MIT
COMPUTER
VISION

Motion Denoising
with Application to Time-lapse Photography

Michael Rubinstein

MIT CSAIL

Ce Liu
Microsoft Research NE

Peter Sand

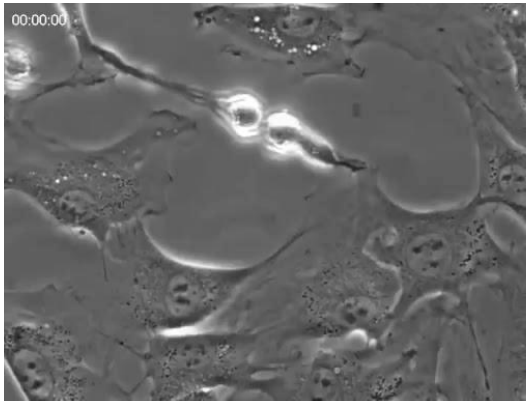
Fredo Durand
MIT

Bill Freeman
MIT

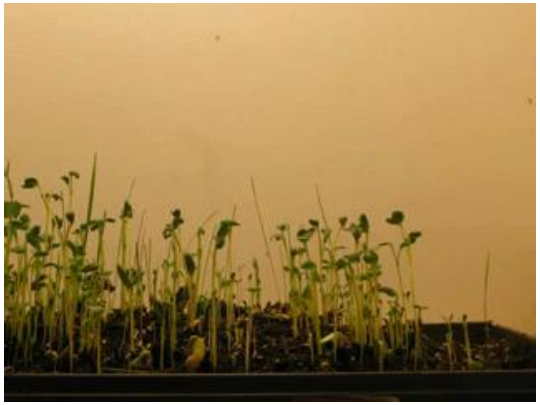
Time-lapse Videos

Construction

Natural phenomena



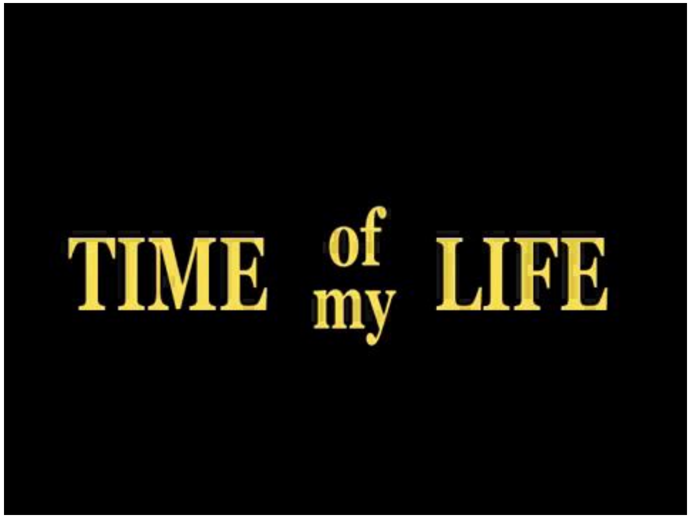
Medical



Biological/Botanical

For Personal Use Too!

9 months



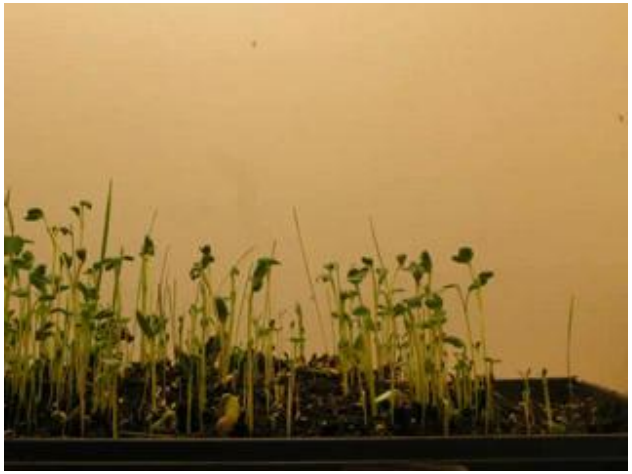
16 years

http://www.danhanna.com/aging_project/p.html

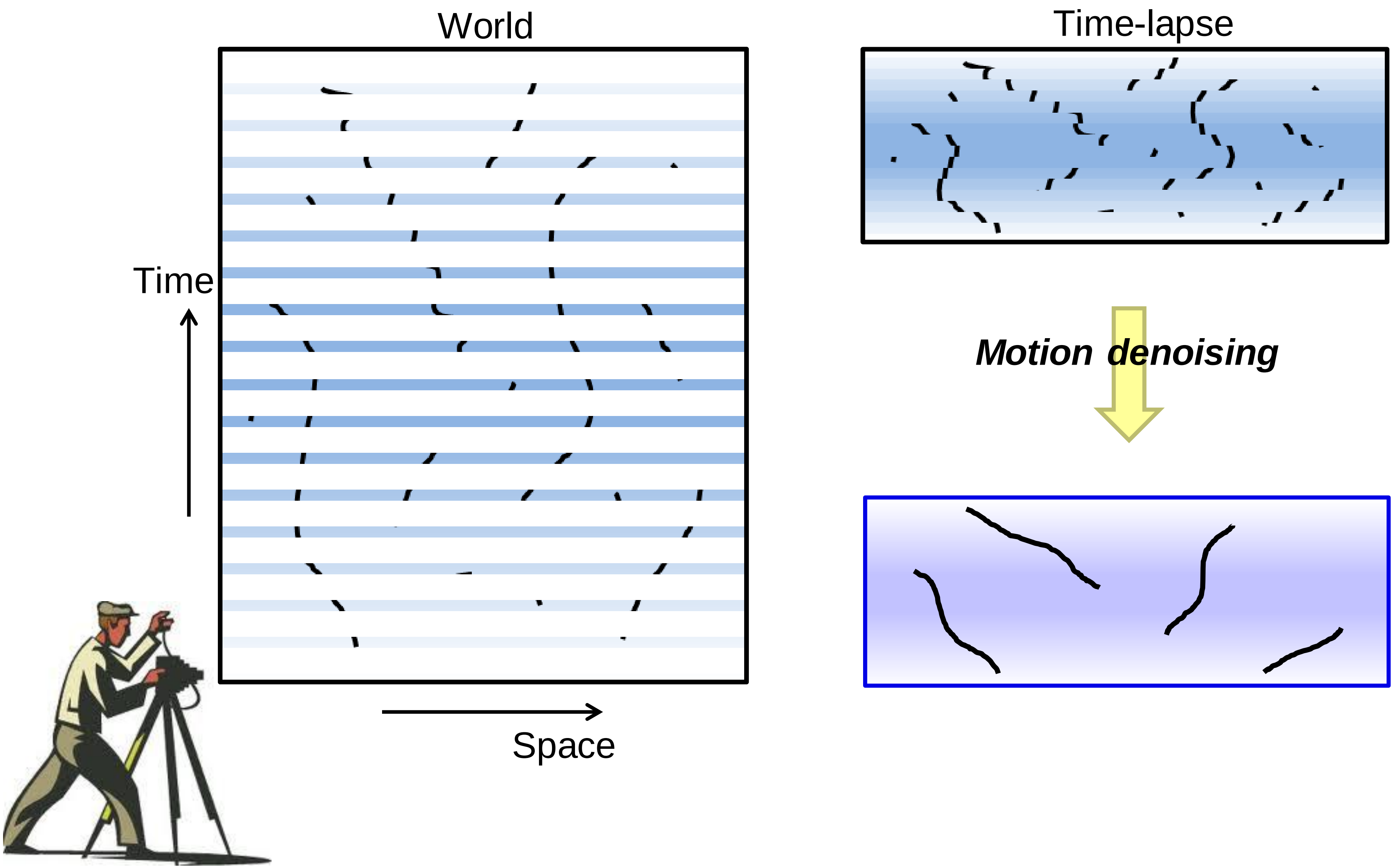
7 years

Source: YouTube

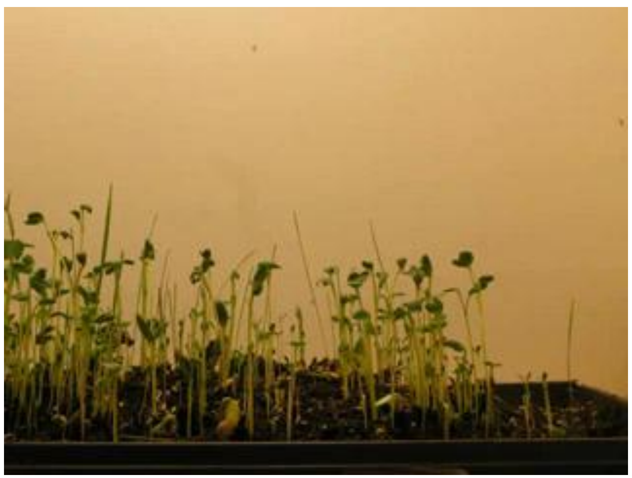
“Stylized Jerkiness”



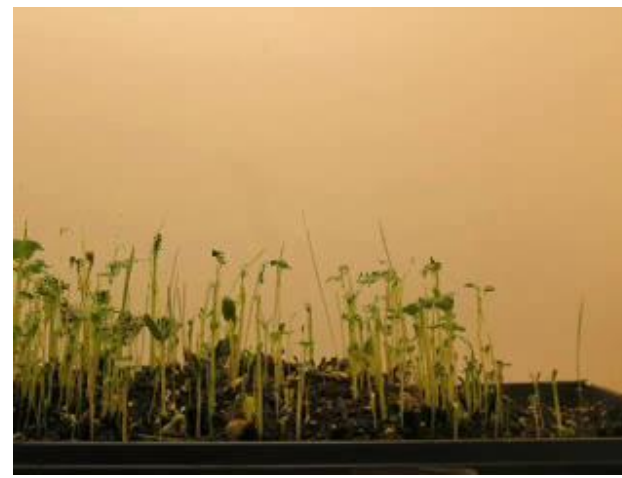
Motion Denoising



Motion Denoising



Motion denoising



Time-lapse in Vision/Graphics Research

- **Video summarization (video → time-lapse)**

[Bennett and McMillan 2007]

[Pritch et al. 2008]

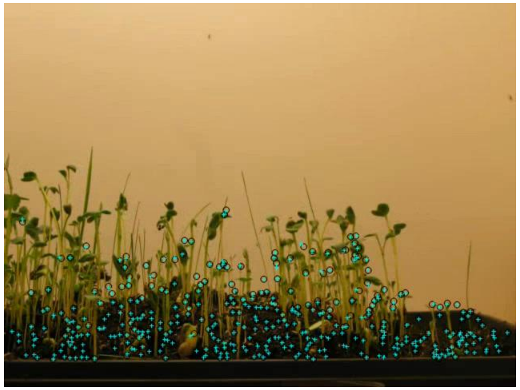
- **Time-lapse editing**

[Sunkavalli et al. 2007]

Motion Denoising is Challenging!

- **Naïve low-pass (temporal) filtering**
  - Pixels of different objects are averaged

- **Smoothing motion trajectories**
  - Motion estimation in time-lapse videos is hard!
    - * **Motion discontinuities**
    - * **Color inconsistencies**



KLT tracks

Formulation

- **Key idea:** long-term events in videos can be statistically explained within some local spatiotemporal support, while short-term events are more distinctive
  - Assumption: world is smooth
  - Short-term variation = *noise*, long-term variation = *signal*

- Our algorithm **reshuffles** the pixels in both space and time to maintain long-term events in the video, while removing short-term noisy motions

Formulation

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

Fidelity (to input)

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

Temporal coherence
(of the result)

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

Regularization
(of the warp)

$p = (x, y, t)$
I – input video, $I(p + w(p))$ – output video
$N_t(p)$ - Temporal neighbors of p, $N(p)$ - Spatiotemporal neighbors of p

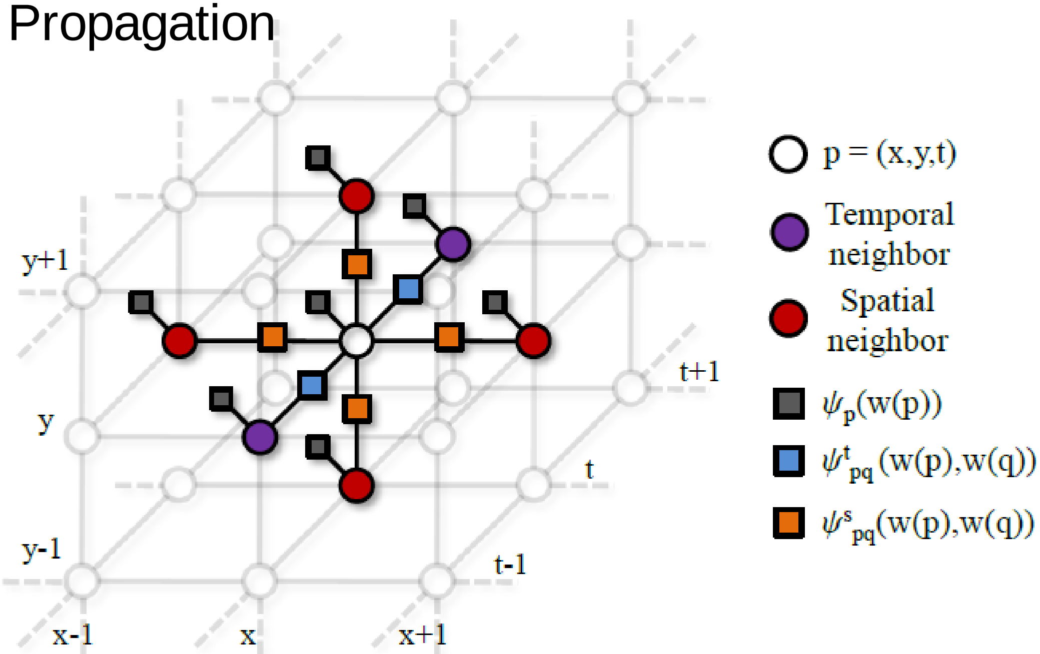
$w(p) \in \{(\delta_x, \delta_y, \delta_t) : |\delta_x| \leq \Delta_s, |\delta_y| \leq \Delta_s, |\delta_t| \leq \Delta_t\}$ - displacement field

$\lambda_{pq} = \exp(-\beta \|I(p) - I(q)\|^2), \quad \beta = (2\langle \|I(p) - I(q)\|^2 \rangle)^{-1}$

Optimization

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

- **Complicated (huge!) inference problem**
  - E.g. 500^3 nodes, 10^3 states per node
  - Optimize using Loopy Belief Propagation



Optimization

- **Potential functions** **message passing**
  - Message structure stored on disk; read and write message chunks on need

$$\psi_p\big(w(p)\big) = \big|I\big(p + w(p)\big) - I(p)\big|$$

Linear in state space +
Pre-compute

$$\psi_{pr}^t\big(w(p), w(r)\big) = \alpha\big\|I\big(p + w(p)\big) - I\big(r + w(r)\big)\big\|^2 + \gamma\lambda_{pr}|w(p) - w(r)|$$

Quadratic in state space
(non convex)

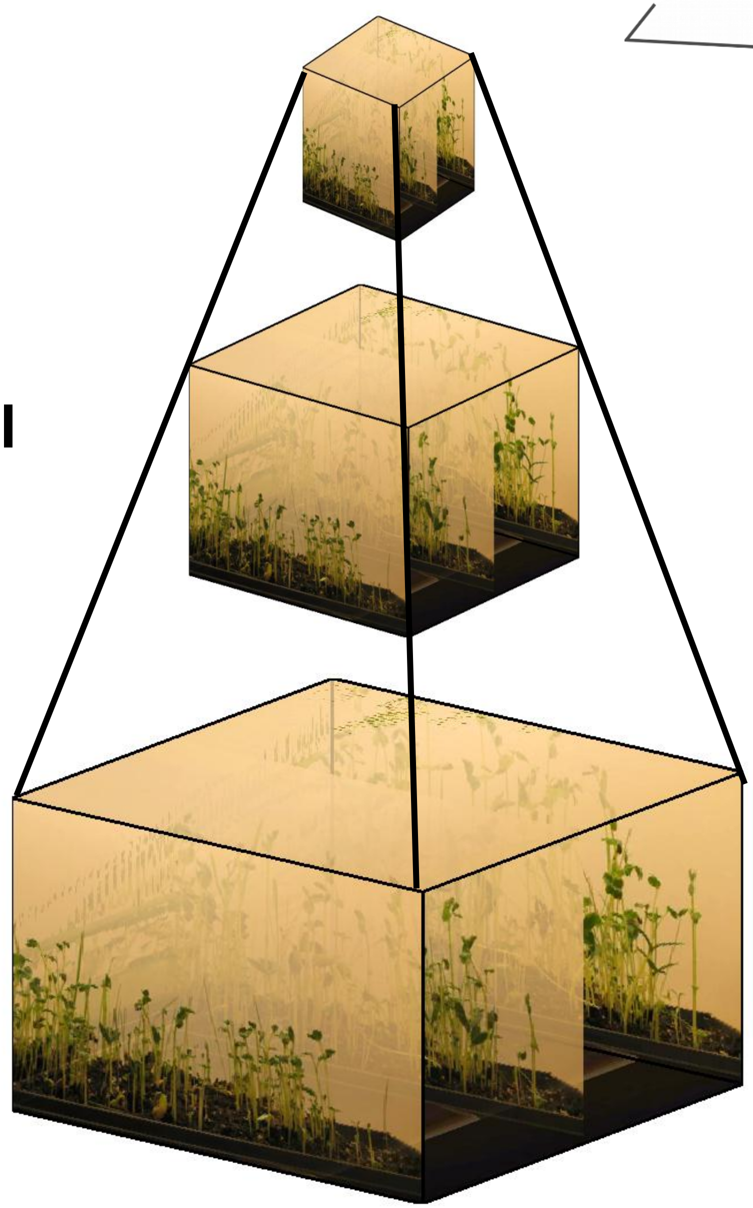
$$\psi_{pq}^t\big(w(p), w(q)\big) = \gamma\lambda_{pq}|w(p) - w(q)|$$

Quadratic in state space
But can be computed in linear time (distance transforms)

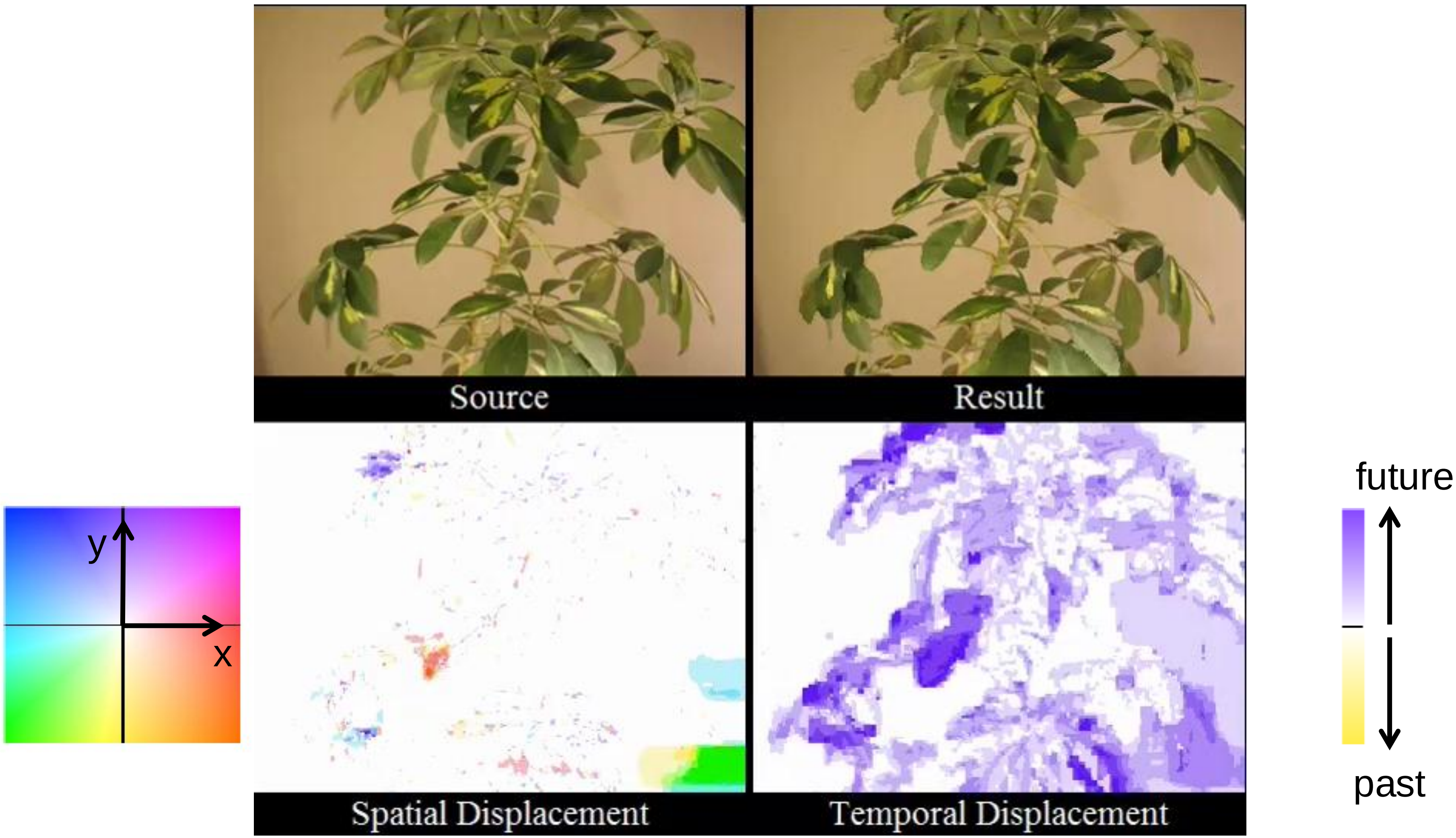
Multi-scale Processing

- **Spatiotemporal video pyramid**
  - Smooth spatially
  - Sample temporally

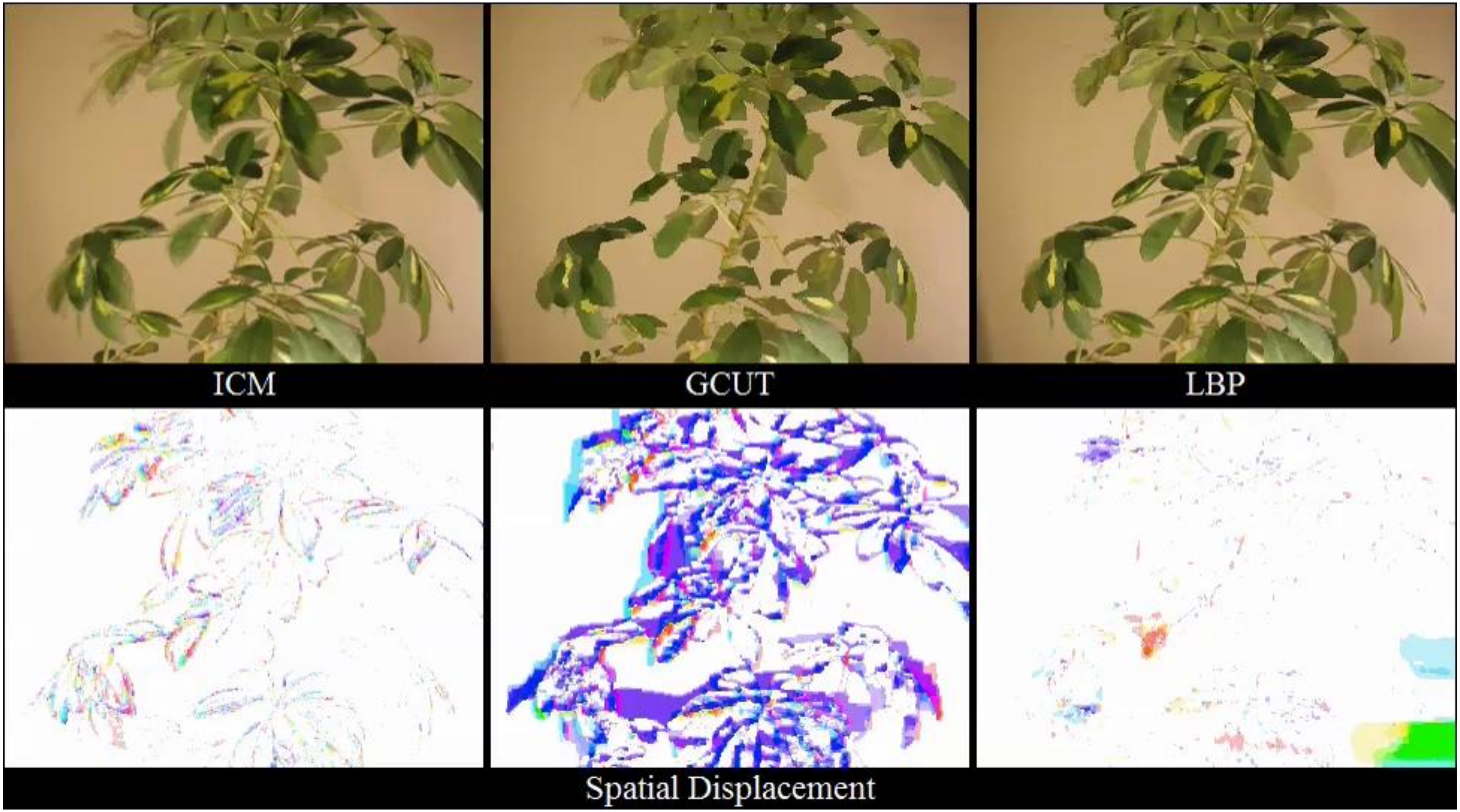
- **Displacements in the coarse level used as centers for the search volume in the finer level**



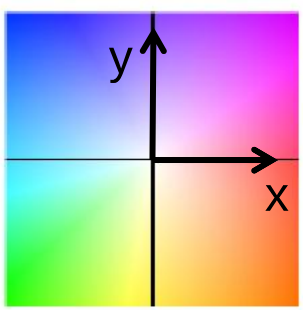
Results

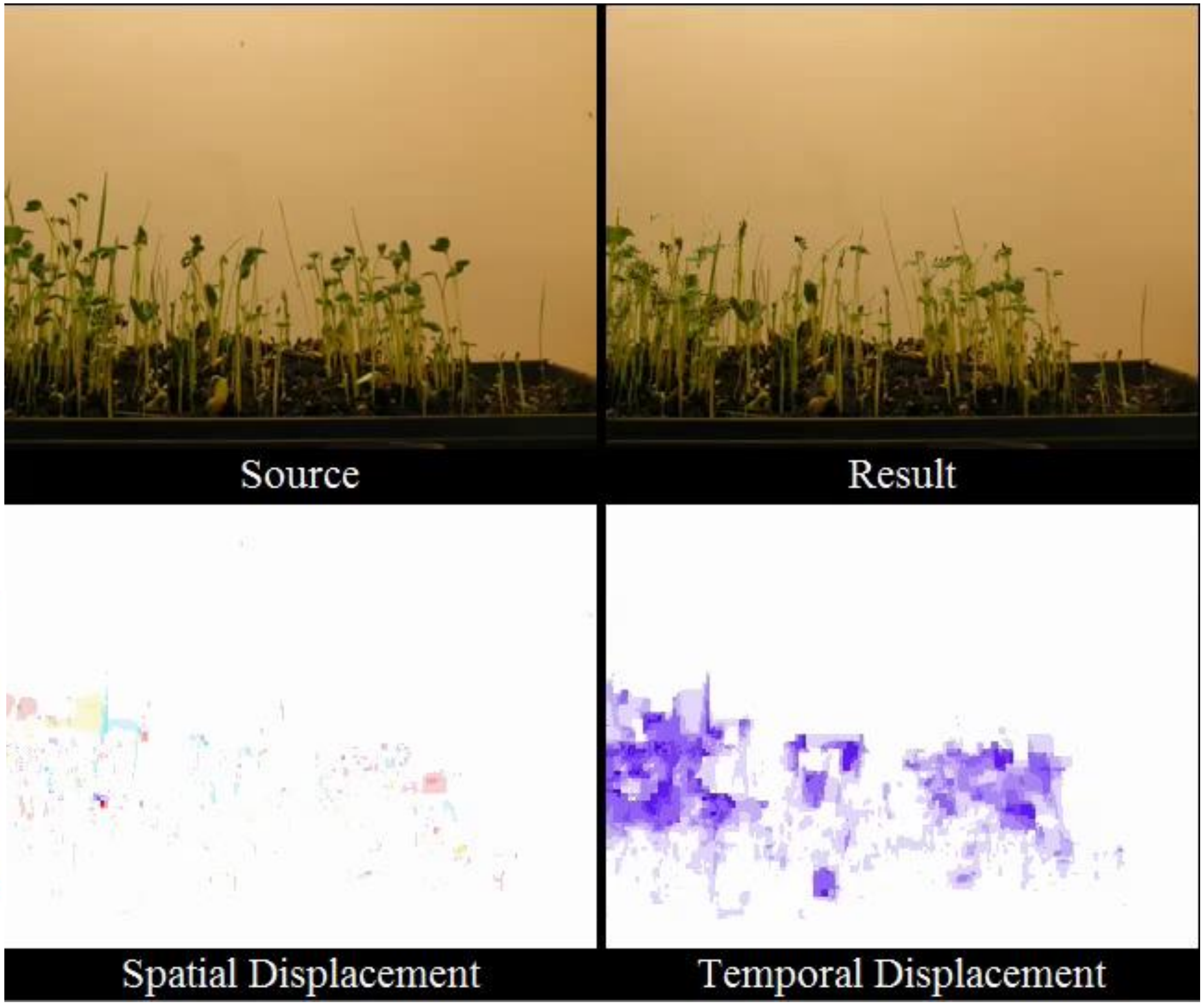

Source
Result
Spatial Displacement
Temporal Displacement
y
x
future
past

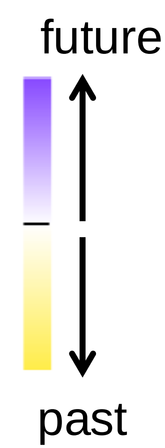
Comparing with Other Optimization Techniques



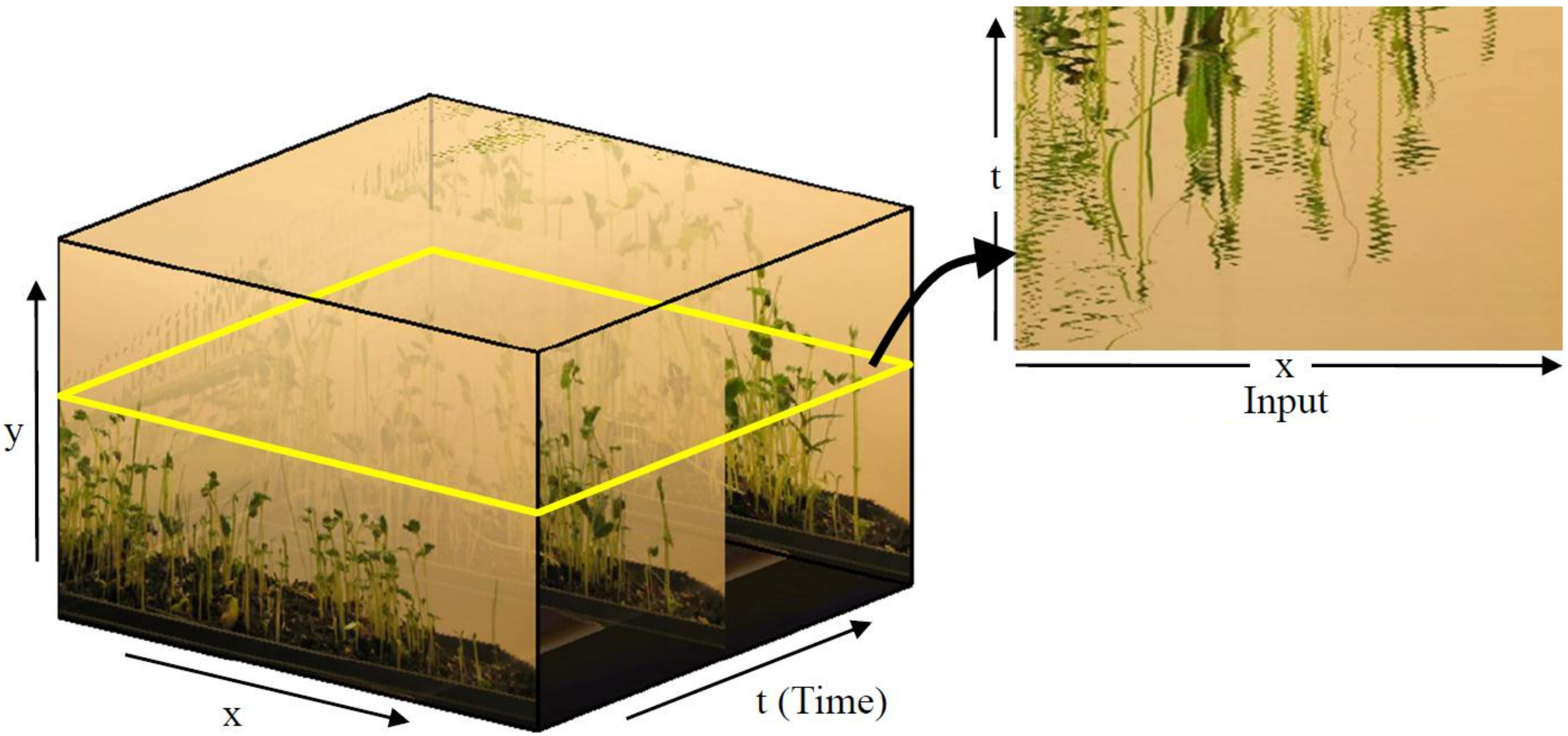
Results


y
x

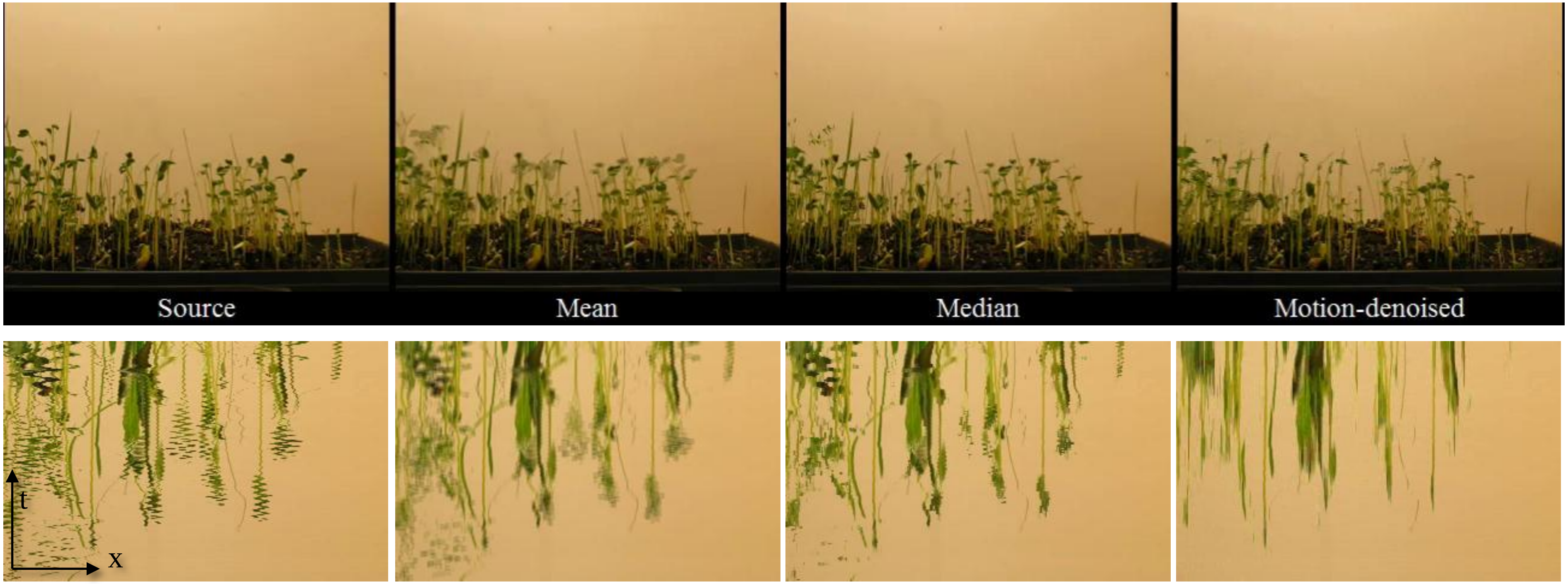

Source
Result
Spatial Displacement
Temporal Displacement


future
past

Results



Comparison with Naïve Temporal Filtering



Support Size

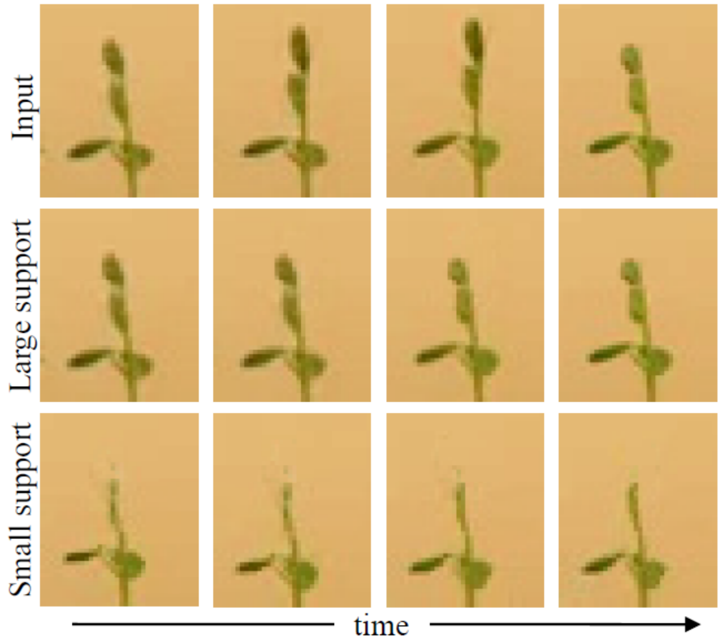
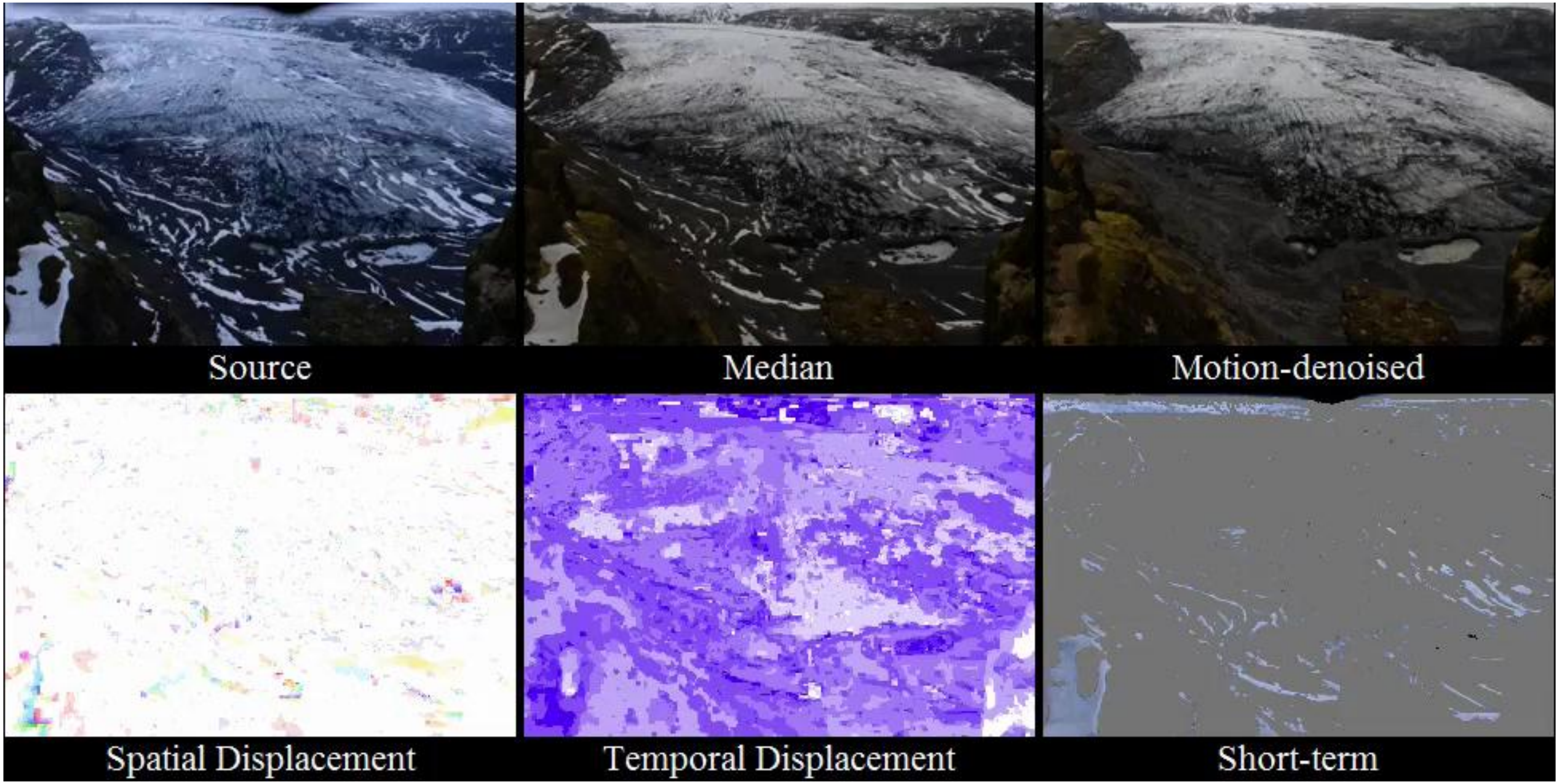


Figure 7. Zoom-in on the rightmost plant in the `sprouts` sequence in four consecutive frames shows that enlarging the search volume used by the algorithm can greatly improve the results. "Large support" corresponds to a $31 \times 31 \times 5$ search volume, while "small support" is the $7 \times 7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

Motion-scale Decomposition

Motion-scale Decomposition

Other Scenarios



Future Work

- **User-controlled motion scales**
  - Not necessarily binary decomposition into long-term and short-term
- **Modify the time-lapse capturing process to help post-processing**
  - E.g. use short videos instead of still images and find best “path” through the video
- **Explore motion-denoising with time-lapse from other domains**
  - Embryos research, satellite imagery

Thank you!

http://csail.mit.edu/mrub/timelapse