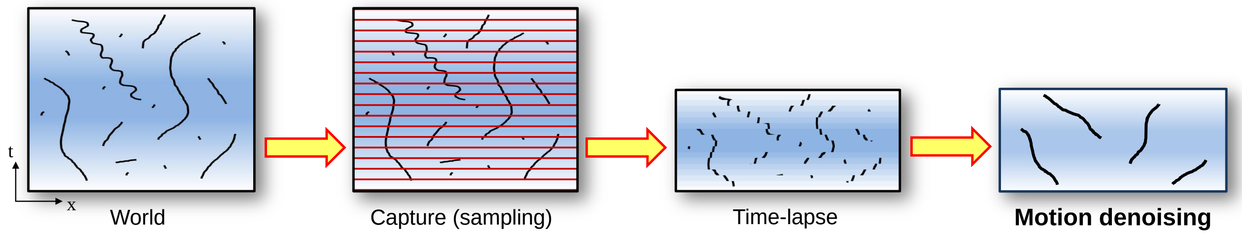
Motion Denoising with Application to Time-lapse Photography

Michael Rubinstein[1], Ce Liu[2,1], Peter Sand, Frédo Durand[1], William T. Freeman[1,2]
[1]MIT CSAIL [2]Microsoft Research New England

Problem Statement

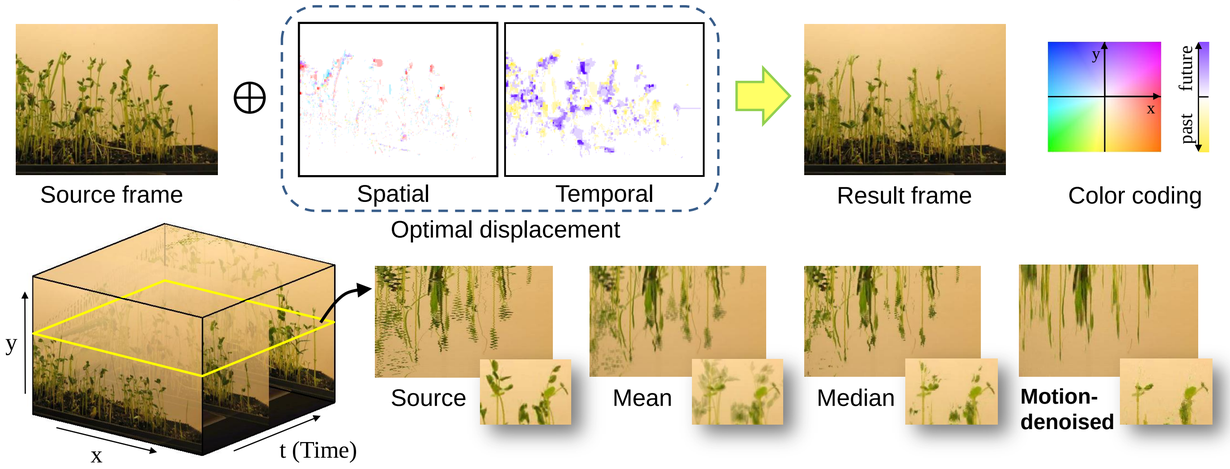
Motions can occur over both short and long time scales. We present motion denoising – the process of **decomposing videos into long-term and short-term motions**, with applications in *time-lapse* videos.



Main contributions:

1. A computational approach to motion denoising without explicit motion analysis
2. A novel application for re-synthesizing time-lapse videos with typical jerky motions *removed* and long-term changes in the scene *preserved*

Motion Re-synthesis



Formulation

- **Key idea:** long-term events in videos can be statistically explained within some local spatiotemporal support, while short-term events are more distinctive
- **Our algorithm reshuffles the pixels in both space and time to maintain long-term events in the video, while removing short-term noisy motions**

$$E(\boldsymbol{w}) = \underbrace{\sum_{p} \left|\boldsymbol{I}\left(p+\boldsymbol{w}(p)\right)-\boldsymbol{I}(p)\right|}_{\text{Fidelity (to input)}} + \alpha \underbrace{\sum_{p,r\in N_t(p)} \left\|\boldsymbol{I}\left(p+\boldsymbol{w}(p)\right)-\boldsymbol{I}\left(r+\boldsymbol{w}(r)\right)\right\|^2}_{\text{Temporal coherence (of the result)}} + \gamma \underbrace{\sum_{p,q\in N(p)} \lambda_{pq}\left|\boldsymbol{w}(p)-\boldsymbol{w}(q)\right|}_{\text{Regularization}}$$

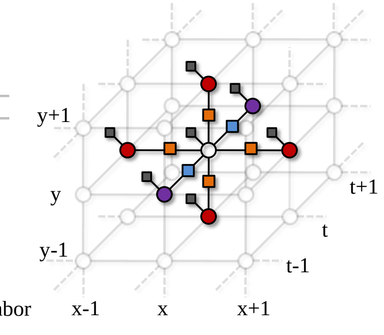
$p=(x,y,t)$, $\boldsymbol{I}$ – input video, $\boldsymbol{J}(p)=\boldsymbol{I}\left(p+\boldsymbol{w}(p)\right)$ – output video, $N(p), N_t(p)$ – spatiotemporal, temporal neighbors resp.

$\boldsymbol{w}(p) \in \left\{\left(\delta_x,\delta_y,\delta_t\right): |\delta_x| \le \Delta_s, \left|\delta_y\right| \le \Delta_s, |\delta_t| \le \Delta_t\right\}$ – displacement field

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

Optimization

- Massive discrete MRF inference (huge grid graph and state space)
  - State space size relates to the space-time extent of events captured
- Pairwise potentials partially non-metric
- **Space-time optimized Loopy Belief Propagation**
  - Sequential message passing backed by the disk
  - Spatial messages in linear time (in the support size) using 3D distance transforms
  - Multi-scale: spatial smoothing, temporal sampling



$\psi_p\left(\boldsymbol{w}(p)\right) = \left|\boldsymbol{I}\left(p+\boldsymbol{w}(p)\right)-\boldsymbol{I}(p)\right|$

$\psi^t_{pr}\left(\boldsymbol{w}(p),\boldsymbol{w}(r)\right) = \alpha\left\|\boldsymbol{I}\left(p+\boldsymbol{w}(p)\right)-\boldsymbol{I}\left(r+\boldsymbol{w}(r)\right)\right\|^2 + \gamma\lambda_{pr}\left|\boldsymbol{w}(p)-\boldsymbol{w}(r)\right|$

$\psi^s_{pq}\left(\boldsymbol{w}(p),\boldsymbol{w}(q)\right) = \gamma\lambda_{pq}\left|\boldsymbol{w}(p)-\boldsymbol{w}(q)\right|$

Experimental Results

We present results on time-lapse sequences of different nature and scenes. All videos and results can be found on our project website http://csail.mit.edu/mrub/timelapse.