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]

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

[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)

© Michael Rubinstein, MIT (mrub@mit.edu)
Distracting Temporal Variation

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

Remove changes to make long-term variation more visible!

© Extreme Ice Survey

© Michael Rubinstein, MIT (mrub@mit.edu)
Imperceptible Temporal Variation

- Changes are too small

*Magnify the variation to make it visible!*

[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

© Michael Rubinstein, MIT (mrub@mit.edu)
For Personal Use Too

9 months

TIME of my LIFE

16 years

1 year

10 years
Stylized Jerkiness

July 2008

© Michael Rubinstein, MIT (mrub@mit.edu)
Timescale Separation

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

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

© Michael Rubinstein, MIT (mrub@mit.edu)
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 inconsistencies
    • Motion discontinuities
**Basic Idea**

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

- **Approach:** 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, $\mathbf{w}$:

$$E(\mathbf{w}) = \sum_p |I(p + \mathbf{w}(p)) - I(p)| + \alpha \sum_{p,r \in N_t(p)} \|I(p + \mathbf{w}(p)) - I(r + \mathbf{w}(r))\|^2 + \gamma \sum_{p,q \in N(p)} \lambda_{pq}|\mathbf{w}(p) - \mathbf{w}(q)|$$

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

Fidelity (to input video)
Temporal coherence (of the output video)
Regularization (of the warp)
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 ($\alpha$-expansion)

Belief Propagation

© Michael Rubinstein, MIT (mrub@mit.edu)

[Boykov et al. 2002]
Comparison with Pixel-wise Temporal Filtering

Source

Mean

Median

Motion-denoised

Source

Sliding mean

Sliding median

Motion denoising

© Michael Rubinstein, MIT (mrub@mit.edu)
Results
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

© Michael Rubinstein, MIT (mrub@mit.edu)
Timescale Decomposition
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

© Michael Rubinstein, MIT (mrub@mit.edu)
Imperceptible Changes in the World

Blood flow

Camera motion due to shutter and mirror

Breathing

Micro-expressions

© Michael Rubinstein, MIT (mrub@mit.edu)
Lagrangian and Eulerian Perspectives (Fluid Dynamics)

• Specifications of physical measurements through space and time:

  Lagrangian
  Track particles

  Eulerian
  Measure changes within fixed voxels in space

© Michael Rubinstein, MIT (mrub@mit.edu)
Basic Idea

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

© Michael Rubinstein, MIT (mrub@mit.edu)
Subtle Color Variations

1. Average spatially to overcome sensor and quantization noise

Input frame

Spatially averaged luminance trace
Amplifying Subtle Color Variations

2. Filter temporally to extract the signal of interest

\[
\text{Input frame} \quad \implies \quad \text{Spatially averaged luminance trace} \quad \ast \quad \text{Temporal filter} \quad \implies \quad \text{Temporally bandpassed trace}
\]
Color Amplification Results

Source

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

© Michael Rubinstein, MIT (mrub@mit.edu)
Bruce Wayne’s Pulse

Christian Bale, Batman Begins (2005)

Courtesy of Warner Bros. Pictures

© Michael Rubinstein, MIT (mrub@mit.edu)
Heart Rate Extraction

Peak detection

Temporally bandpassed trace (one pixel)

Pulse locations
Extracting Heart Rate

With Dr. Donna Brezinski and the Winchester Hospital staff

© Michael Rubinstein, MIT (mrub@mit.edu)
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

\[
\frac{\partial x}{\partial t}
\]

© Michael Rubinstein, MIT (mrub@mit.edu)
Differential Brightness Constancy

- Measure temporal variation $\frac{\partial I}{\partial t}$ (at each pixel)

\[
\frac{\partial I}{\partial t} \approx \frac{\partial I}{\partial x} \frac{\partial x}{\partial t}
\]

Lucas-Kanade, 1981
Horn-Schunck, 1981
Eulerian Motion Magnification

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

$$\alpha \frac{\partial I}{\partial t}$$

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

Intensity

0

Space ($x$)

© Michael Rubinstein, MIT (mrub@mit.edu)
Relating Temporal and Spatial Changes

Signal at time $t$
Signal at time $t + \Delta t$
Motion-magnified

Courtesy of Lili Sun

© Michael Rubinstein, MIT (mrub@mit.edu)
Synthetic 2D Example

Source
Method Pipeline

Laplacian pyramid
[Burt and Adelson 1983]
Motion Magnification Results

Source

© Michael Rubinstein, MIT (mrub@mit.edu)
Selective Motion Magnification

Source
(600 fps)

72-92 Hz Amplified

100-120 Hz Amplified

Low E (82.4 Hz)

A (110 Hz)

© Michael Rubinstein, MIT (mrub@mit.edu)
Related Work: Motion Magnification [Liu 2005]

Source

Motion-magnified

Liu et al. *Motion Magnification*, SIGGRAPH 2005

© Michael Rubinstein, MIT (mrub@mit.edu)
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

Liu et al. *Motion Magnification*, SIGGRAPH 2005

© Michael Rubinstein, MIT (mrub@mit.edu)
Lagrangian vs. Eulerian

- See my thesis for more details!

© Michael Rubinstein, MIT (mrub@mit.edu)
Limitations

Amplified noise

Source

Motion-magnified

Intensity clipping

Source (300 fps)

© Michael Rubinstein, MIT (mrub@mit.edu)
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

© Michael Rubinstein, MIT (mrub@mit.edu)
Fourier Decomposition

\[ I(x) \quad \xrightarrow{\text{FFT}} \quad \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega x} \]

\[ A_1 \times \text{Intensity} \quad + \quad A_2 \times \text{Intensity} \quad + \cdots \]

© Michael Rubinstein, MIT (mrub@mit.edu)
Fourier Shift Theorem

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

Phase shift $\iff$ Translation

\( A_1 \times \) Intensity $\overset{\text{Space (x)}}{\iff} + A_2 \times$ Intensity $\overset{\text{Space (x)}}{\iff} + \cdots$

Phase \((e^{-i \delta})\)  

© Michael Rubinstein, MIT (mrub@mit.edu)
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

© Michael Rubinstein, MIT (mrub@mit.edu)
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

Filter Bank

<table>
<thead>
<tr>
<th>Scale 1</th>
<th></th>
<th>Scale 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orientation 1</td>
<td>Real</td>
<td>Imag</td>
</tr>
<tr>
<td>Orientation 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Real

Imag

Intensity

Space (x)

Complex Sinusoid (Global)

Window

© Michael Rubinstein, MIT (mrub@mit.edu)
Local Phase

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

Local phase shift $\leftrightarrow$ Local translation

© Michael Rubinstein, MIT (mrub@mit.edu)
Linear Pipeline (SIGGRAPH’12)

Laplacian pyramid [Burt and Adelson 1983]

Temporal filtering on **intensities**

© Michael Rubinstein, MIT (mrub@mit.edu)
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

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! (derivation in the thesis)
Improvement #2: Better Noise Performance

Example of motion-magnifying Gaussian white noise

Source
(IID noise, std=0.1)

Noise **amplified**

Noise **translated**
Results: Phase-based vs. Linear

Linear (SIGGRAPH’12)
Motions amplified x10

Phase-based (SIGGRAPH’13)
Motions amplified x10

© Michael Rubinstein, MIT (mrub@mit.edu)
Results: Phase-based vs. Linear
Car Engine

Source
Car Engine

22Hz Magnified

© Michael Rubinstein, MIT (mrub@mit.edu)
Car Engine

Source

© Michael Rubinstein, MIT (mrub@mit.edu)
Car Engine

22Hz Magnified

© Michael Rubinstein, MIT (mrub@mit.edu)
Vibration Modes

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

© Michael Rubinstein, MIT (mrub@mit.edu)
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/]

**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 that 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.

[Images of different videos provided.]

© Michael Rubinstein, MIT (mrub@mit.edu)
“VideoScope” by Quanta Research Cambridge

http://videoscope.qrclab.com/
EVM in the Wild: Pregnancy

Original

Processed

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

© Michael Rubinstein, MIT (mrub@mit.edu)
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

© Michael Rubinstein, MIT (mrub@mit.edu)
EVM in the Wild: Guinea Pig

“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”

By Erin Knutson (Graphic Design student at Yale)

© Michael Rubinstein, MIT (mrub@mit.edu)
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
  – ...

© Michael Rubinstein, MIT (mrub@mit.edu)
Identifying Temporal Signals Automatically

Source frame

Dominant frequency
Seeing and Measuring Refractive Flow (hot air, gas)

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

Source (2 KHz)

© Michael Rubinstein, MIT (mrub@mit.edu)
Neck Skin Vibrations

Source (2 KHz)  100 Hz Amplified x100

Fundamental frequency: ~100Hz

**Unpublished**

© Michael Rubinstein, MIT (mrub@mit.edu)
Can We Recover Sound From Video?

© Michael Rubinstein, MIT (mrub@mit.edu)
Recovering Sound from Video

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

Low-passed 0-2KHz

© Michael Rubinstein, MIT (mrub@mit.edu)
Recovering Sound from Video

Source

Reconstructed

2Khz

Frequency

Time

© Michael Rubinstein, MIT (mrub@mit.edu)
Natural Microphones

Reconstruction from:

Water

Latex membrane

Cardboard

Brick

© Michael Rubinstein, MIT (mrub@mit.edu)
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

© Michael Rubinstein, MIT (mrub@mit.edu)
Acknowledgements

Motion Denoising
(CVPR 2011)

Eulerian Video Magnification
(SIGGRAPH 2012)

Phase-based Motion Processing
(SIGGRAPH 2013)

Refractive Flow

Visual Microphone

© Michael Rubinstein, MIT (mrub@mit.edu)
Acknowledgements
Acknowledgements
Acknowledgements
Acknowledgements
Acknowledgments

NSF CGV-1111415
“Images Through Time”
Thank you!

Michael Rubinstein
MIT CSAIL