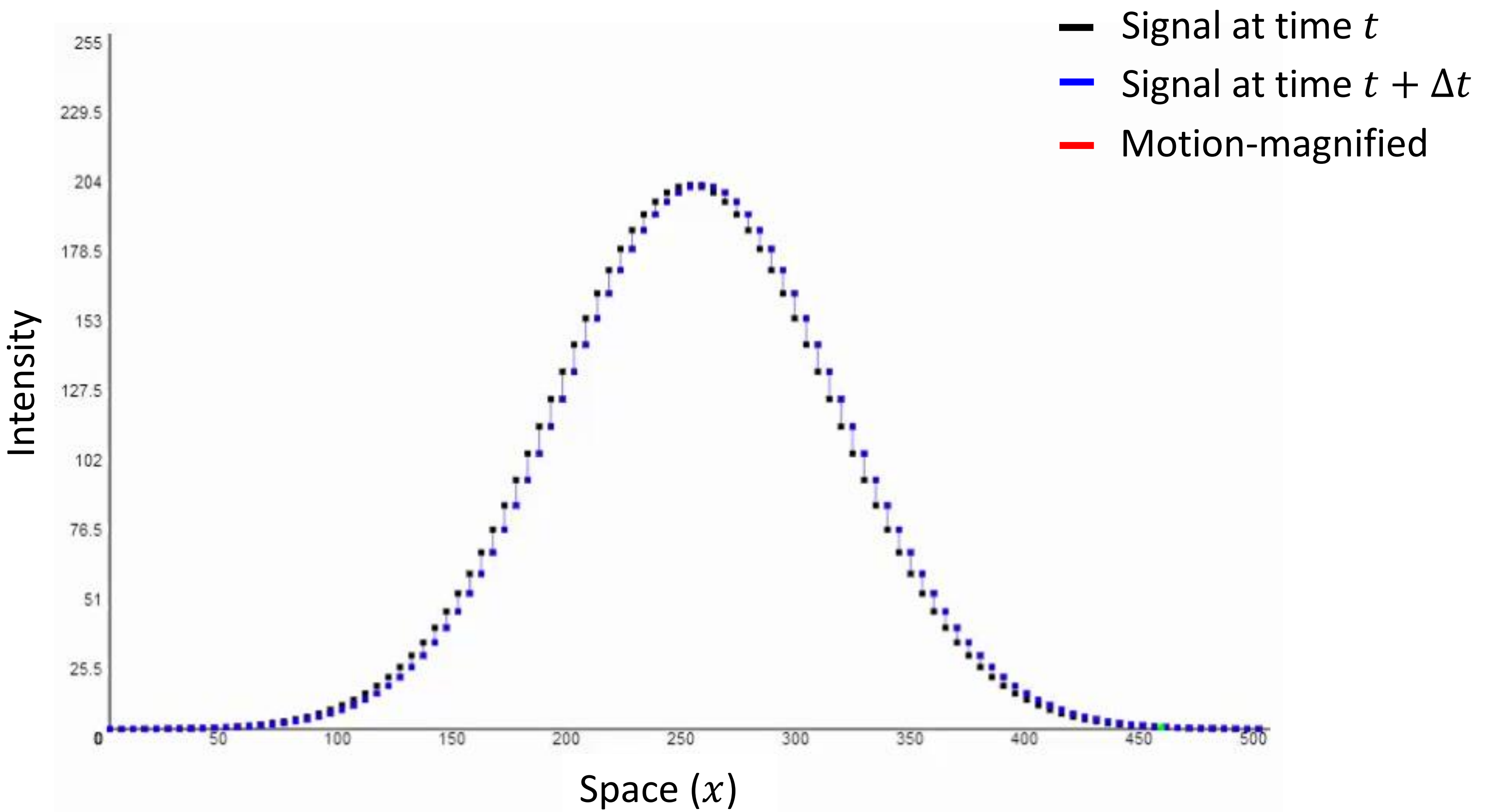
Eulerian Motion Magnification

- Amplify temporal variation $\partial I/\partial t$ (at each pixel)

First-order (linear) approximation to the true magnified motion
(derivation in the thesis)

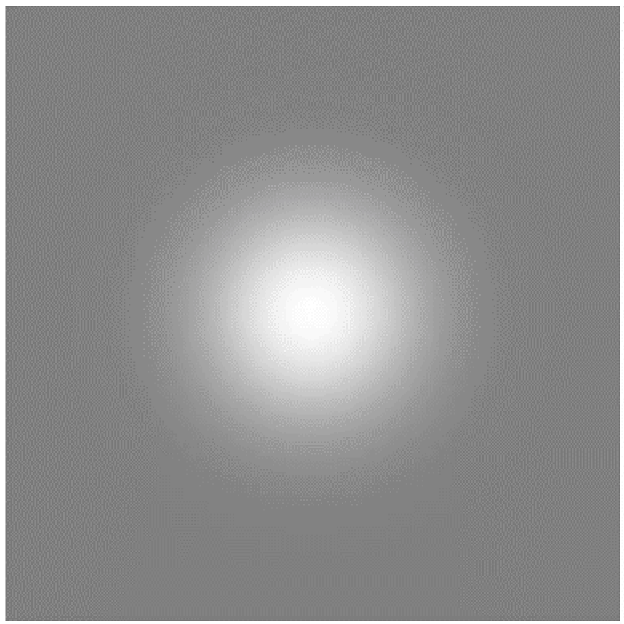


Relating Temporal and Spatial Changes



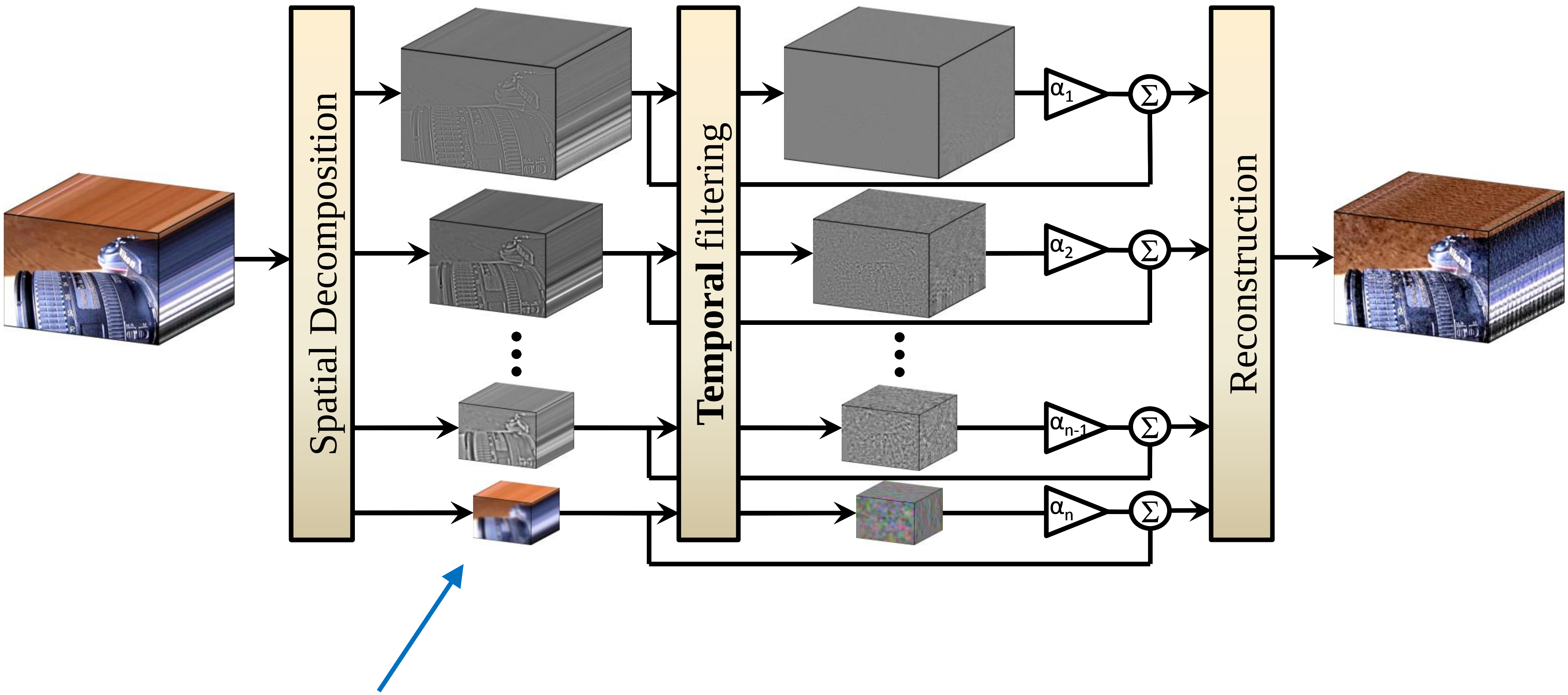
Courtesy of Lili Sun

Synthetic 2D Example



Source

Method Pipeline



Laplacian pyramid
[Burt and Adelson 1983]

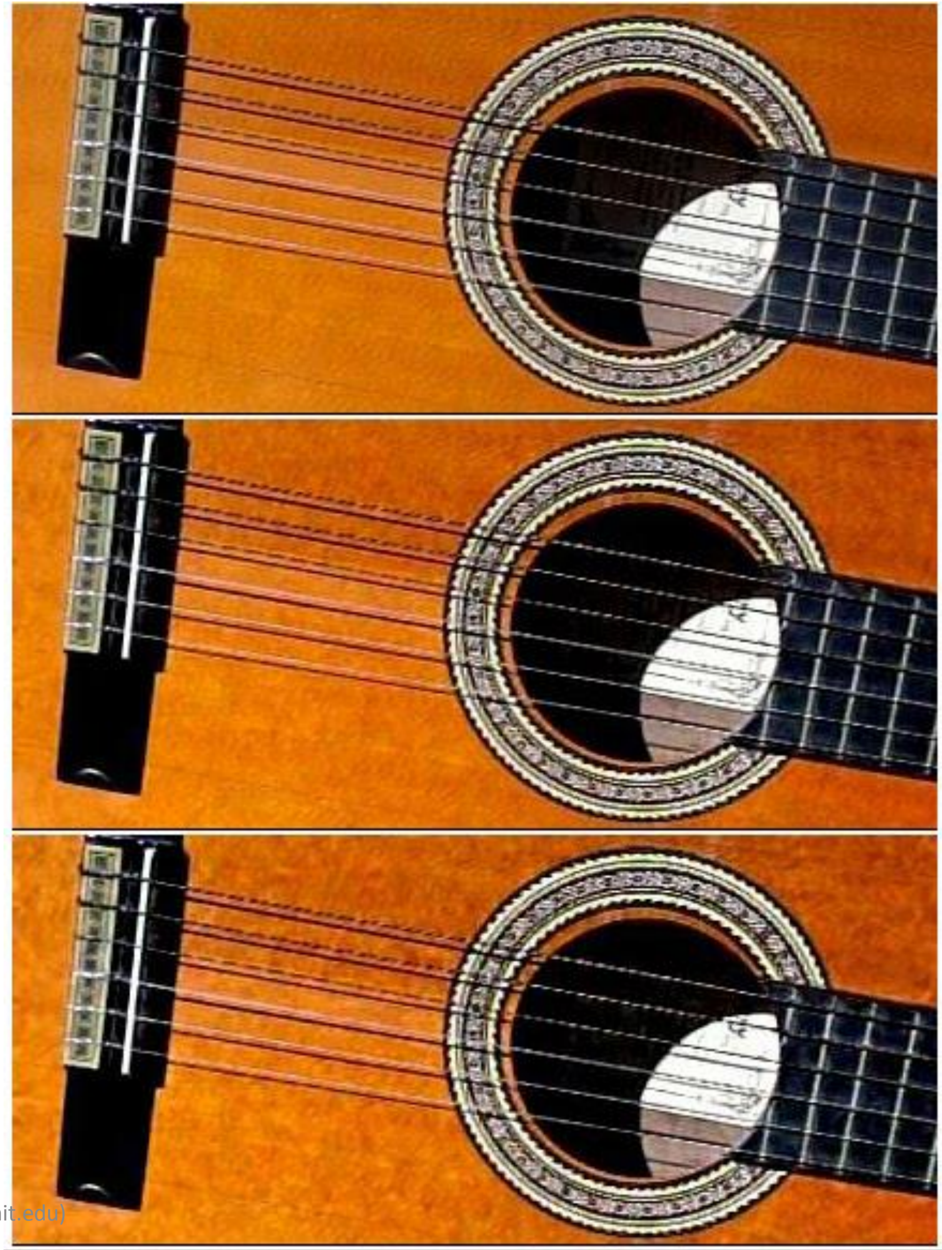
Motion Magnification Results

Source

Selective Motion Magnification

Source
(**600 fps**)

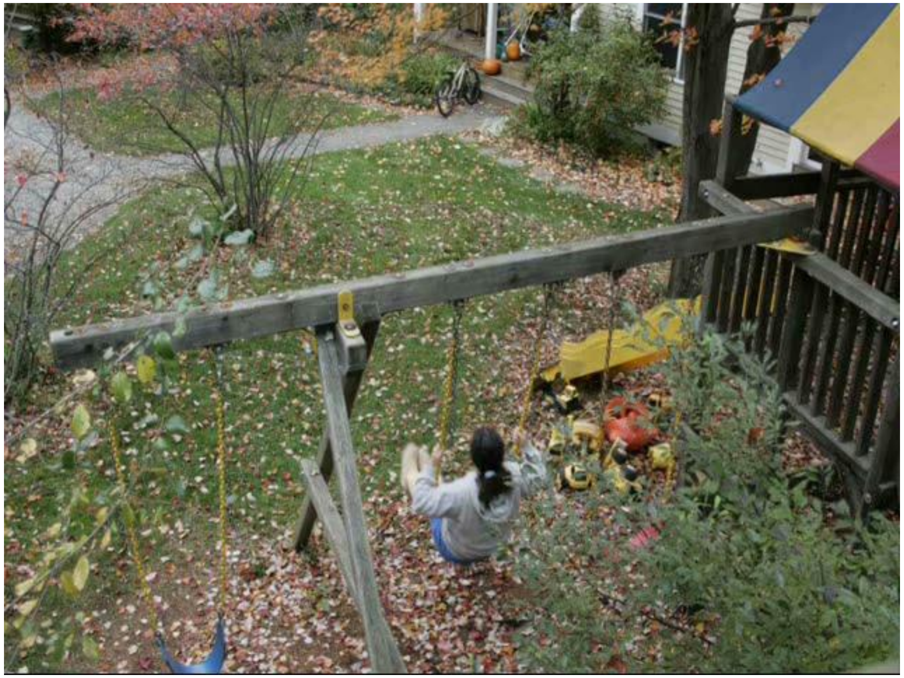
72-92 Hz
Amplified

100-120 Hz
Amplified

Low E (82.4 Hz)

A (110 Hz)

Related Work: Motion Magnification [Liu 2005]

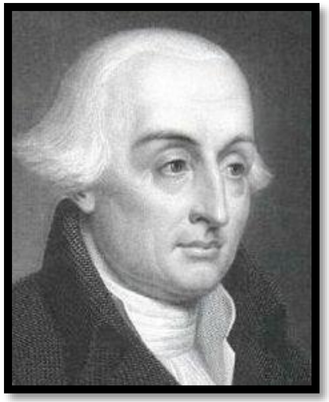
Source



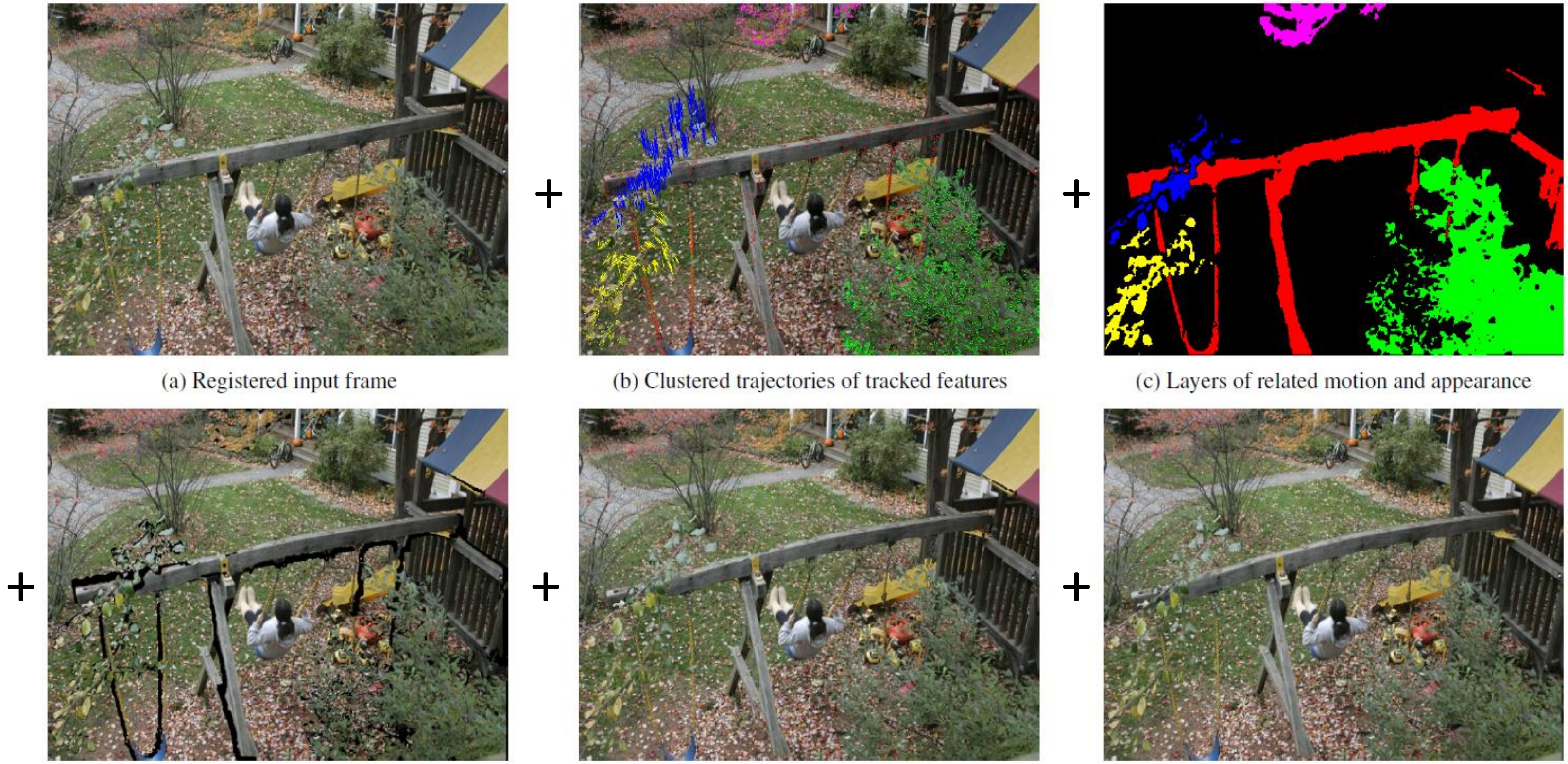
Motion-magnified

Liu et al. *Motion Magnification*, SIGGRAPH 2005

Related Work: Motion Magnification [Liu 2005]



- Better for large motions, point features, occlusions, but…
- Requires motion analysis, motion segmentation, inpainting
  - Nontrivial to do artifact-free
  - Computationally intensive



(a) Registered input frame

(b) Clustered trajectories of tracked features

(c) Layers of related motion and appearance

(d) Motion magnified, showing holes

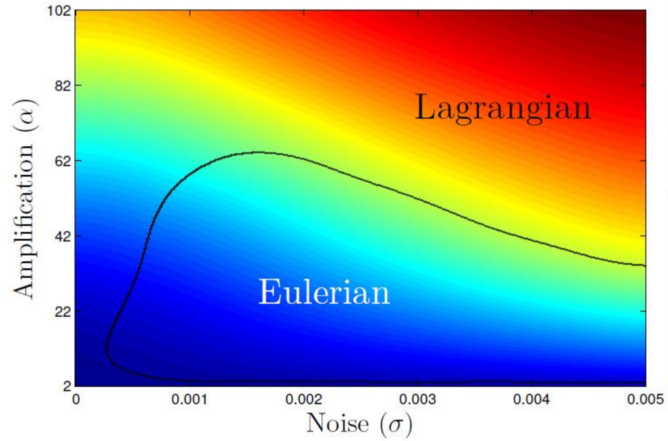
(e) After texture in-painting to fill holes

(f) After user's modifications to segmentation map in (c)

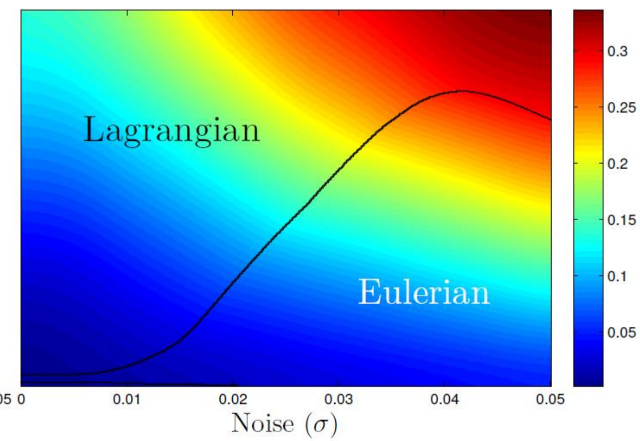
Liu et al. *Motion Magnification*, SIGGRAPH 2005

Lagrangian vs. Euleriann

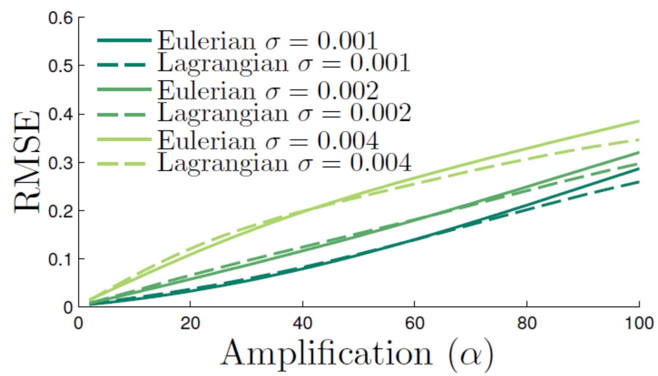
- See my thesis for more details!



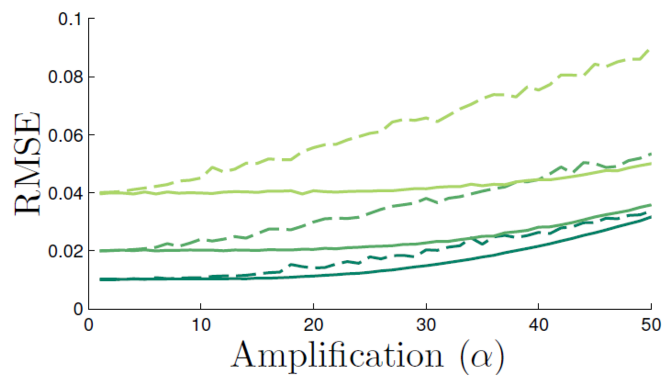
Point-wise optical flow



Spatially-regularized optical flow



Spatiotemporal noise



Spatial noise only

Limitations

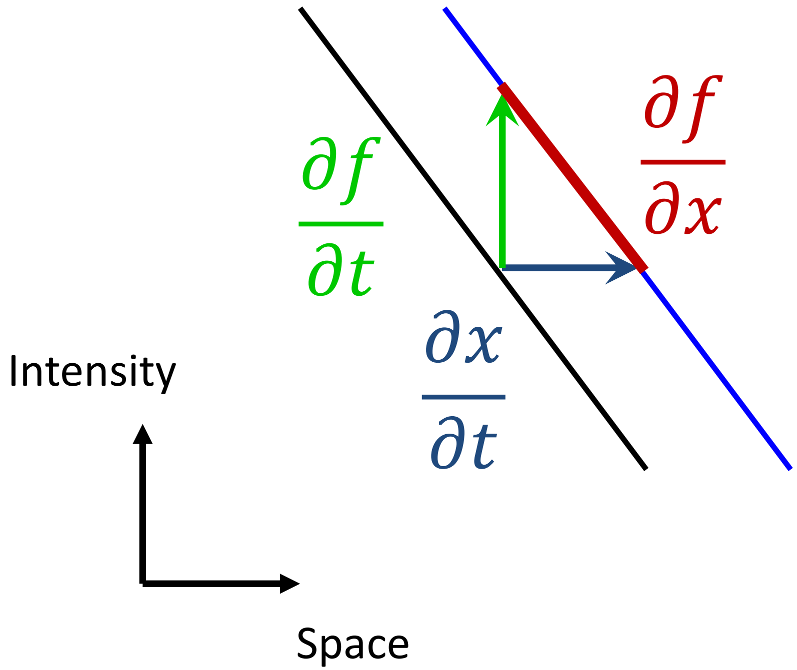
Amplified noise

Source

Motion-magnified

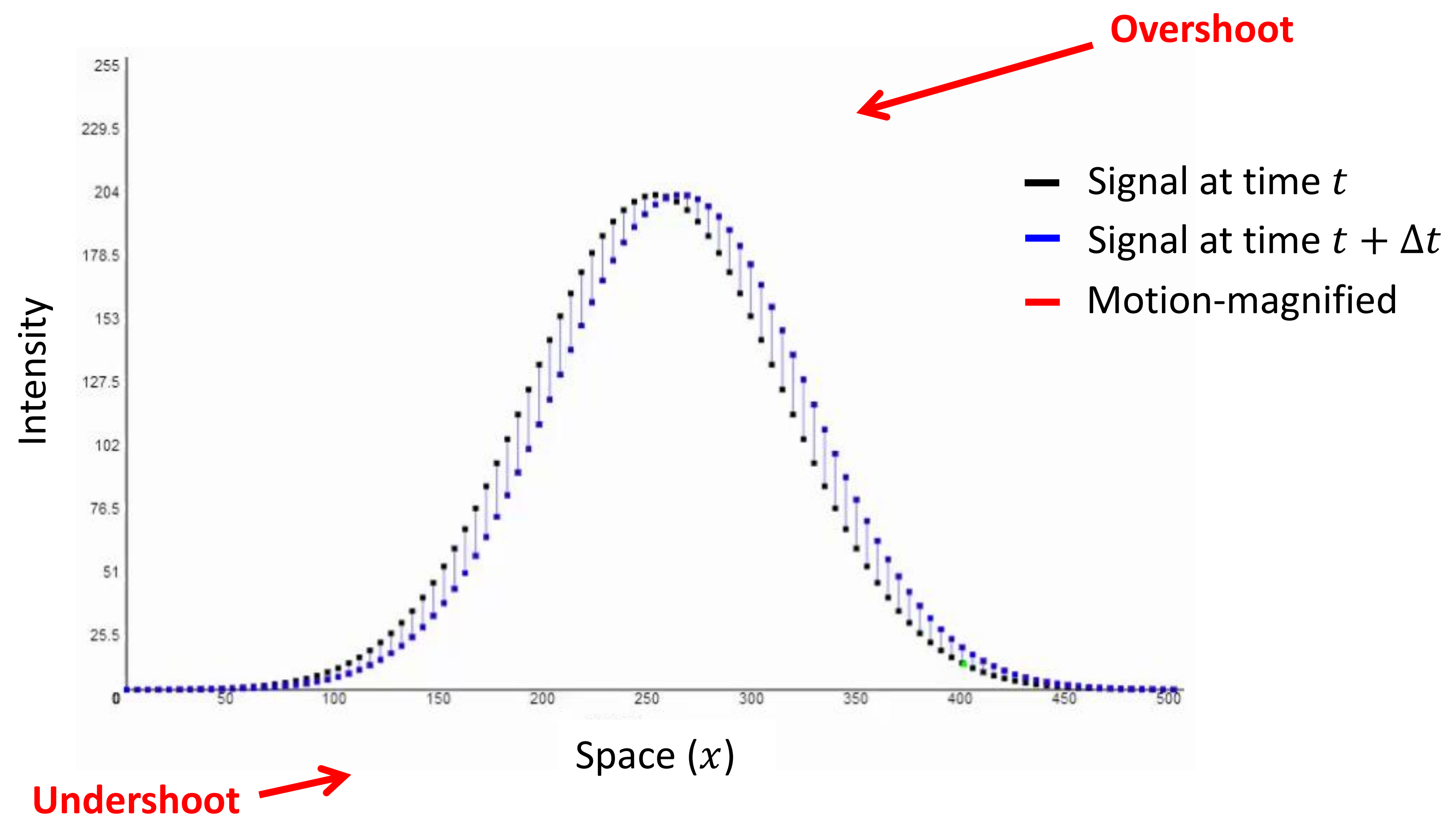
Limitations of Linear Motion Processing

- Assumes image intensity is locally linear



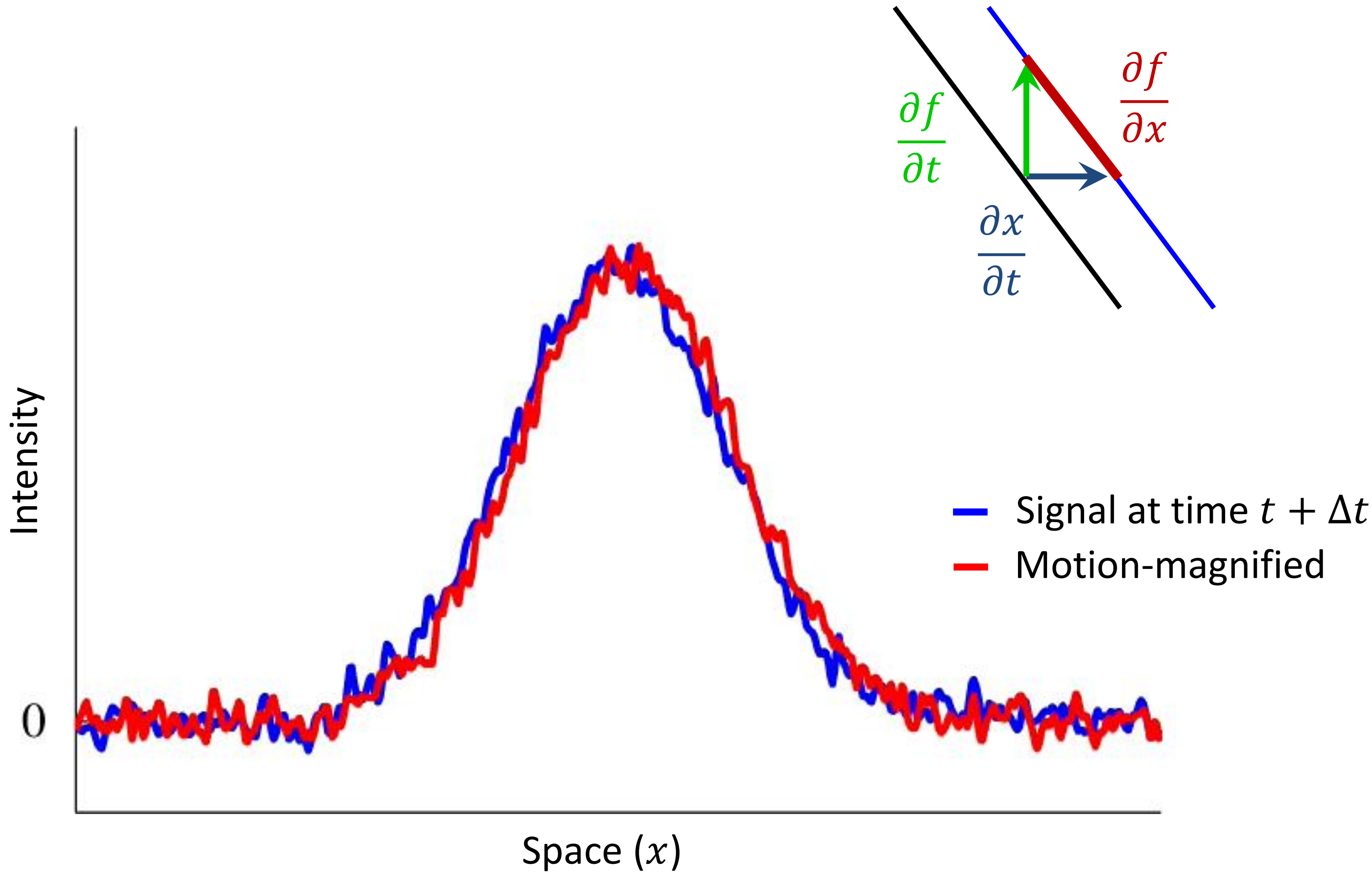
Limitations of Linear Motion Processing

- Breaks at high spatial frequencies and large motions



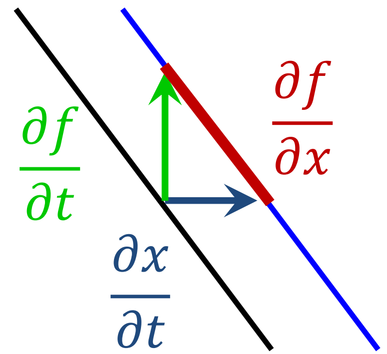
Limitations of Linear Motion Processing

- Noise amplified with signal

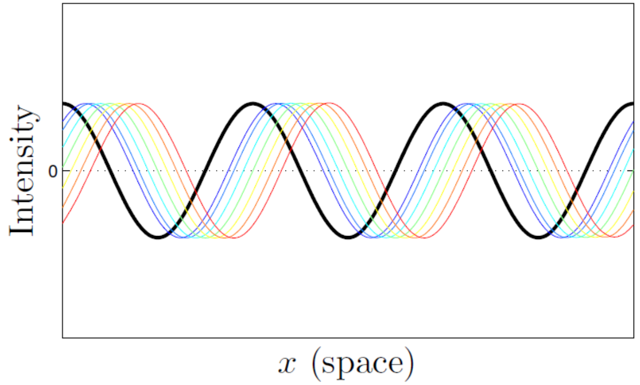


Linear vs. Phase-Based Motion Processing

- Linear motion processing
  - Assumes images are **locally linear**
  - Translate by **changing intensities**



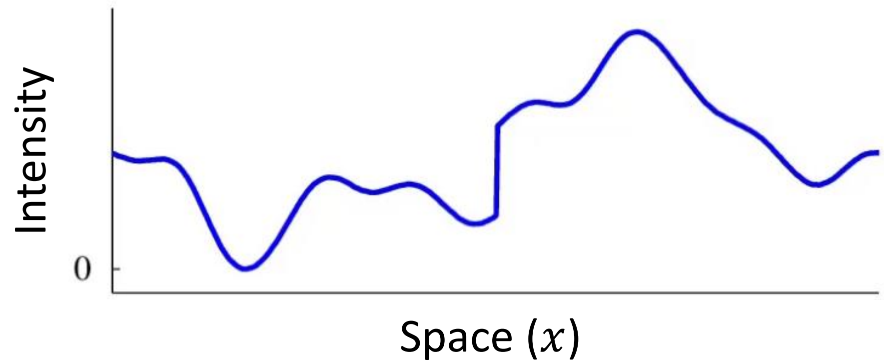
- **NEW** phase-based motion processing
  - Represents images as collection of **local sinusoids**
  - Translate by **shifting phase**

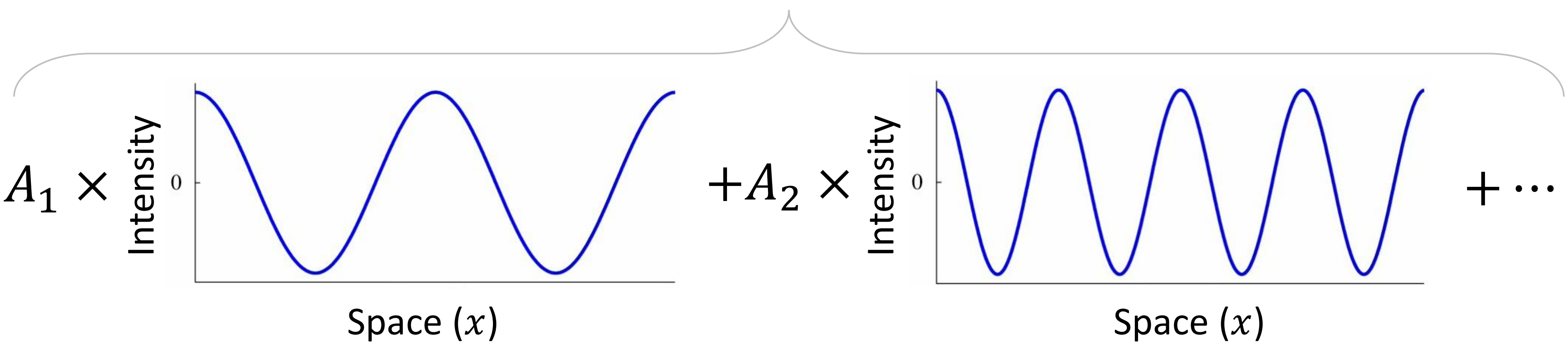


Phase-Based Video Motion Processing (SIGGRAPH 2013)
With Neal Wadhwa, Fredo Durand, Bill Freeman

Fourier Decomposition

$$I(x) \overset{\text{FFT}}{\Longleftrightarrow} \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega x}$$

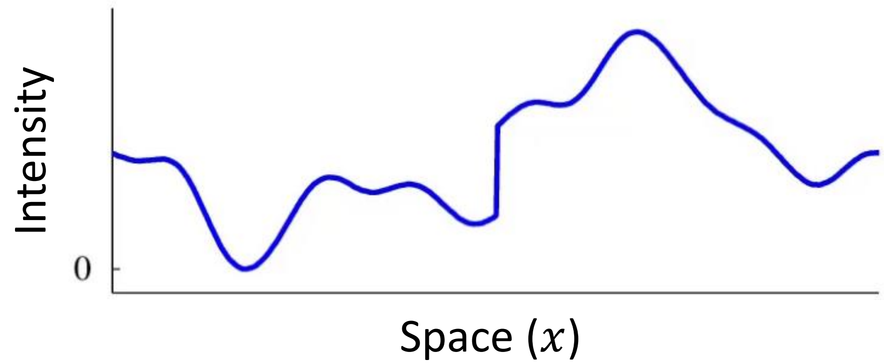


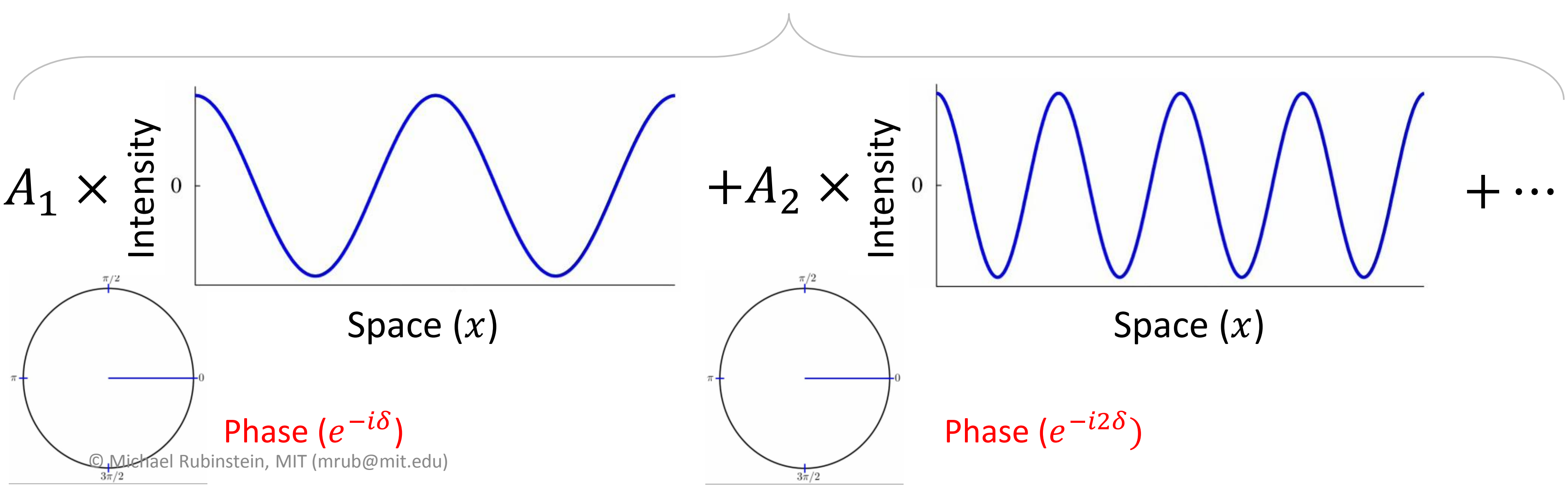


Fourier Shift Theorem

Phase shift ⇔ Translation

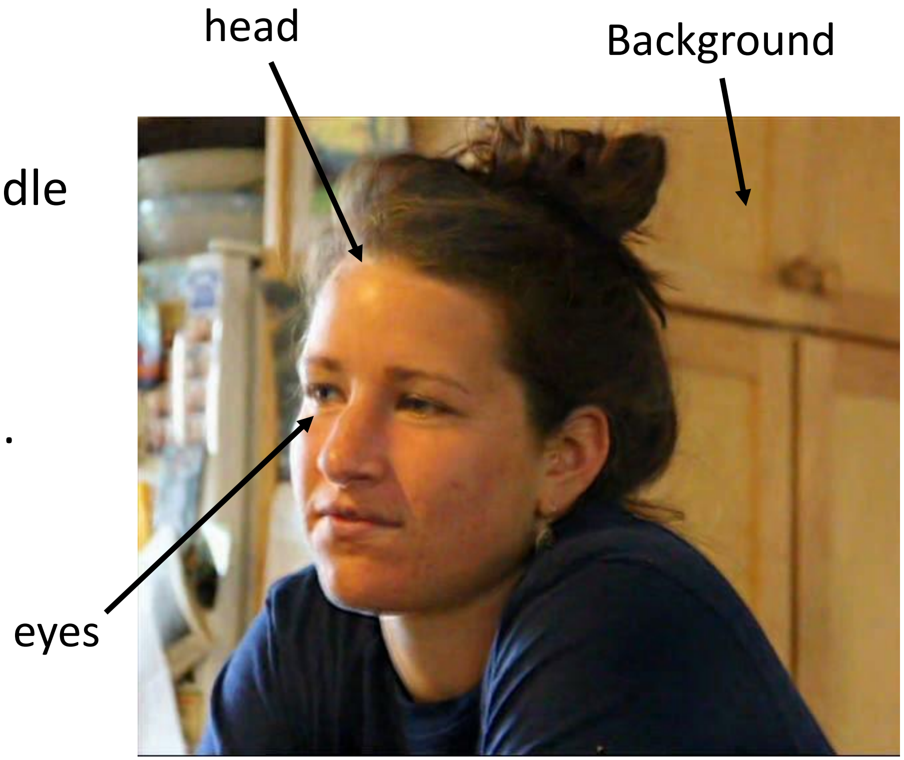
$$I(x-\delta) \overset{\text{FFT}}{\Longleftrightarrow} \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega x} e^{-i\omega\delta}$$





Local Motions

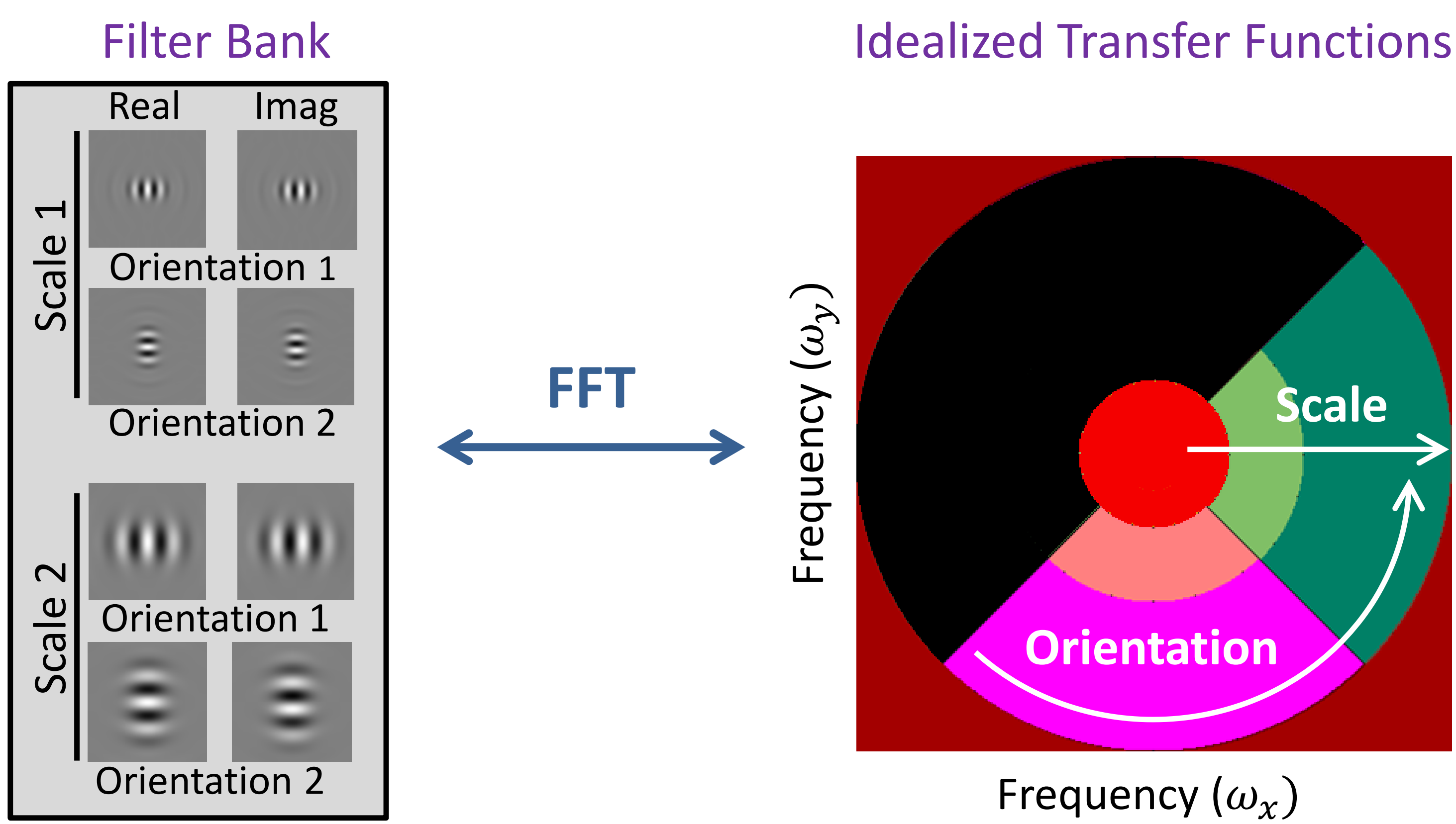
- Fourier shift theorem only lets us handle **global** motion
- But, videos have many local motions…



→ Need a localized Fourier Series for **local** motion

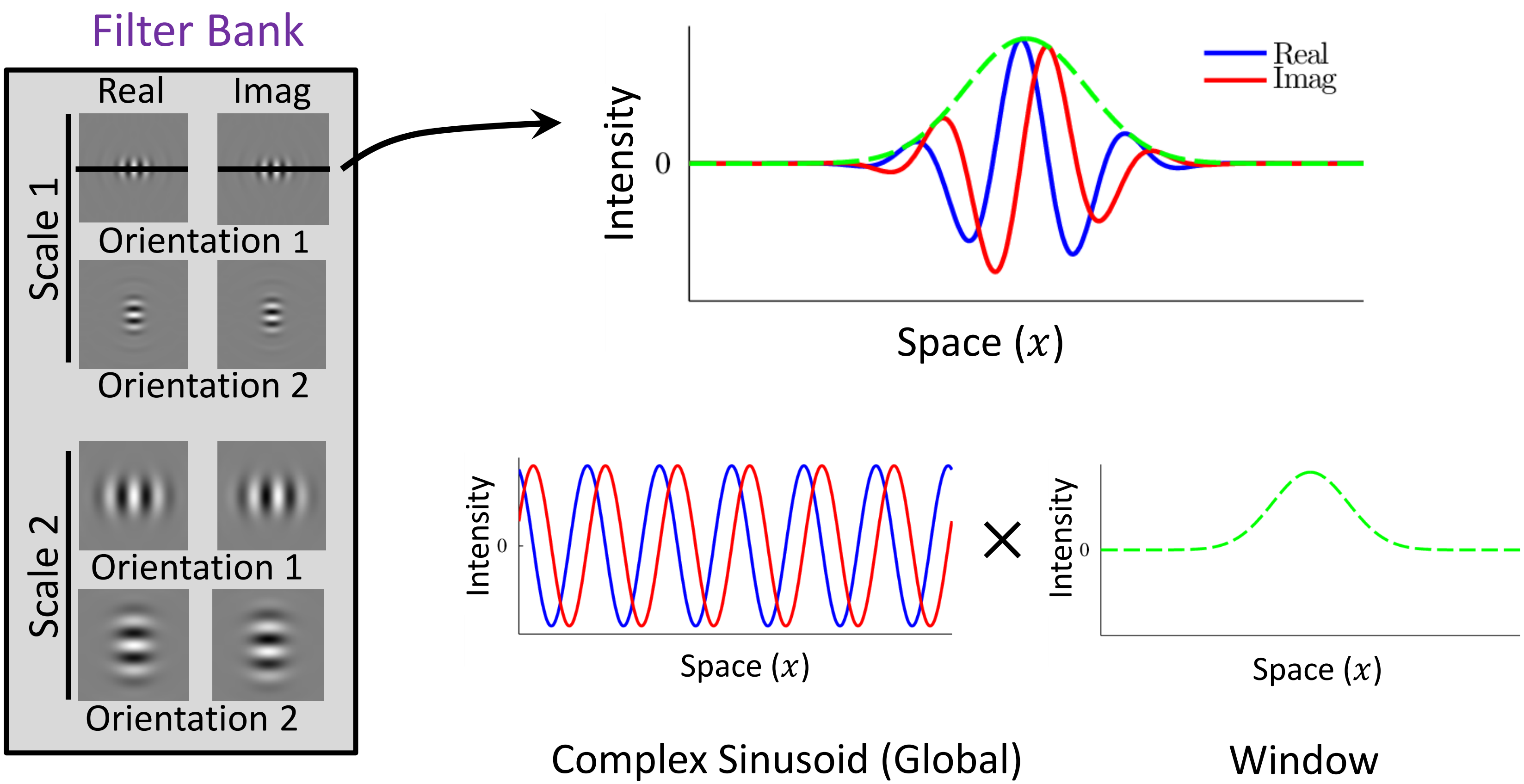
Complex Steerable Pyramid [Simoncelli, Freeman, Adelson, Heeger 1992]

- **Localized Fourier transform** that breaks the image into spatial structures at different scales and orientations



Complex Steerable Pyramid [Simoncelli, Freeman, Adelson, Heeger 1992]

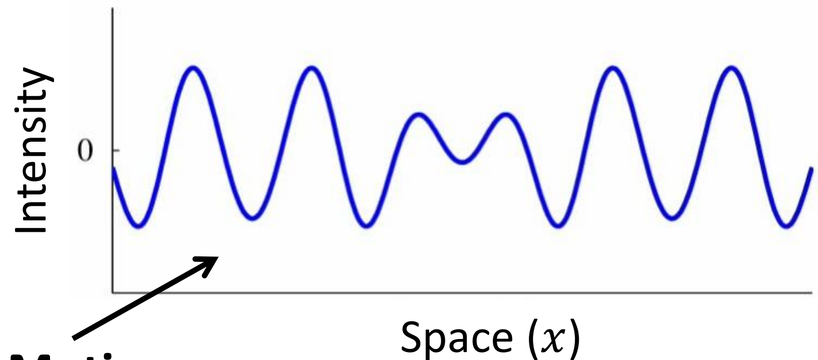
- Basis functions are wavelets with even (cosine) and odd (sine) components which give local amplitude and phase



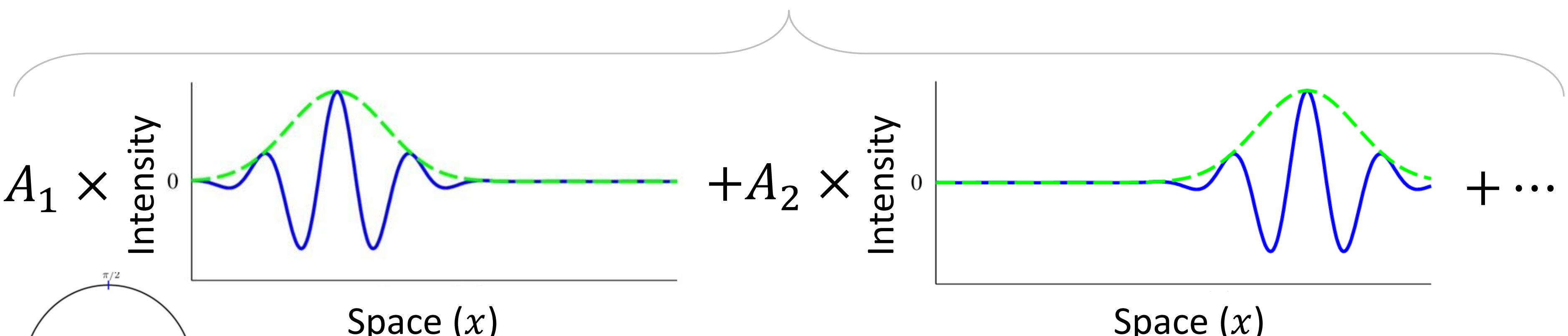
Local Phase

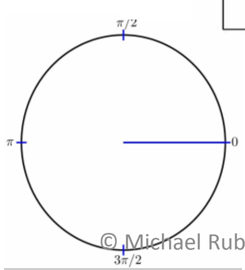
Local phase shift ↔ Local translation

- In a single subband, image is coefficients times translated copies of basis functions



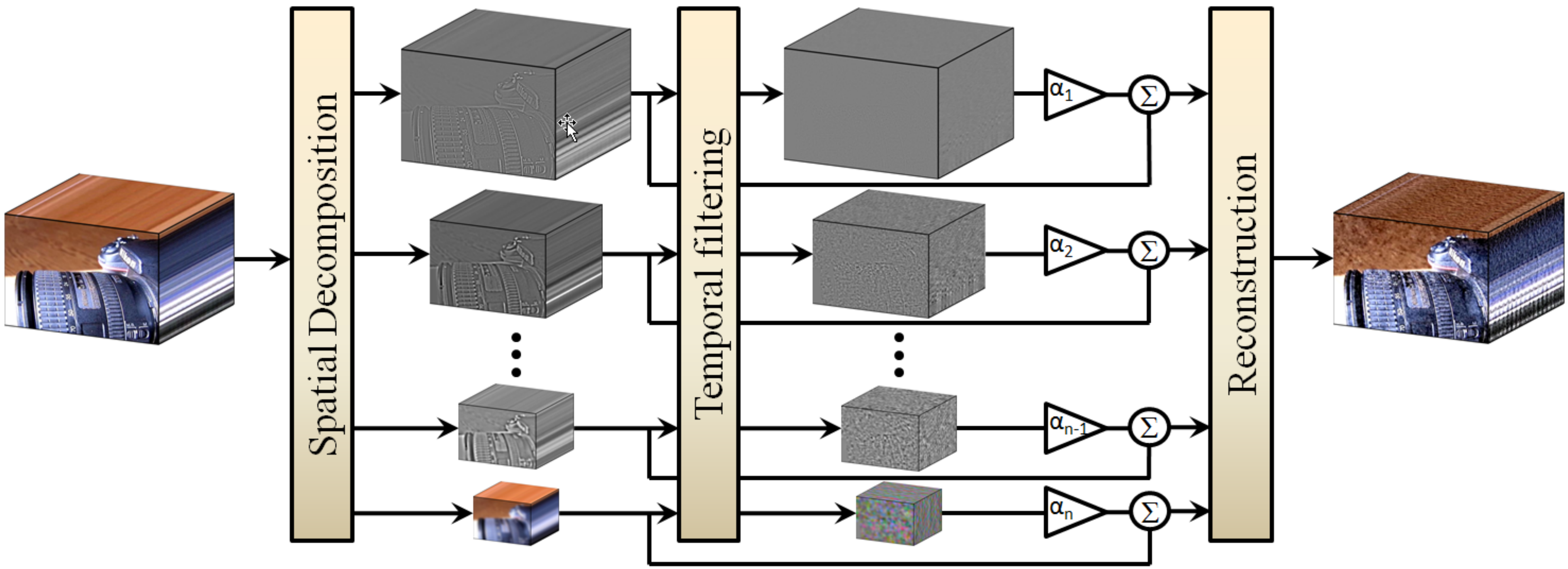
Local Motions





Local phase

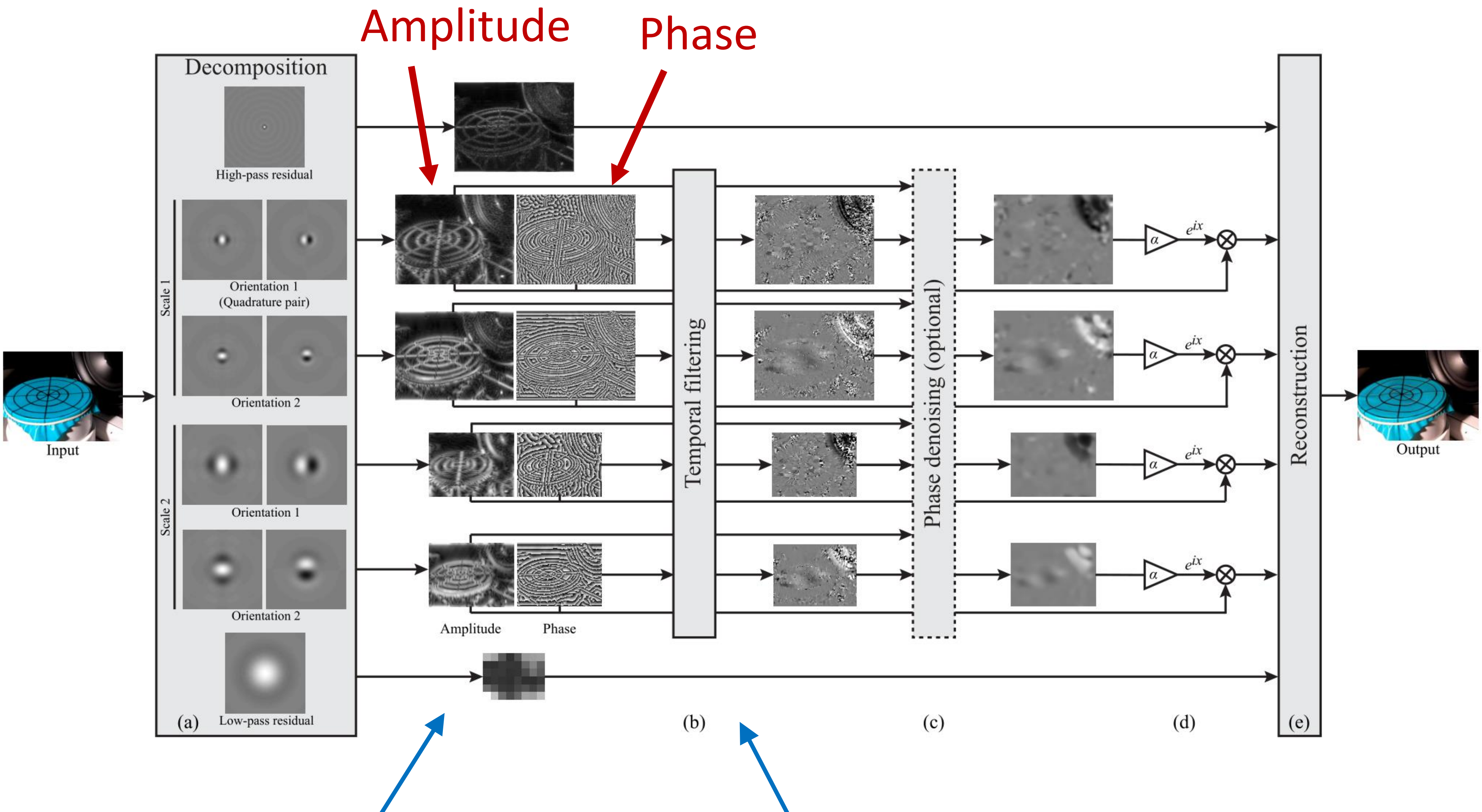
Linear Pipeline (SIGGRAPH'12)



Laplacian pyramid
[Burt and Adelson 1983]

Temporal filtering on **intensities**

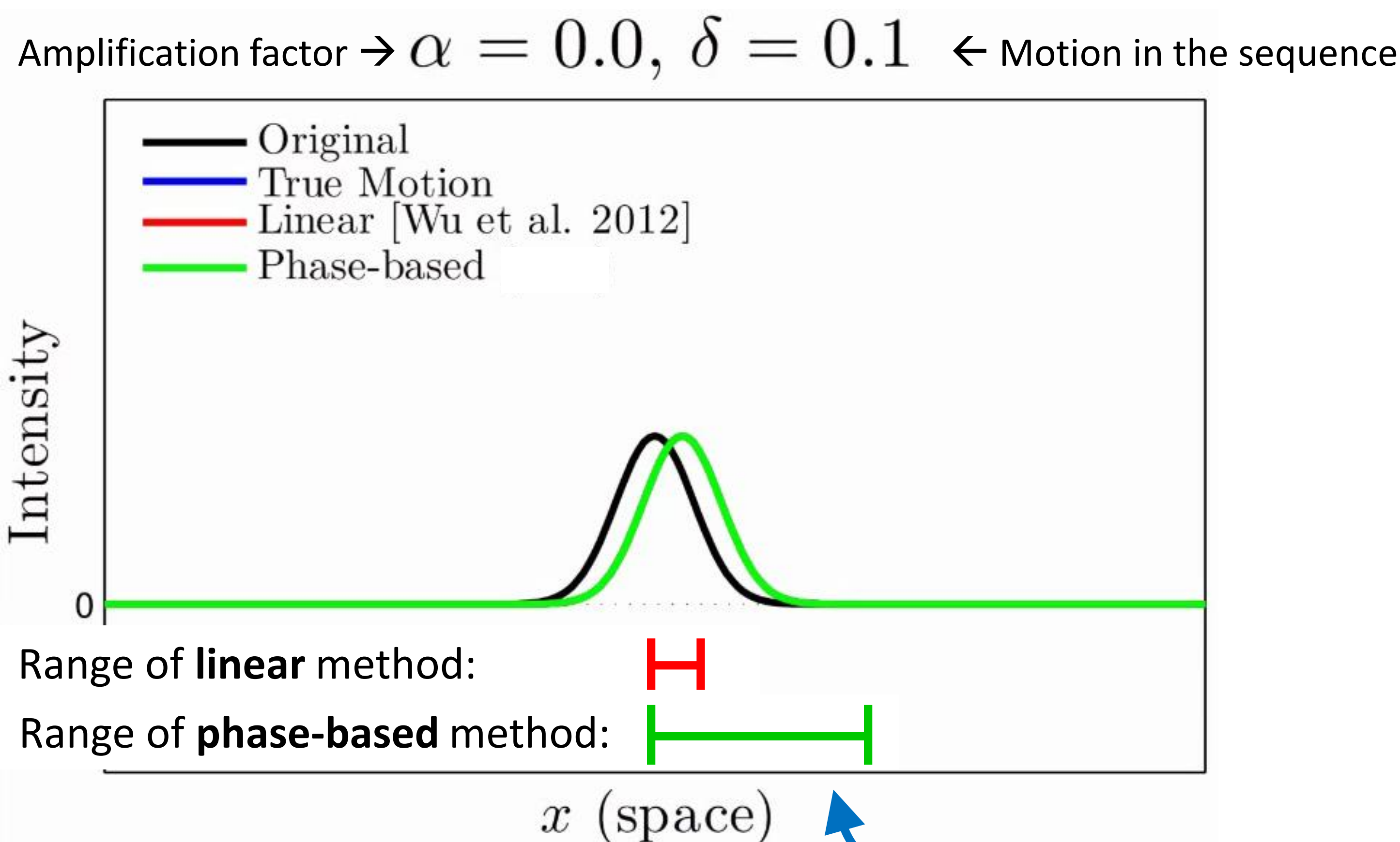
Phase-based Pipeline (SIGGRAPH'13)



Complex steerable pyramid
[Simoncelli, Freeman, Adelson, Heeger 1992]
[Portilla and Simoncelli 2000]

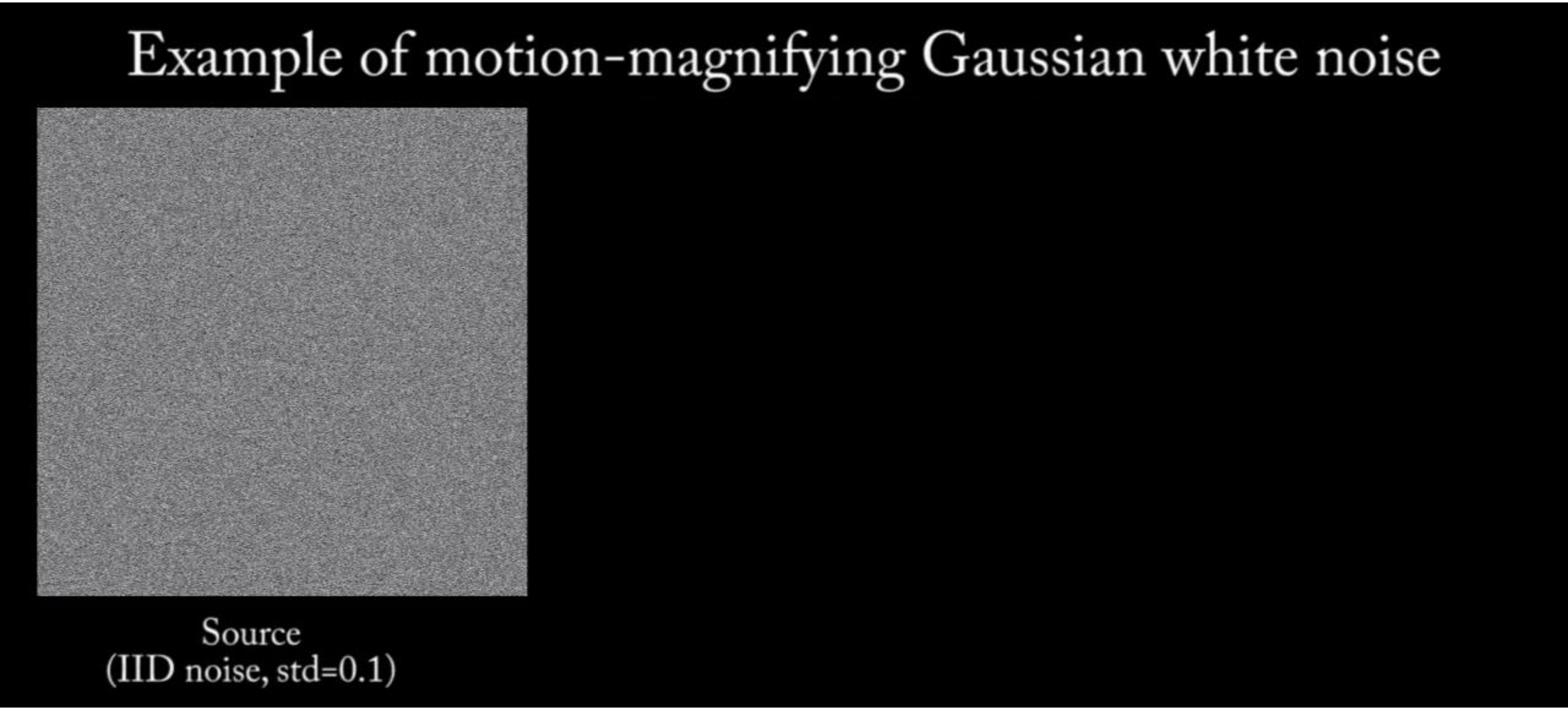
Temporal filtering on **phases**

Improvement #1: More Amplification



4 times the amplification!
(derivation in the thesis)

Improvement #2: Better Noise Performance



Noise **amplified**

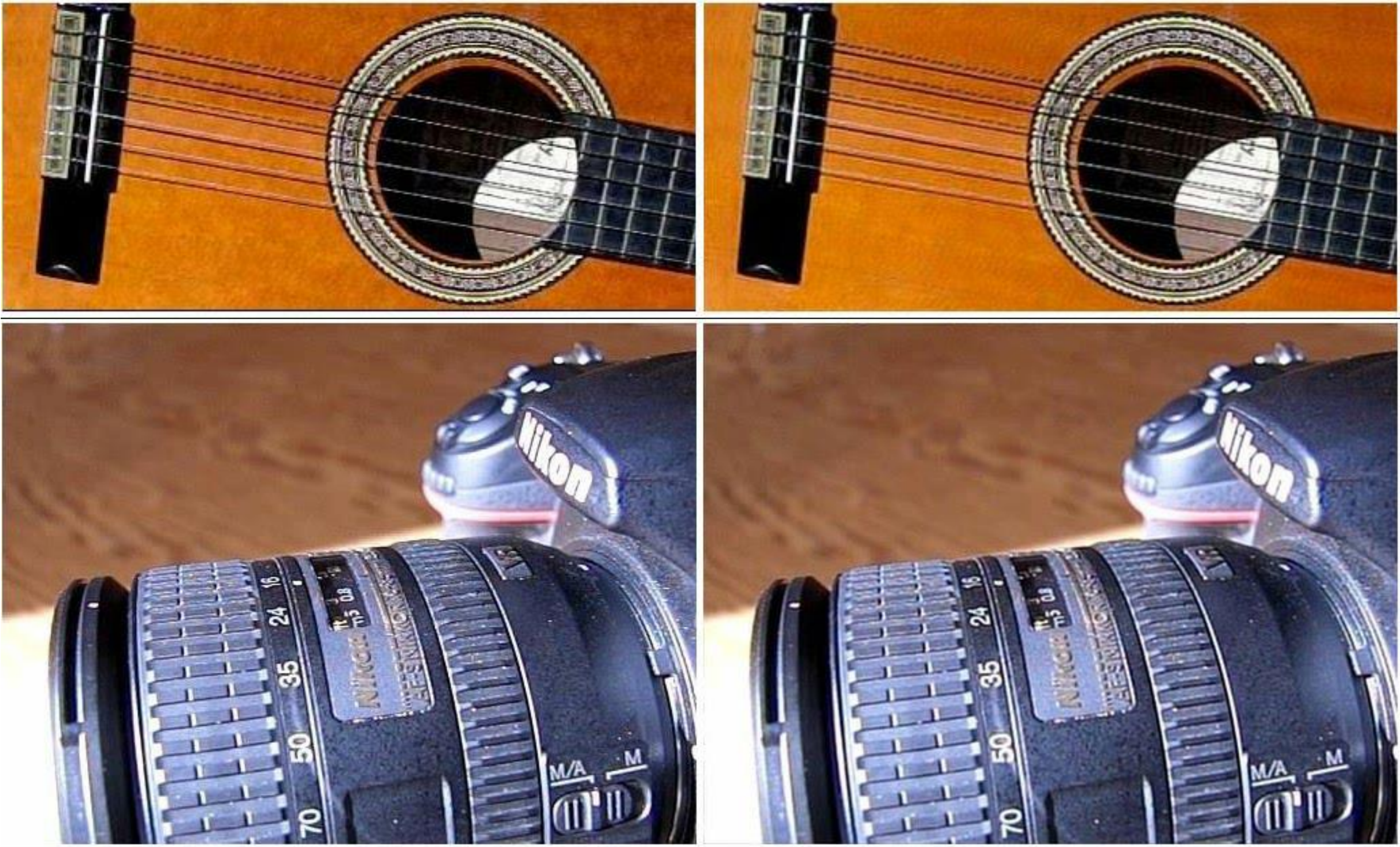
Noise **translated**

Results: Phase-based vs. Linear

Linear (SIGGRAPH'12)
Motions amplified x10

Phase-based (SIGGRAPH'13)
Motions amplified x10

Results: Phase-based vs. Linear



Linear (SIGGRAPH'12) Phase-based (SIGGRAPH'13)

Car Engine

Car Engine

Car Engine

Car Engine

Vibration Modes

Source (20000 FPS)

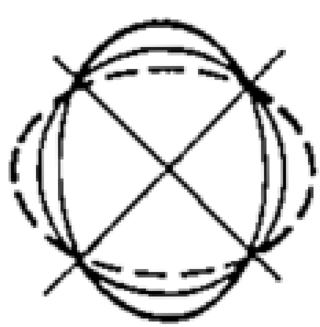
480Hz (x200)

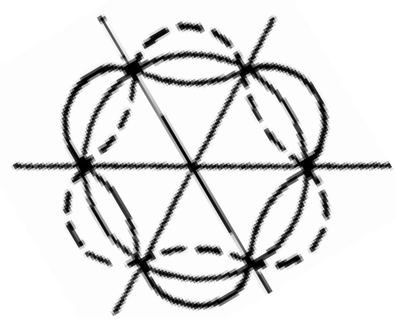
1200Hz (x400)

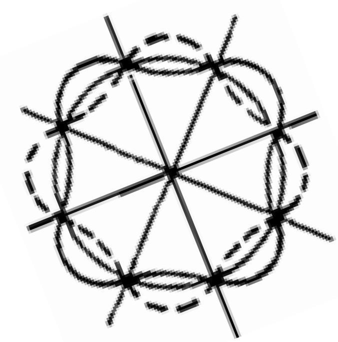
2400Hz (x1200)

Sequence courtesy of Justin Chen

“Piping Vibration Analysis”
[Wachel et al. 1990]

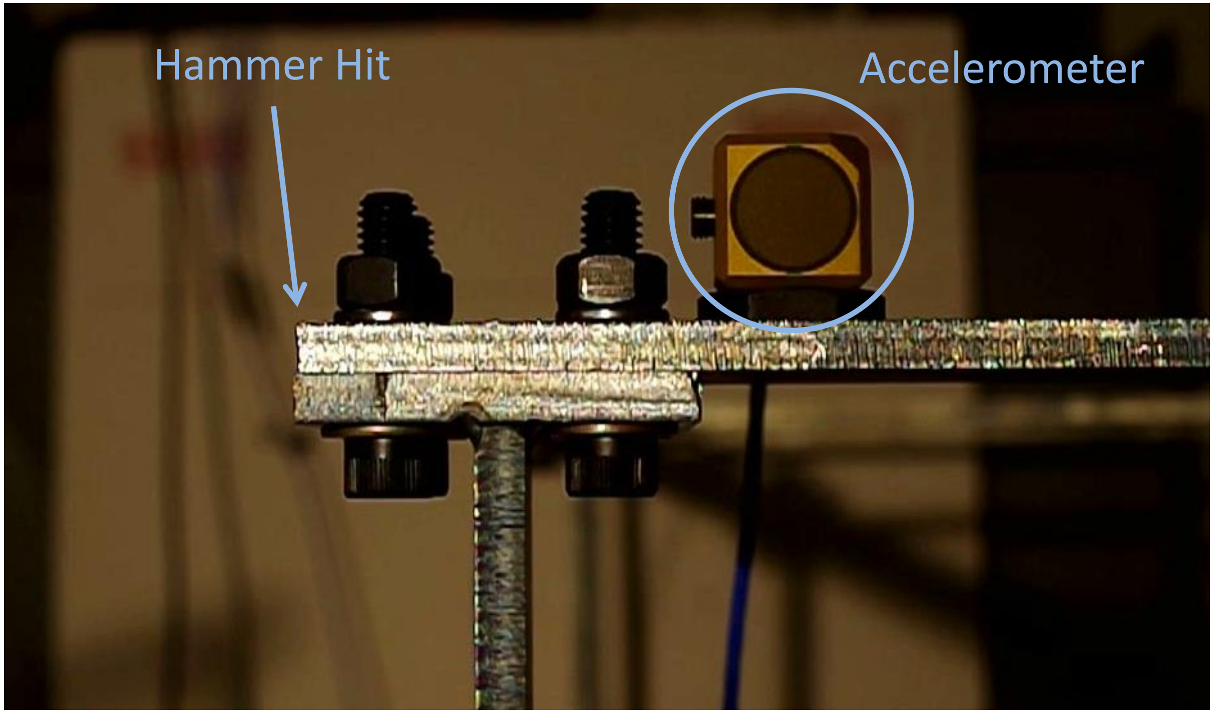




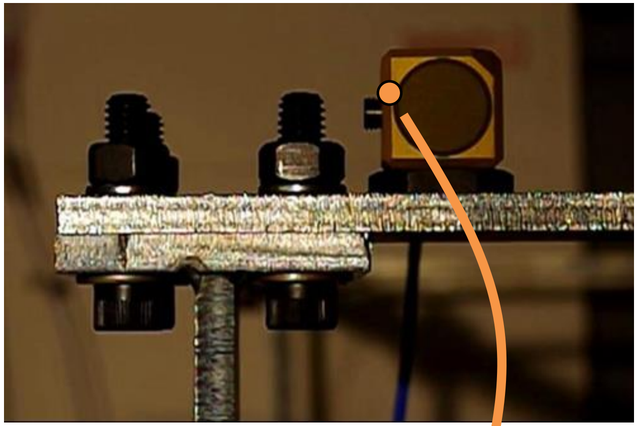


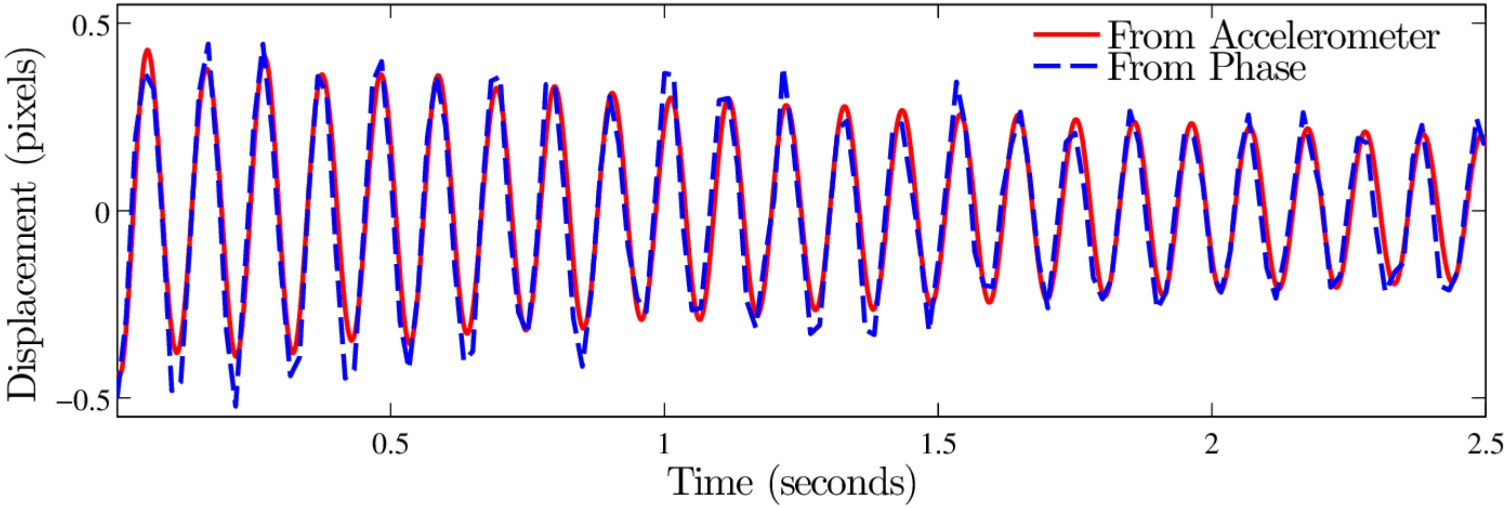
Ground Truth Validation

- Induce motion (with hammer)
- Record true motion with accelerometer

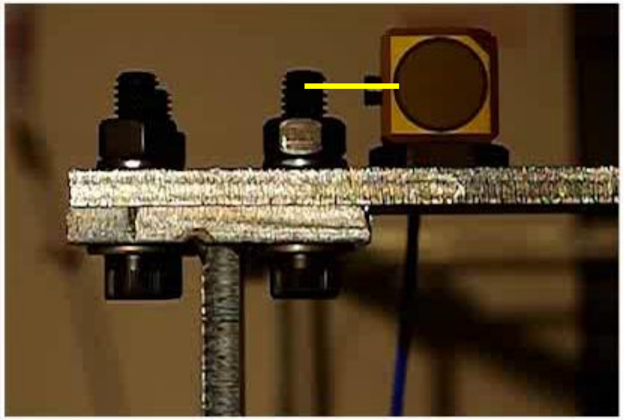


Ground Truth Validation





Qualitative Comparison



Input
(motion of 0.1 px)

Revealing Invisible Changes in the World

- NSF International Science and Engineering Visualization Challenge (SciVis), 2012
- Science Vol. 339 No. 6119 Feb 1 2013

Talk Outline

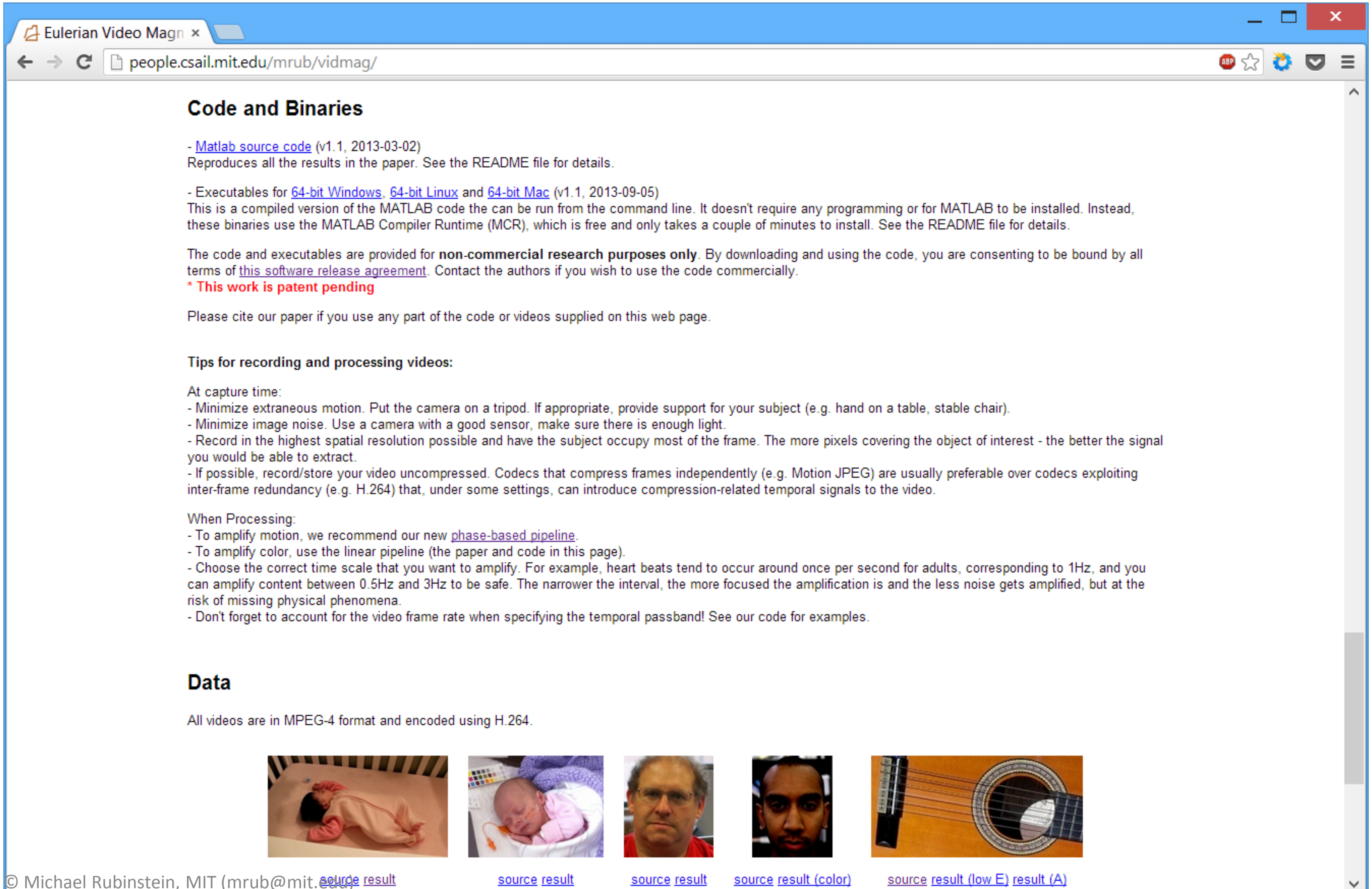
- Removing distracting variations
  - Motion Denoising

- Magnifying imperceptible variations
  - Eulerian Video Magnification
  - Phase-based Video Motion Processing

- Ongoing research and future work

Code Available

- Matlab code + executables http://people.csail.mit.edu/mrub/vidmag/



Eulerian Video Magn ×

people.csail.mit.edu/mrub/vidmag/

Code and Binaries

- Matlab source code (v1.1, 2013-03-02)
Reproduces all the results in the paper. See the README file for details.

- Executables for 64-bit Windows, 64-bit Linux and 64-bit Mac (v1.1, 2013-09-05)
This is a compiled version of the MATLAB code the can be run from the command line. It doesn't require any programming or for MATLAB to be installed. Instead, these binaries use the MATLAB Compiler Runtime (MCR), which is free and only takes a couple of minutes to install. See the README file for details.

The code and executables are provided for **non-commercial research purposes only**. By downloading and using the code, you are consenting to be bound by all terms of this software release agreement. Contact the authors if you wish to use the code commercially.
*** This work is patent pending**

Please cite our paper if you use any part of the code or videos supplied on this web page.

Tips for recording and processing videos:

At capture time:
- Minimize extraneous motion. Put the camera on a tripod. If appropriate, provide support for your subject (e.g. hand on a table, stable chair).
- Minimize image noise. Use a camera with a good sensor, make sure there is enough light.
- Record in the highest spatial resolution possible and have the subject occupy most of the frame. The more pixels covering the object of interest - the better the signal you would be able to extract.
- If possible, record/store your video uncompressed. Codecs that compress frames independently (e.g. Motion JPEG) are usually preferable over codecs exploiting inter-frame redundancy (e.g. H.264) that, under some settings, can introduce compression-related temporal signals to the video.

When Processing:
- To amplify motion, we recommend our new phase-based pipeline.
- To amplify color, use the linear pipeline (the paper and code in this page).
- Choose the correct time scale that you want to amplify. For example, heart beats tend to occur around once per second for adults, corresponding to 1Hz, and you can amplify content between 0.5Hz and 3Hz to be safe. The narrower the interval, the more focused the amplification is and the less noise gets amplified, but at the risk of missing physical phenomena.
- Don't forget to account for the video frame rate when specifying the temporal passband! See our code for examples.

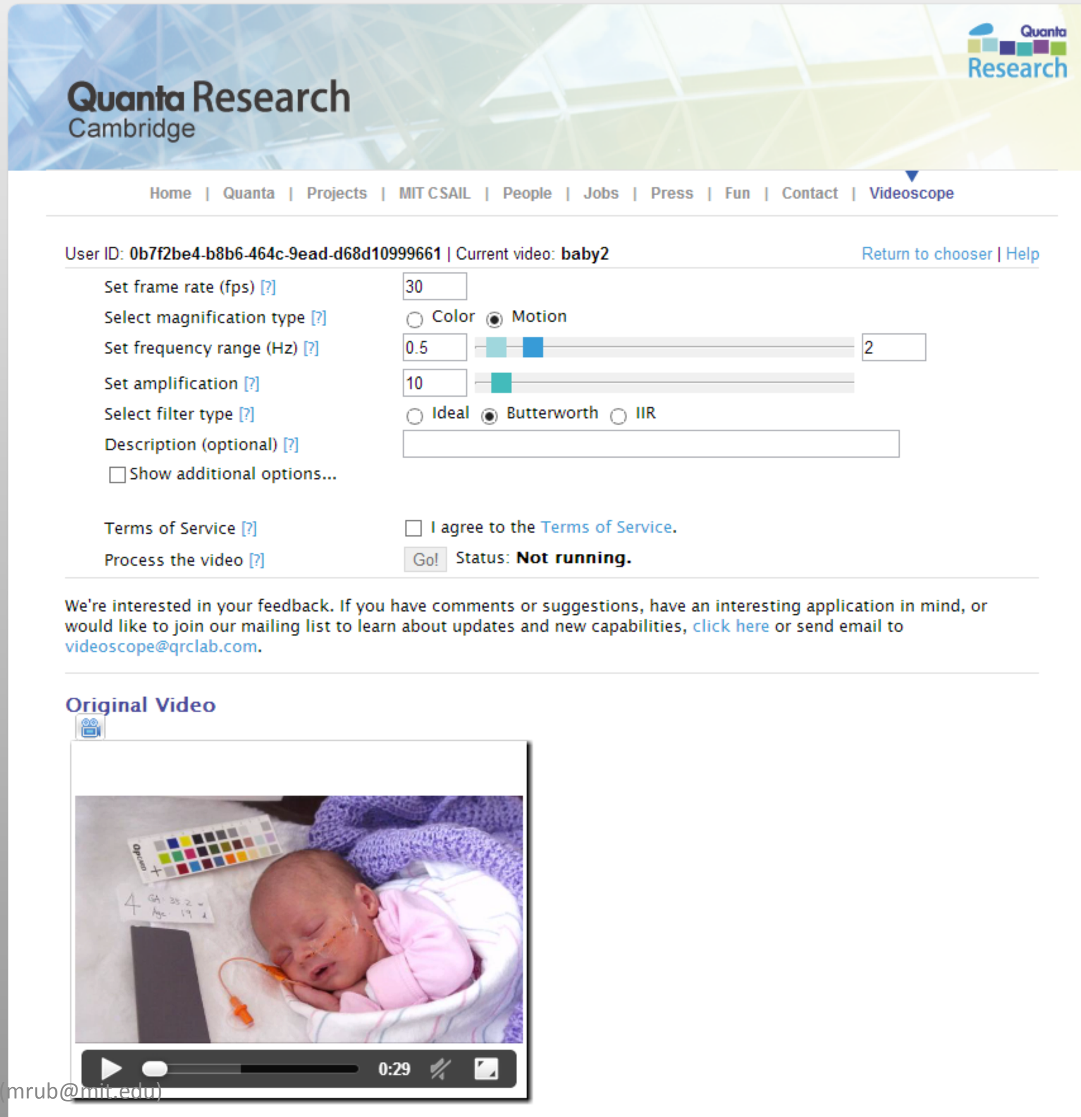
Data

All videos are in MPEG-4 format and encoded using H.264.

source result | source result | source result | source result (color) | source result (low E) result (A)

“VideoScope” by Quanta Research Cambridge

http://videoscope.qrclab.com/

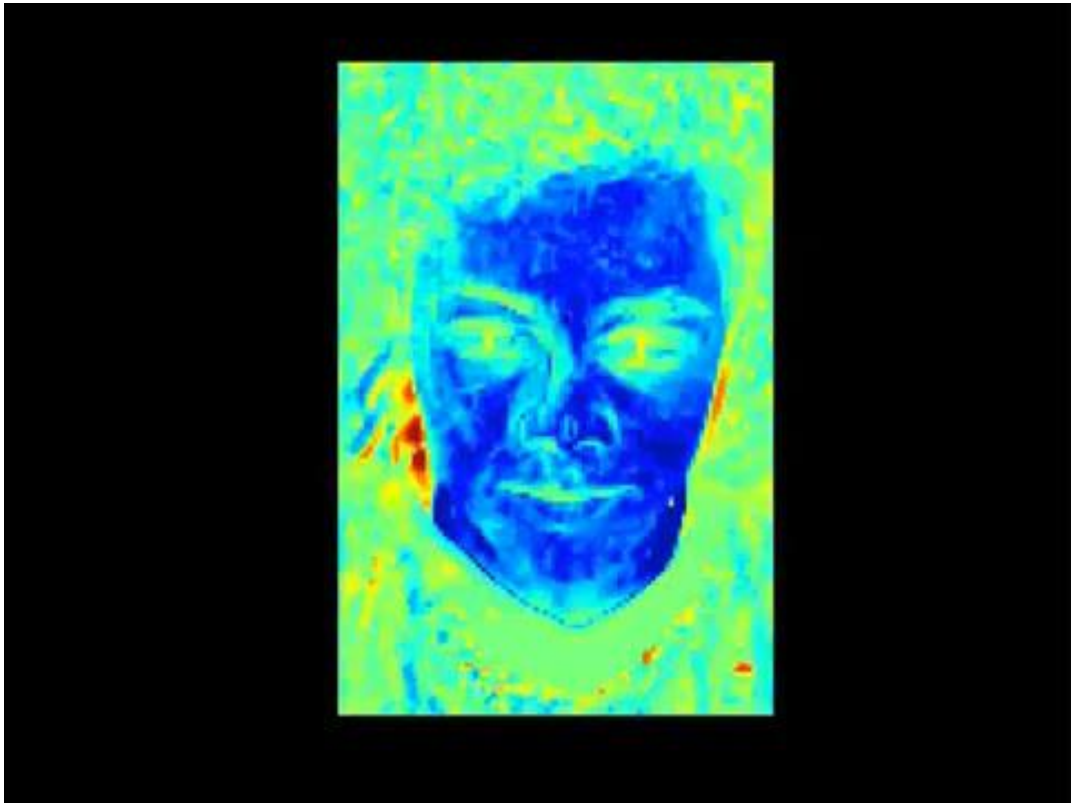


EVM in the Wild: Pregnancy

Original Processed

“Tomez85” https://www.youtube.com/watch?v=J1wvFmWv7zY

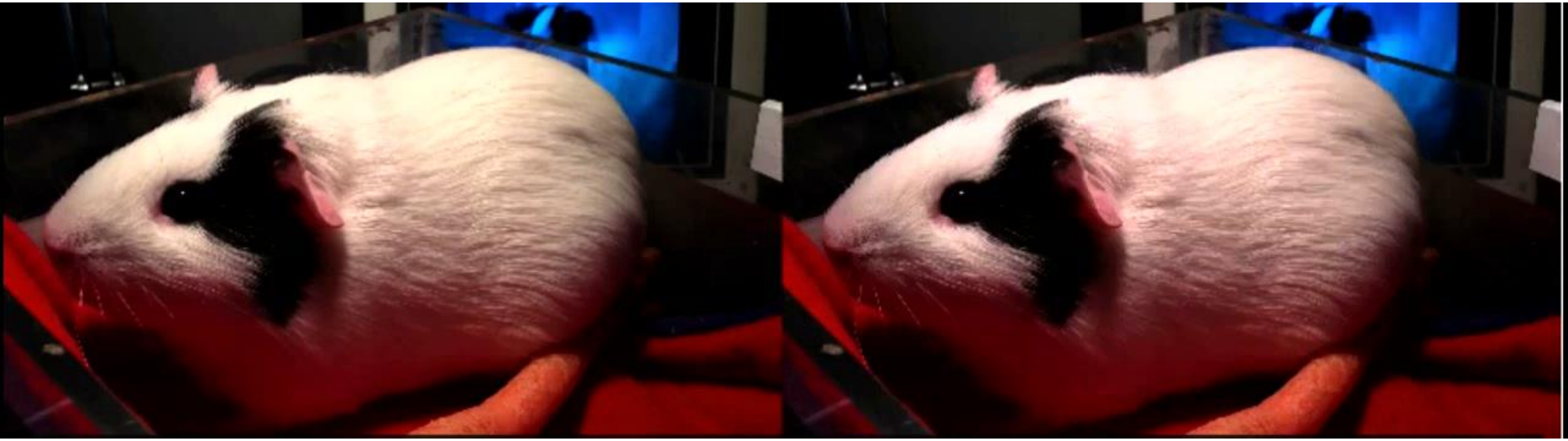
EVM in the Wild: Blood flow Visualization



Red = high blood volume
Blue = low blood volume

Institute for Biomedical Engineering, Dresden Germany
https://www.youtube.com/watch?v=Nb18CRVmXGY

EVM in the Wild: Guinea Pig



Source Motion-magnified

“SuperCreaturefan”: “Guinea pig Tiffany is the first rodent on Earth to undergo Eulerian Video Magnification.”
http://www.youtube.com/watch?v=uXOSJvNwtIk

EVM in the Wild: “Eulerian Profiles”

Source | Motion-amplified | Source | Motion-amplified

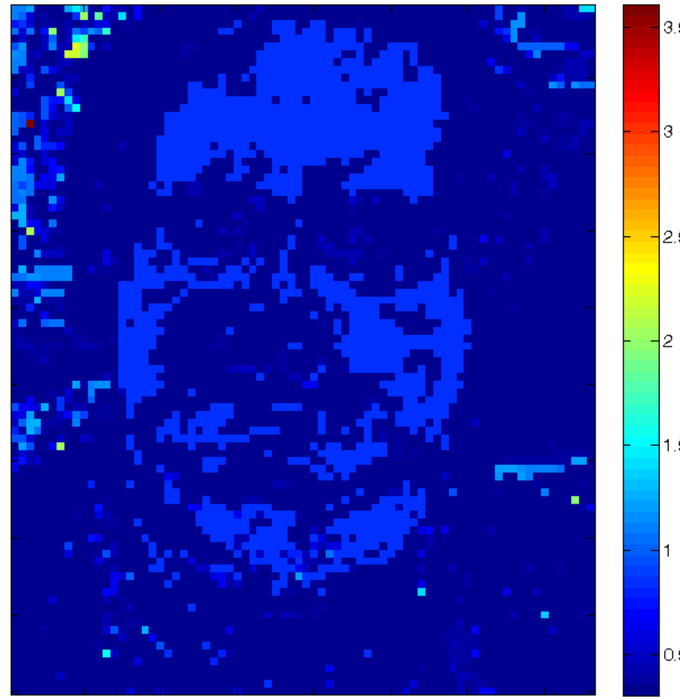
By Erin Knutson (Graphic Design student at Yale)

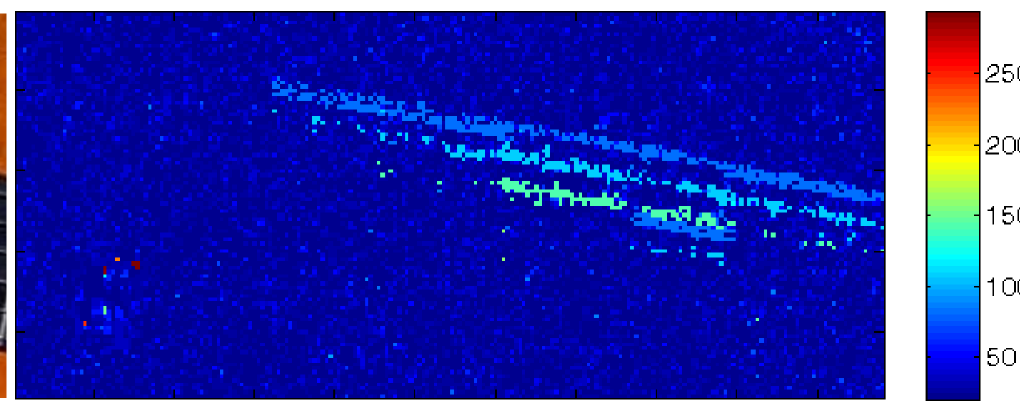
People Interested in…

- **Health care**
  - Contactless monitoring
  - Blood vessel identification
  - Tissue perfusion in plastic surgery
  - …
- **Scientific analysis**
  - Changes in the earth's surface from satellite imagery
  - Seismic data
  - …
- **Engineering**
  - Structural integrity of bridges, buildings
  - …



Identifying Temporal Signals Automatically





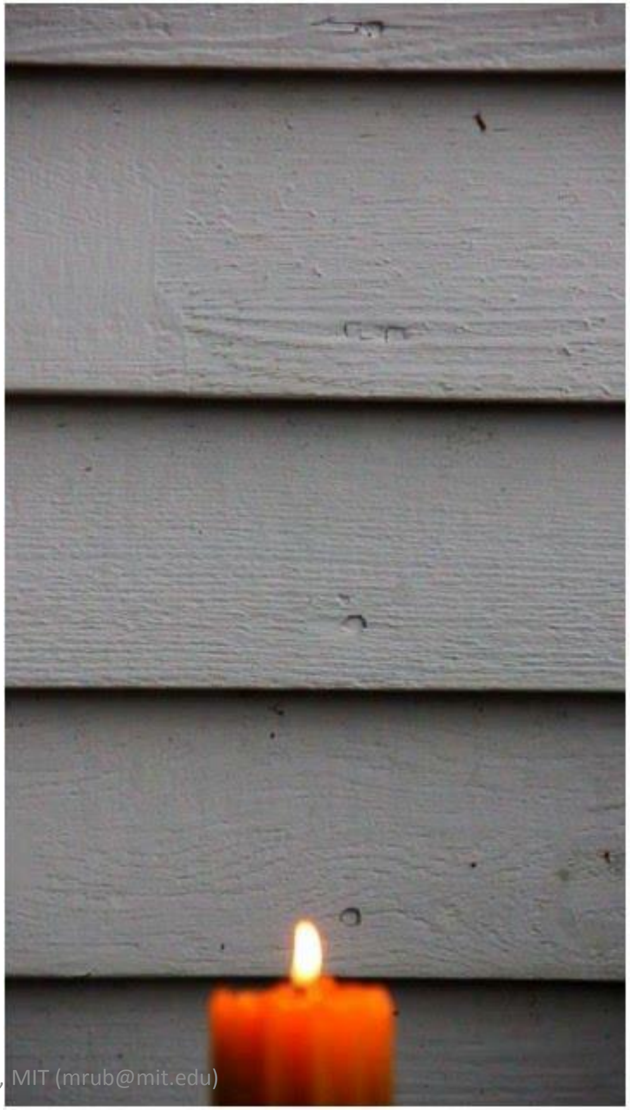
Source frame

Dominant frequency

Seeing and Measuring Refractive Flow (hot air, gas)

****Unpublished****

- Small motions due to changes in refraction index (change in density, temp.)



****Unpublished****

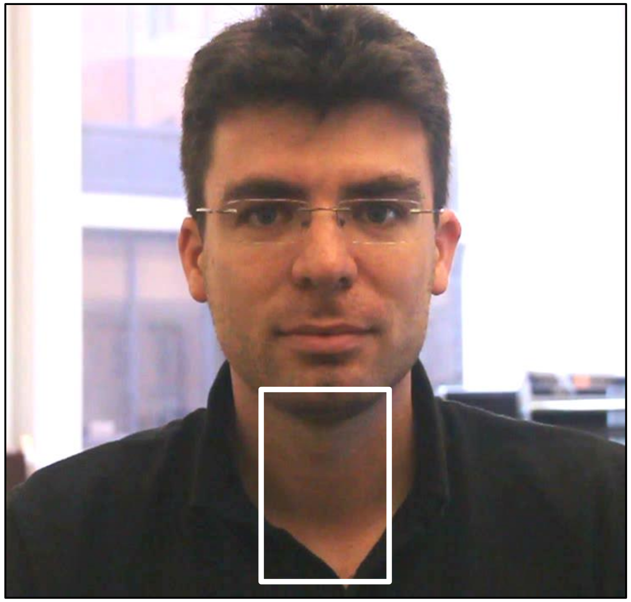
Seeing and Measuring Sound

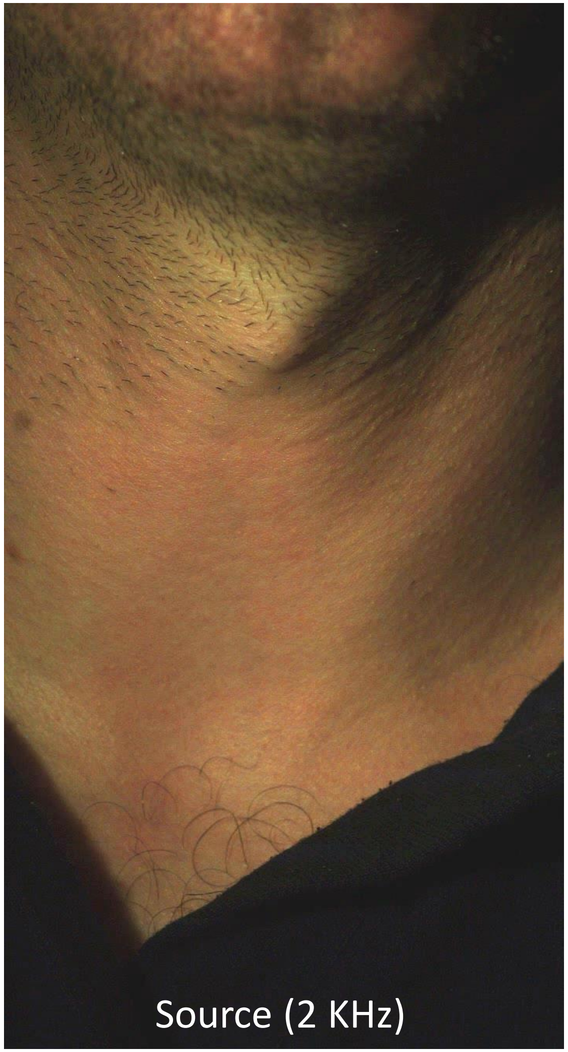
- Sound is fluctuations in air pressure traveling through space
- These pressure waves hit objects and make them vibrate
  - This is how we hear; this is how we record sound

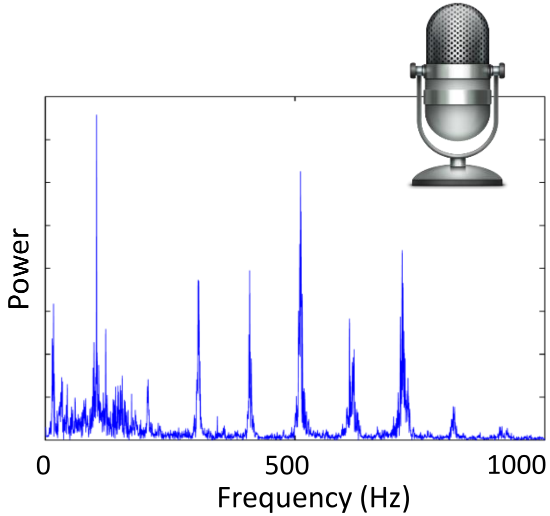
"water sound waves"
xsgianni, http://www.youtube.com/watch?v=xPW3gihYnZE

Unpublished

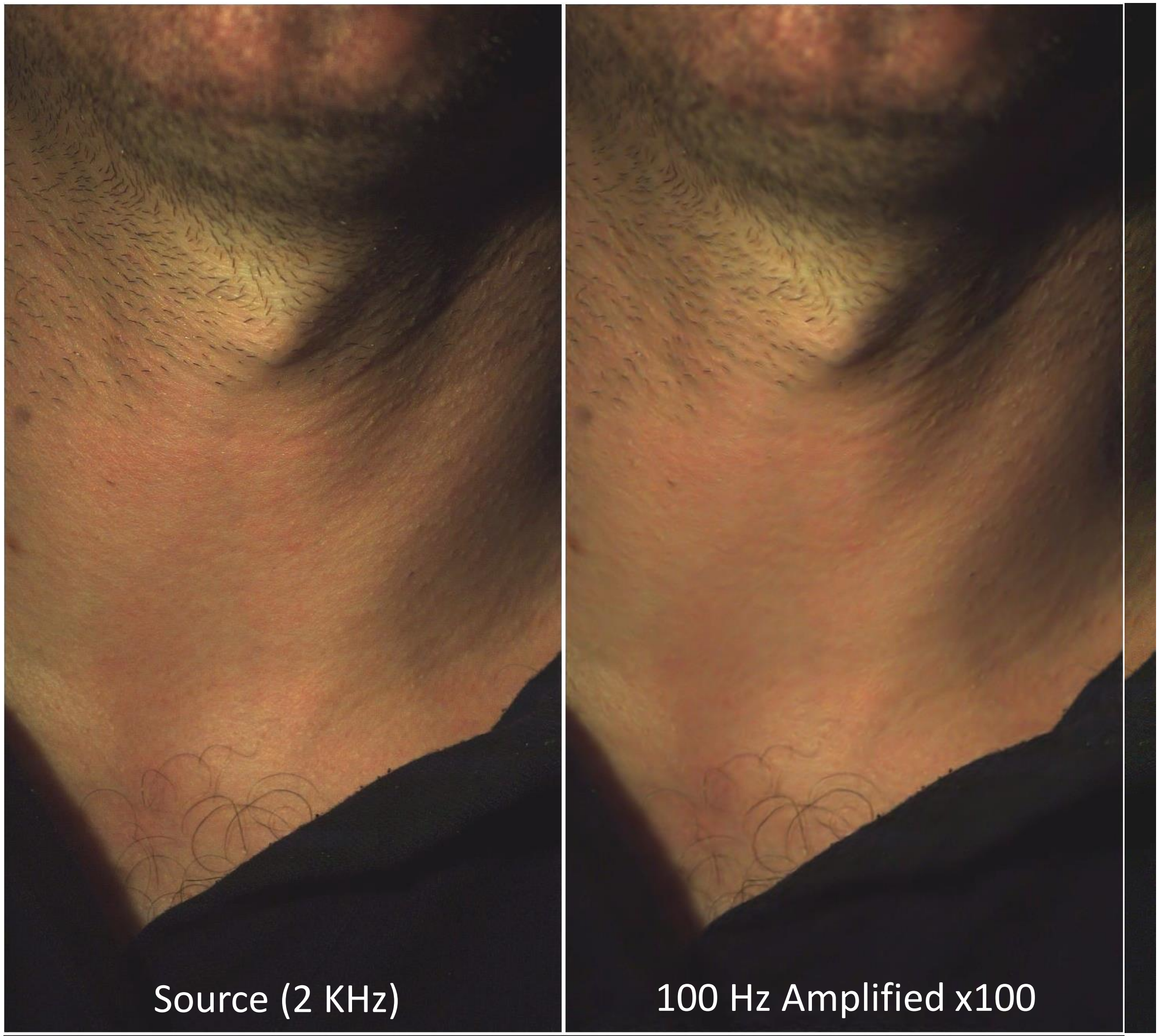
Neck Skin Vibrations

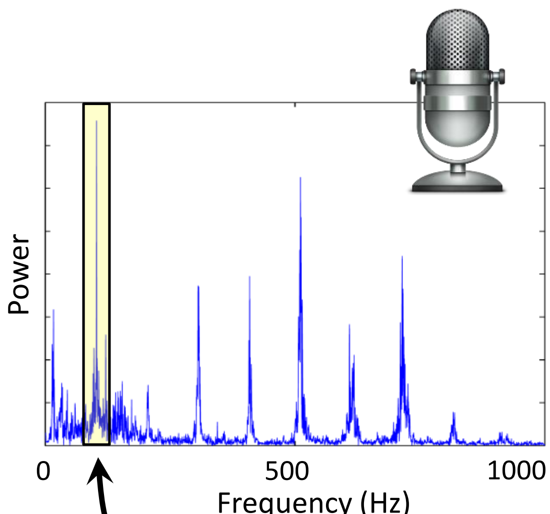






Neck Skin Vibrations





Fundamental frequency: ~100Hz

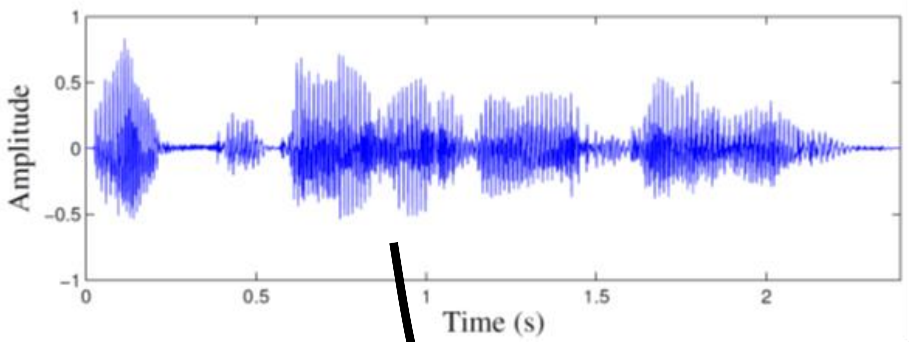
Unpublished

Can We Recover Sound From Video?

****Unpublished****

Recovering Sound from Video

- Assuming scene is static, motions should be well correlated with sound pressure waves

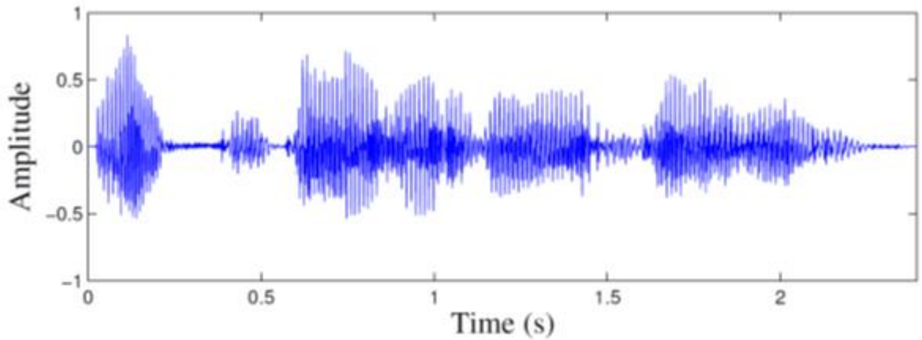


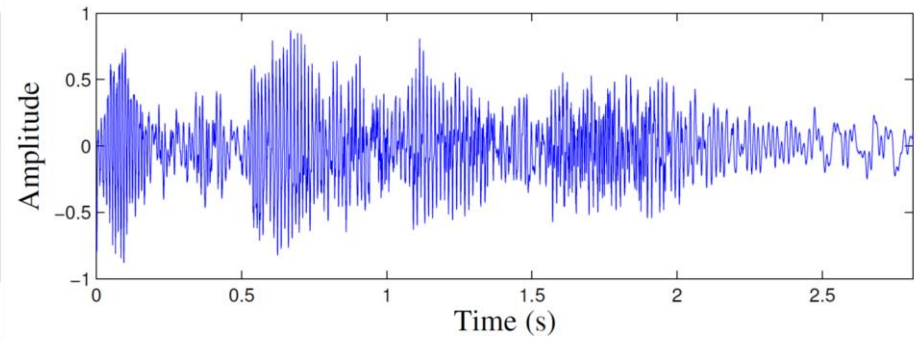
Low-passed 0-2KHz

Recorded video (4 KHz)

Recovering Sound from Video

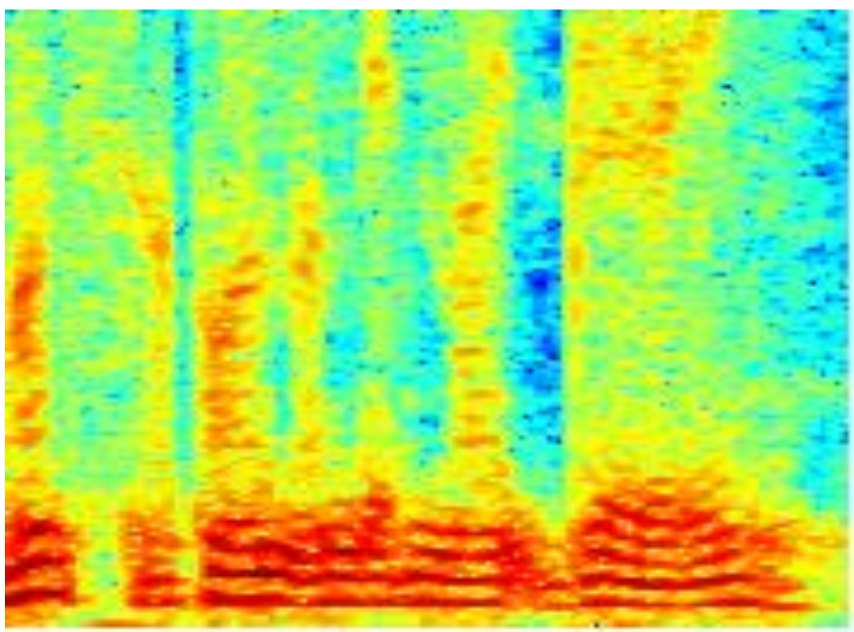
****Unpublished****

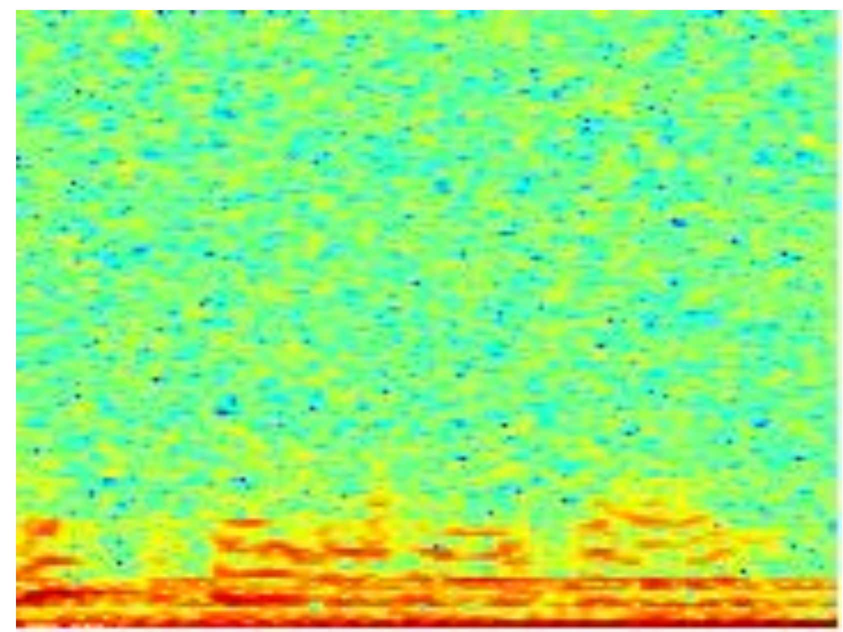




2Khz

Frequency





Time

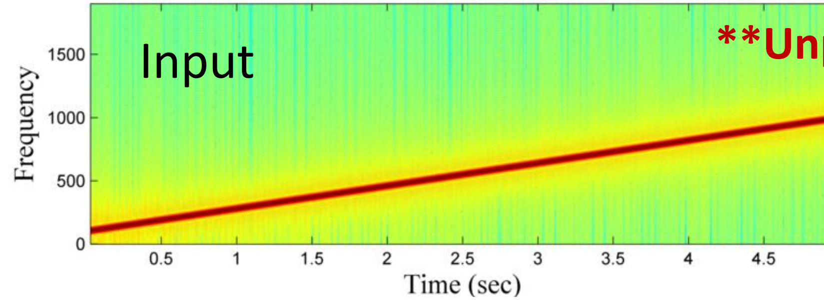
Time

Source

Reconstructed

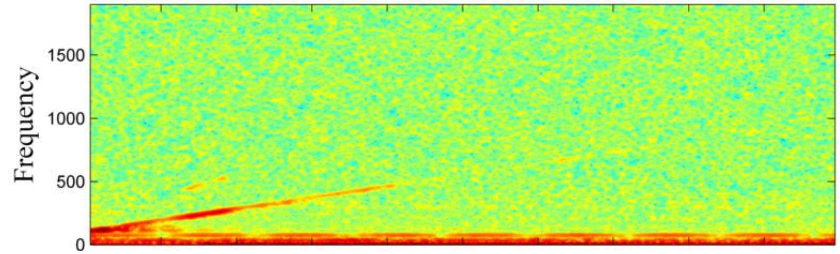
Natural Microphones

Unpublished

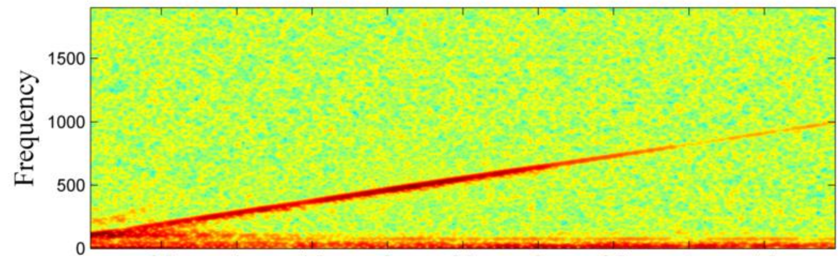


Reconstruction from:

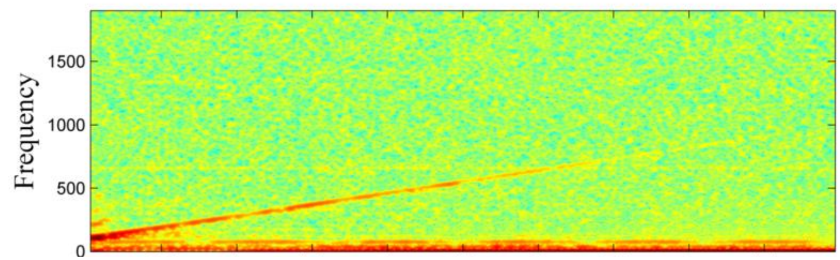
Water



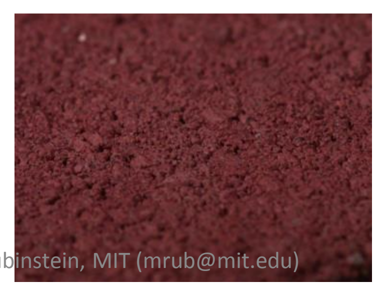
Latex membrane

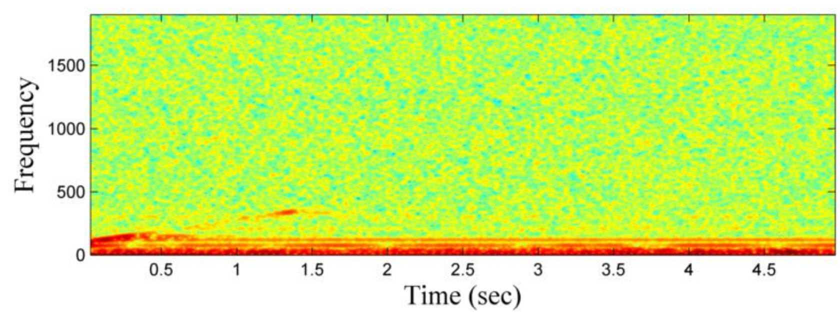


Cardboard



Brick





Conclusions

- We decompose temporal signals in videos into different components and re-render the video to analyze and visualize them separately

- **Removing distracting temporal variation**
  - **Motion denoising** – decomposition into long-term and short-term changes
    - No explicit motion analysis

- **Amplifying imperceptible temporal variation**
  - **Eulerian approaches** for representing, analyzing and visualizing small-amplitude temporal signals
    - No explicit motion analysis

  - The world is full of small, informative motions and changes we cannot normally see, and we can reveal them using regular video

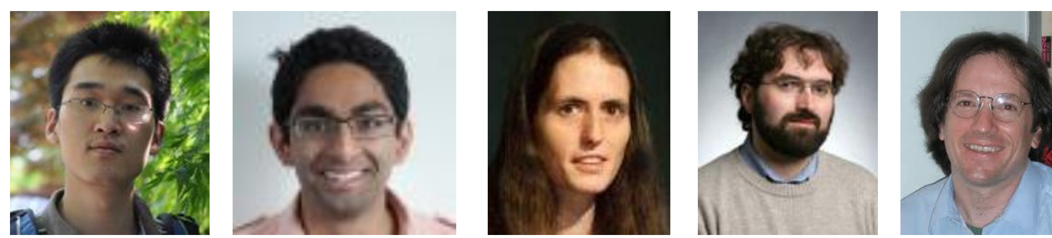
Acknowledgements

Motion Denoising
(CVPR 2011)

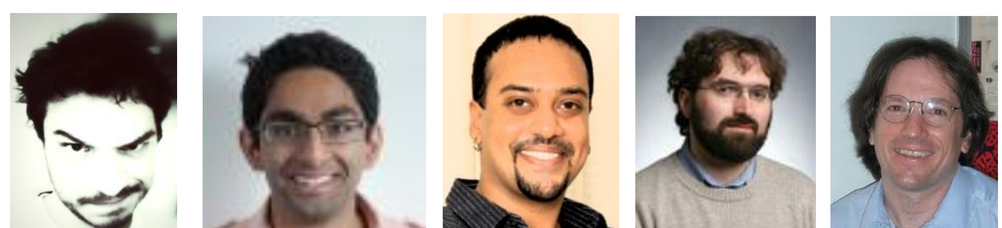
Eulerian Video Magnification
(SIGGRAPH 2012)

Phase-based Motion Processing
(SIGGRAPH 2013)

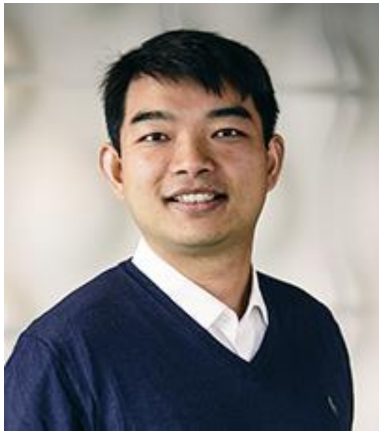
Refractive Flow



Visual Microphone

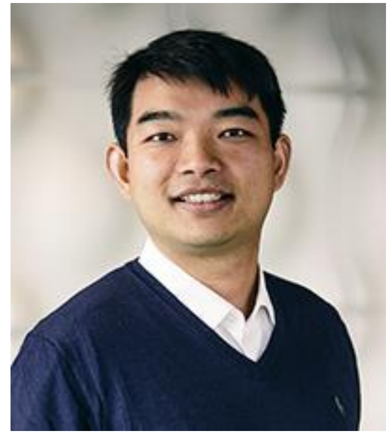


Acknowledgements

Acknowledgements

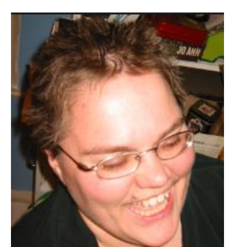










Acknowledgements

Acknowledgements

Acknowledgments

NSF CGV-1111415
"Images Through Time"

Thank you!

Michael Rubinstein

MIT CSAIL