Content-Preserving Warps for 3D Video Stabilization

Feng Liu
Michael Gleicher
University of Wisconsin-Madison

Hailin Jin
Aseem Agarwala
Adobe Systems, Inc.

THE UNIVERSITY of WISCONSIN MADISON

Adobe

Images courtesy Peter Sand and Flickr user Charles W. Brown

Amateur Video

Traditional 2D Video Stabilization

Our Result

Goal

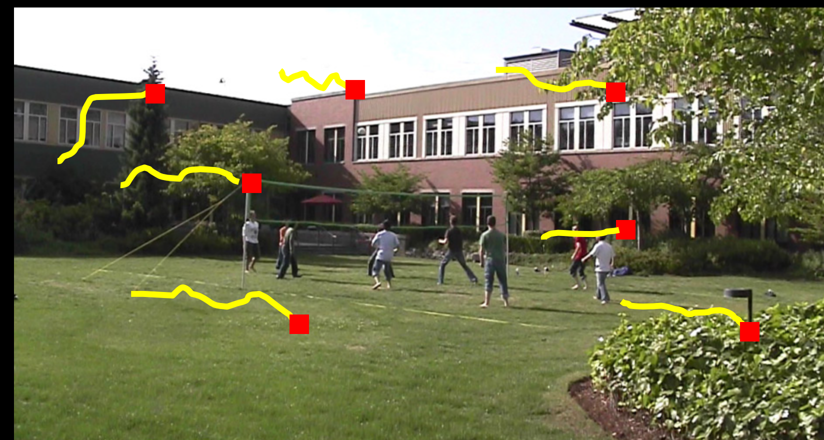
Input

Our result

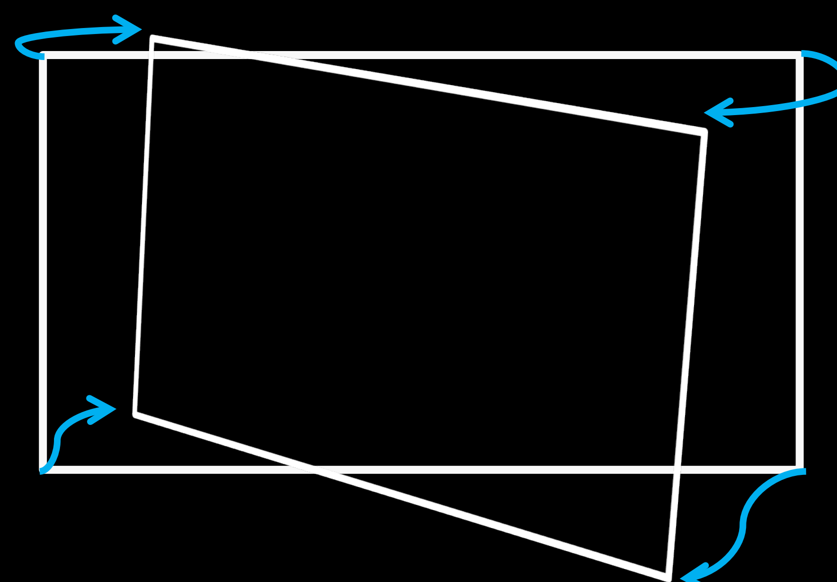
Related Work

Traditional 2D Video Stabilization

Motion estimation



Fit full-frame warps



Limitations

- No knowledge of actual 3D camera path,
  so cannot control desired motion directly

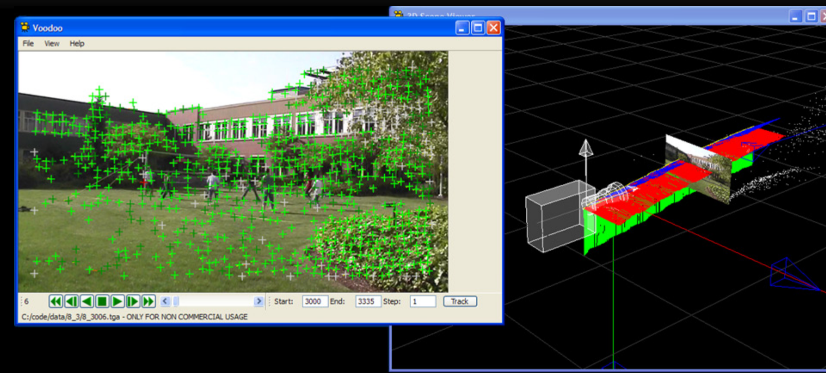
- Homography cannot model parallax in 3D scene

3D Video Stabilization

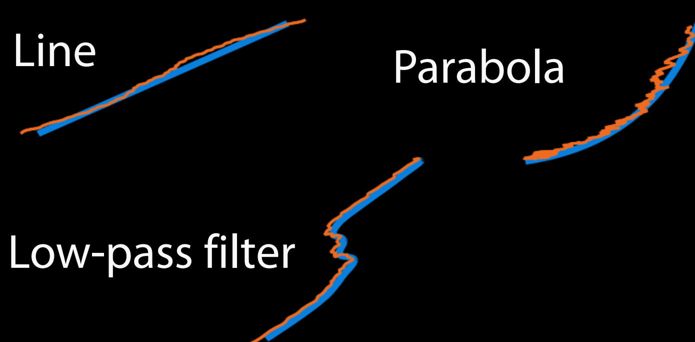
- Non-metric image-based rendering for video stabilization [Buehler et al. 01]
- Image-based rendering using image-based priors [Fitzgibbon et al. 05]
- Using photographs to enhance videos of a static scene [Bhat et al. 07]

3D Video Stabilization: Pipeline

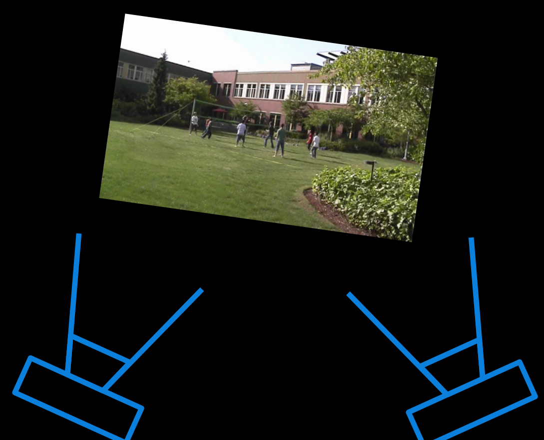
Structure from motion



Camera path planning



Novel view synthesis



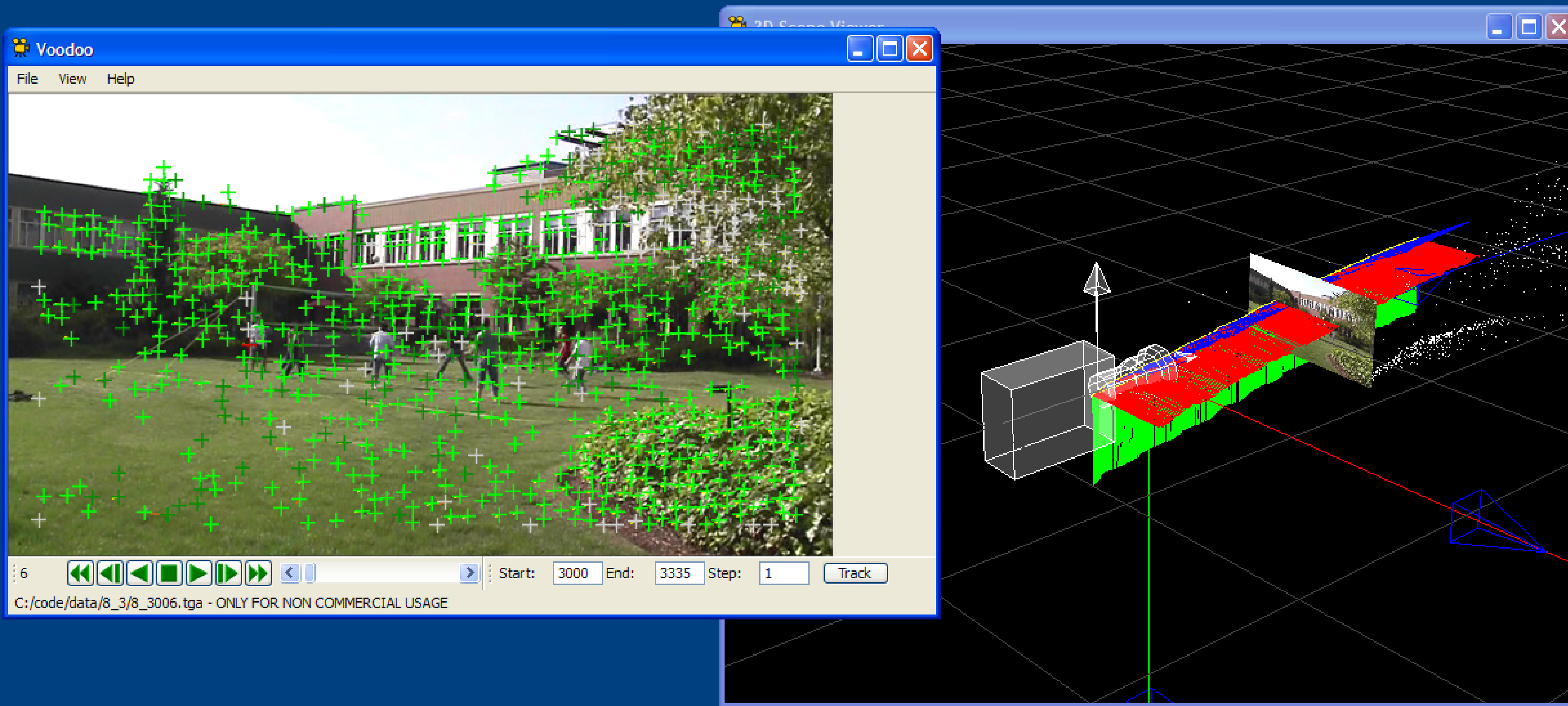
IBR based 3D Video Stabilization

Unstructured lumigraph rendering [Buehler et al. 01]

Our Approach

3D Video Stabilization

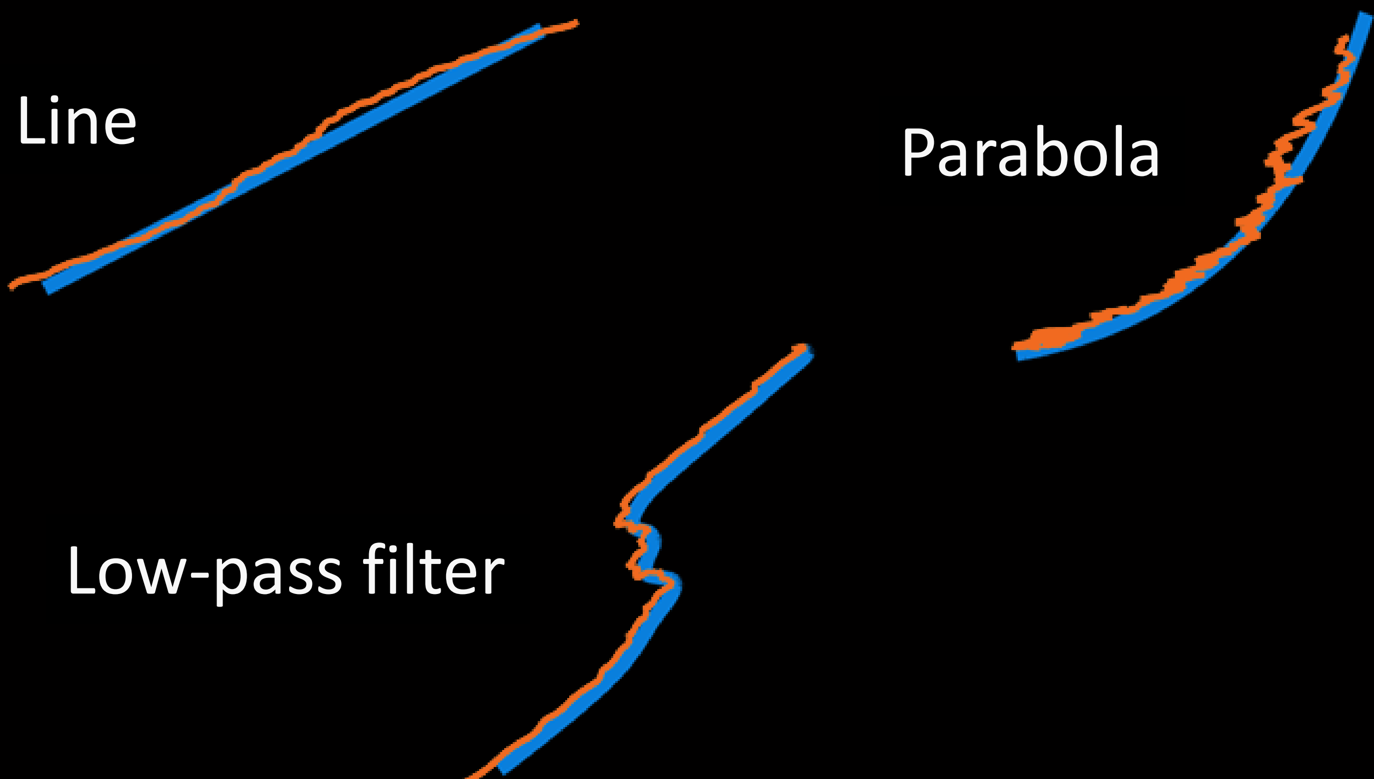
structure from motion



Voodoo Camera Tracker (http://www.digilab.uni-hannover.de)

3D Video Stabilization

camera path planning



Temporal Constraint

One input frame

One output frame

Novel View from One Frame

- A Series of Vision Challenges!
  - Segment out layers
  - Determine depth
  - Shift and re-composite layers
  - Fill holes
- Cannot achieve accurate dis-occlusions, non-lambertian reflection, etc

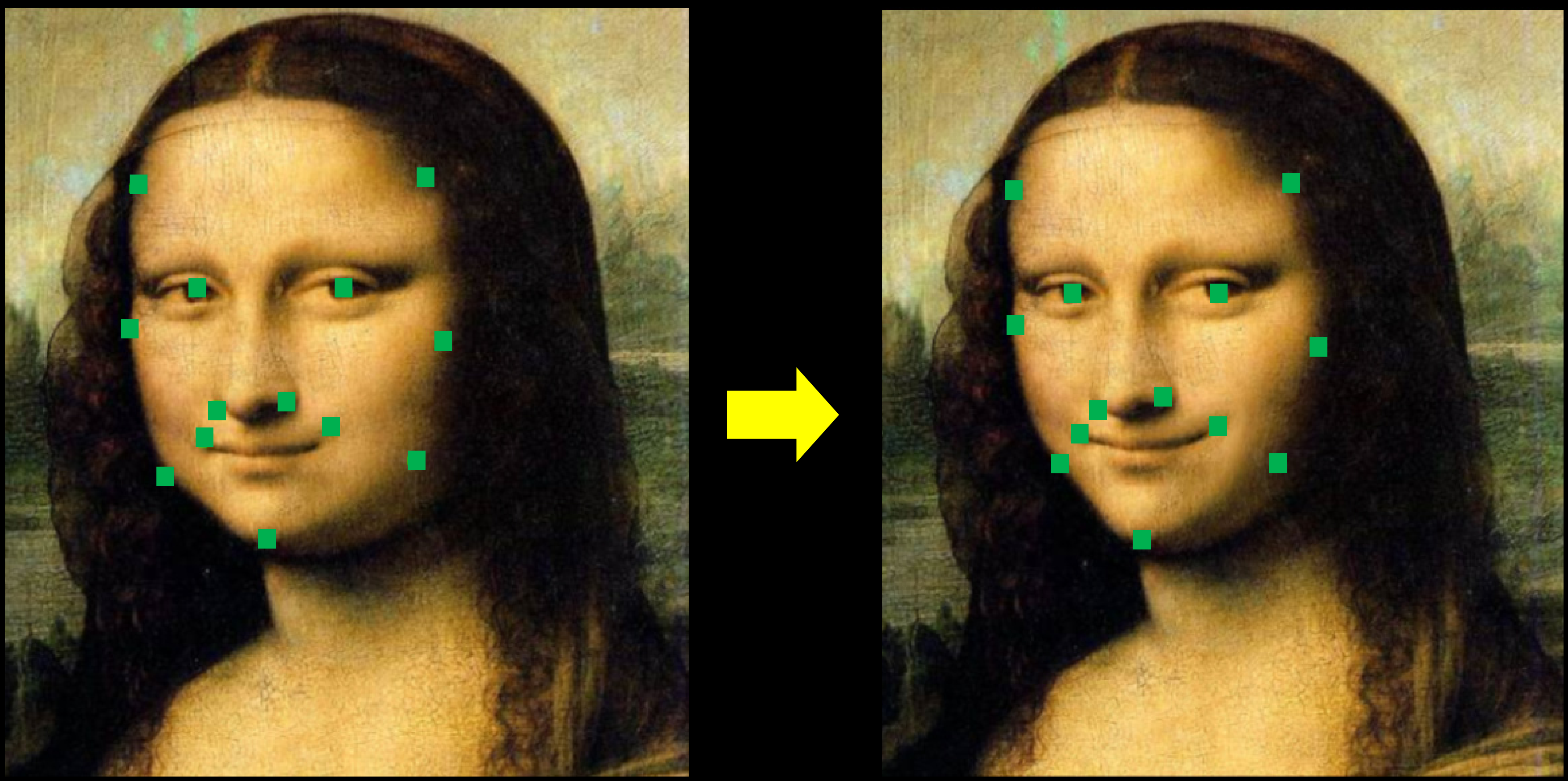
Can We Fake It?

- Viewpoint shifts will be small
- Aim for perceptual plausibility rather than accurate novel view synthesis
  - Salient content move along stabilized paths
  - No noticeable artifacts

Inspirations

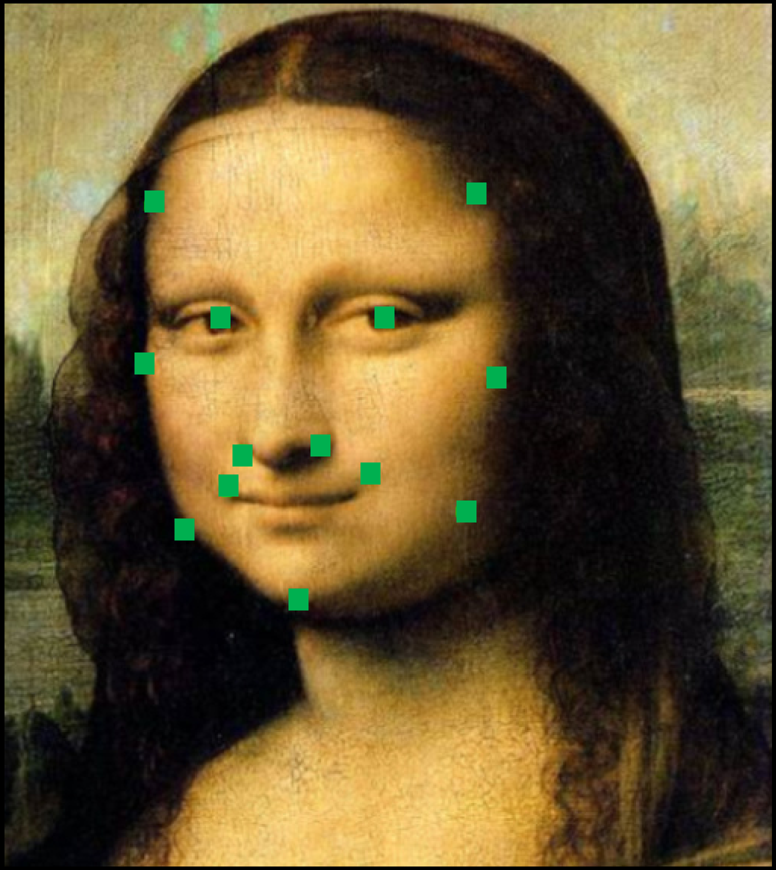
- As-rigid-as possible deformation
  - [Igarashi et al. 05, Schaeffer et al. 06, Gal et al. 06]
- Image retargeting
  - [Wolf et al. 07, Avidan et al. 07, Wang et al. 08]

As Rigid As Possible Deformation



[Schaeffer et al. 06]

As Rigid As Possible Deformation

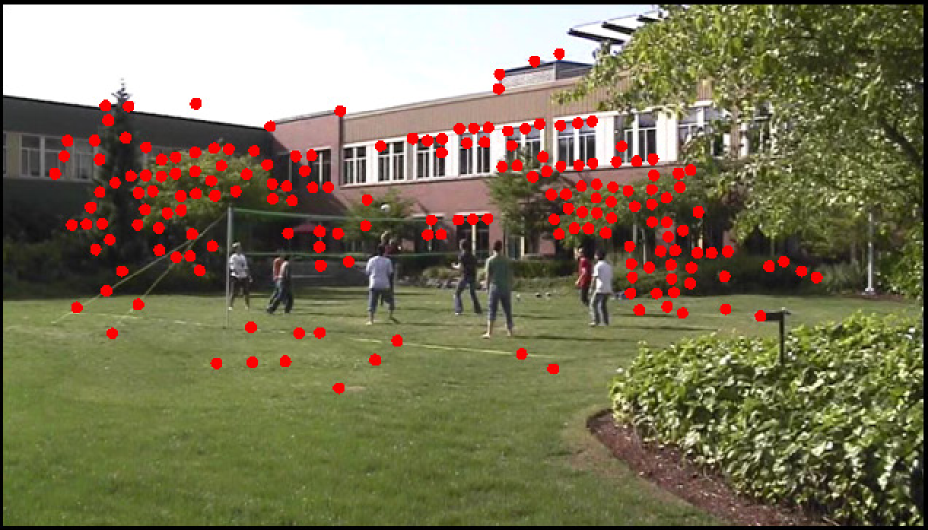


[Schaeffer et al. 06]

Image Retargeting

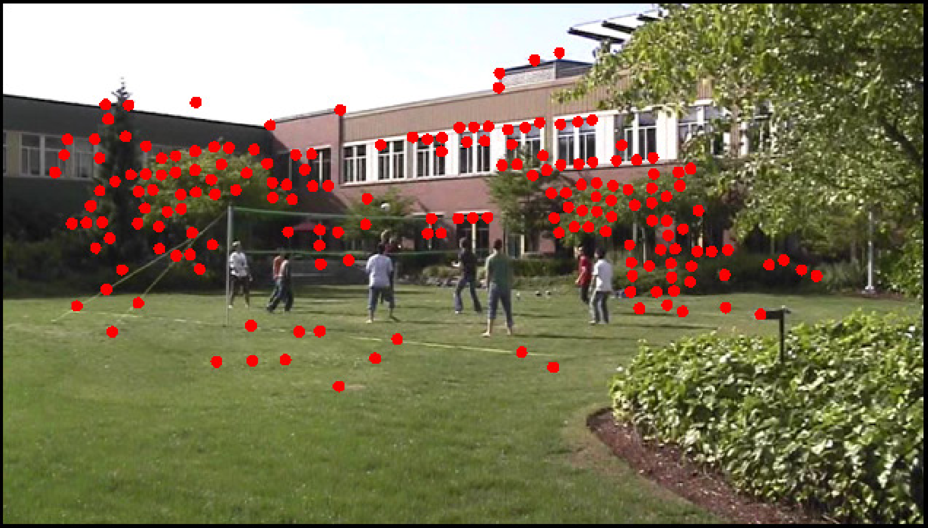
[Avidan et al. 07]

Problem Setup

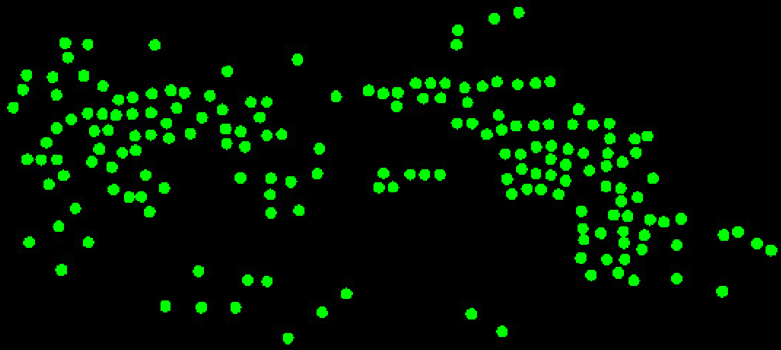


input frame and points

Problem Setup

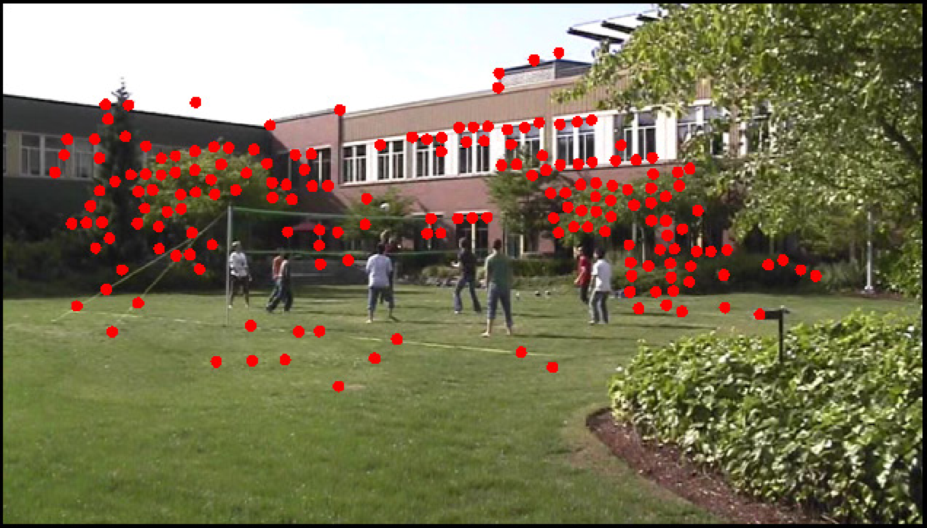


input frame and points



output points

Problem Setup



input frame and points

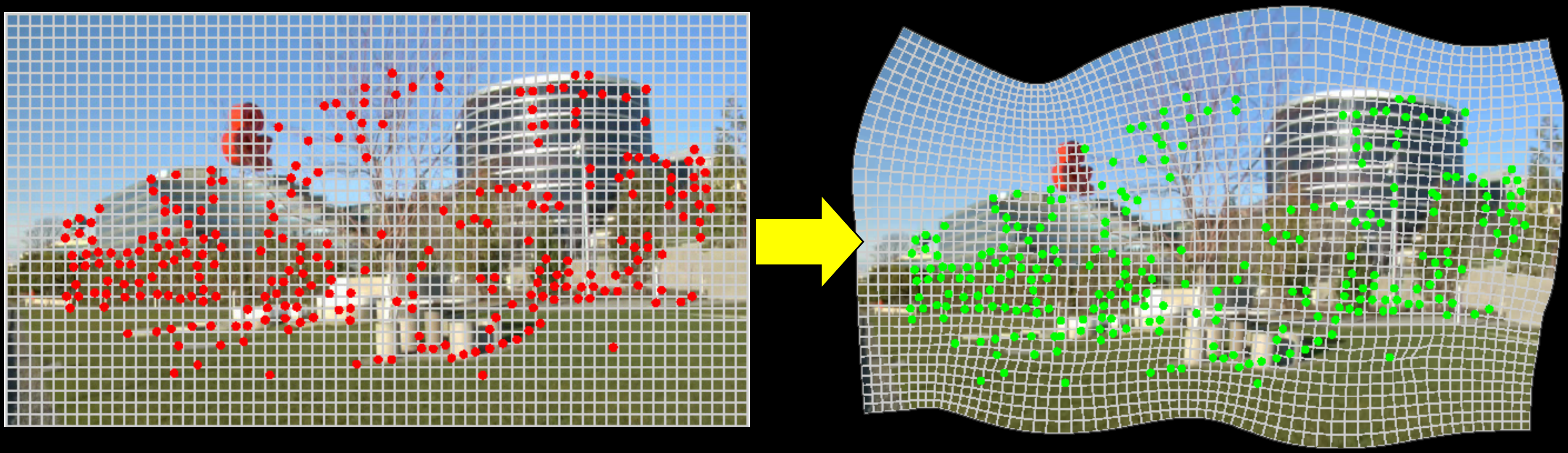
output frame

Option 1: Scattered Data Interpolation

Option 2: Full-frame Warping with Homography

A Less Successful Result

Our Approach

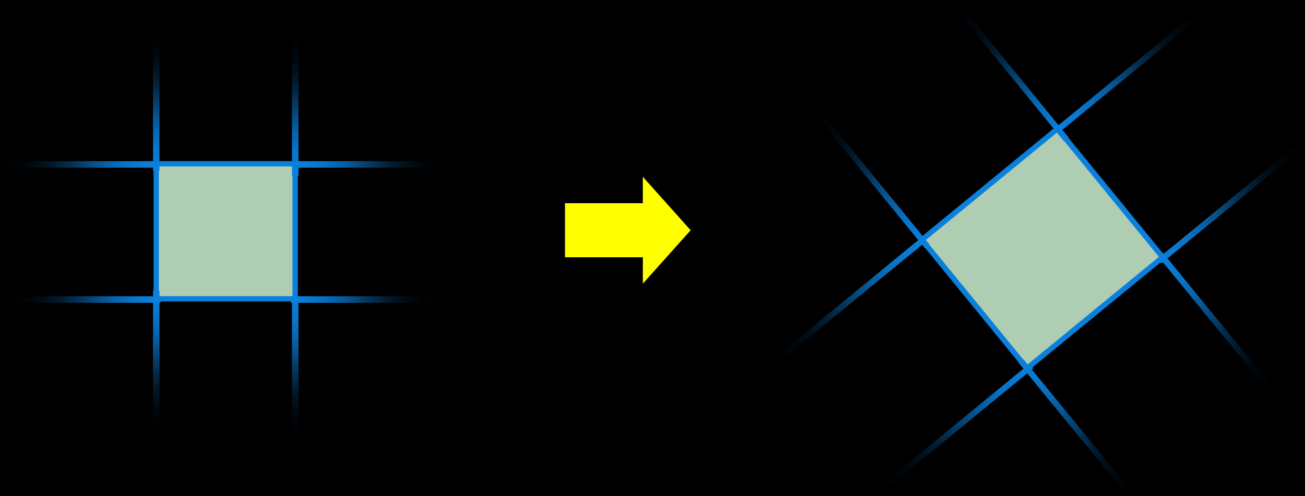


Warp each input frame to create the output frame by least-squares minimization

- ✓ Data term: Soft, sparse displacement constraint
- ✓ Smoothness term: Local similarity transformation constraint

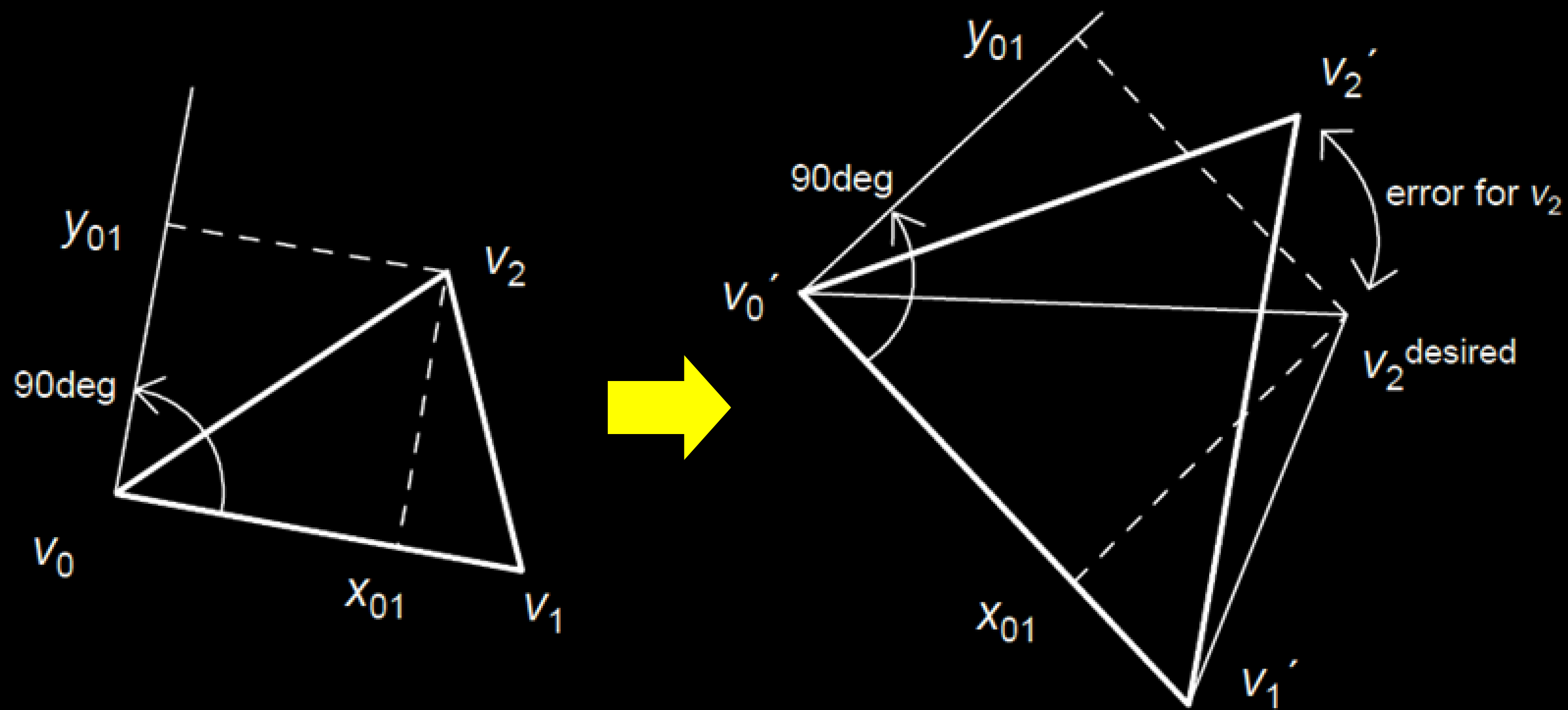
Smoothness Term: Minimize Visual Distortion

Local similarity transformation constraint



Smoothness Term: Minimize Visual Distortion

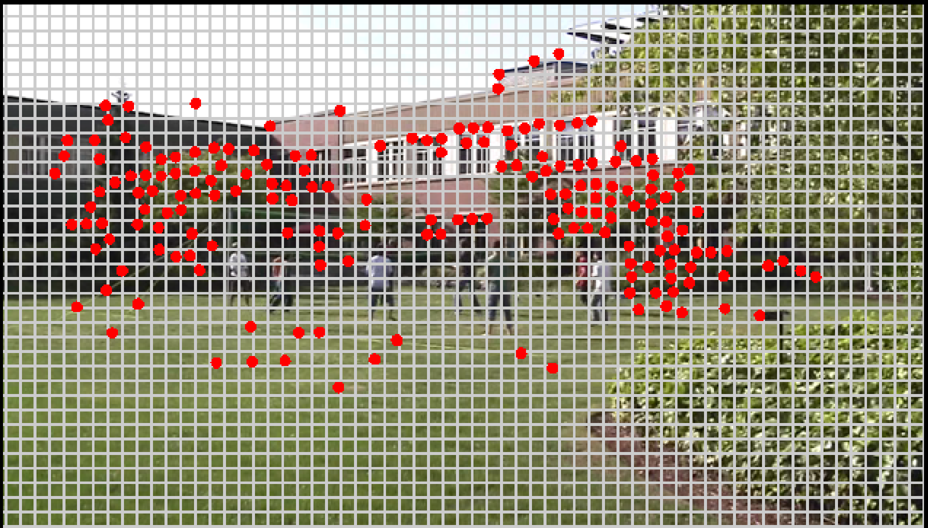
Local similarity transformation constraint



[Igarashi et al. 05]

Saliency Weight

Concentrate distortion to non-salient regions

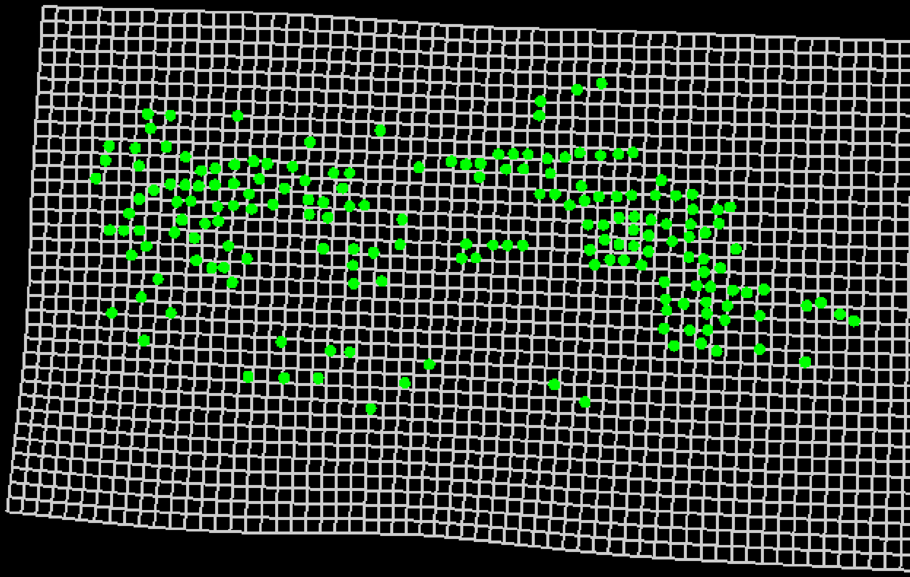


Input

Saliency map

Content-Preserving Warping

Input



Output

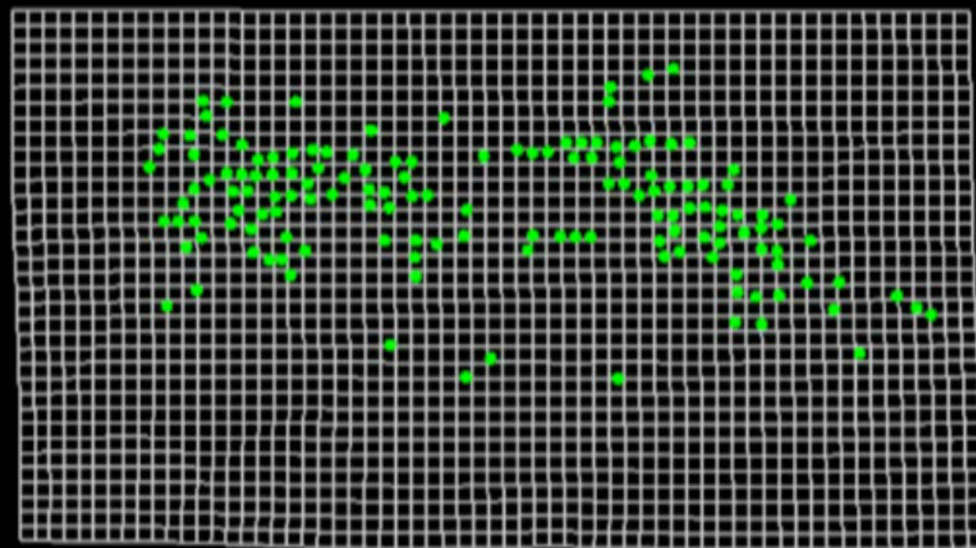
Content-Preserving Warping

Input

Output

Content-Preserving Warping

Grid mesh
& points



Output

More Issues

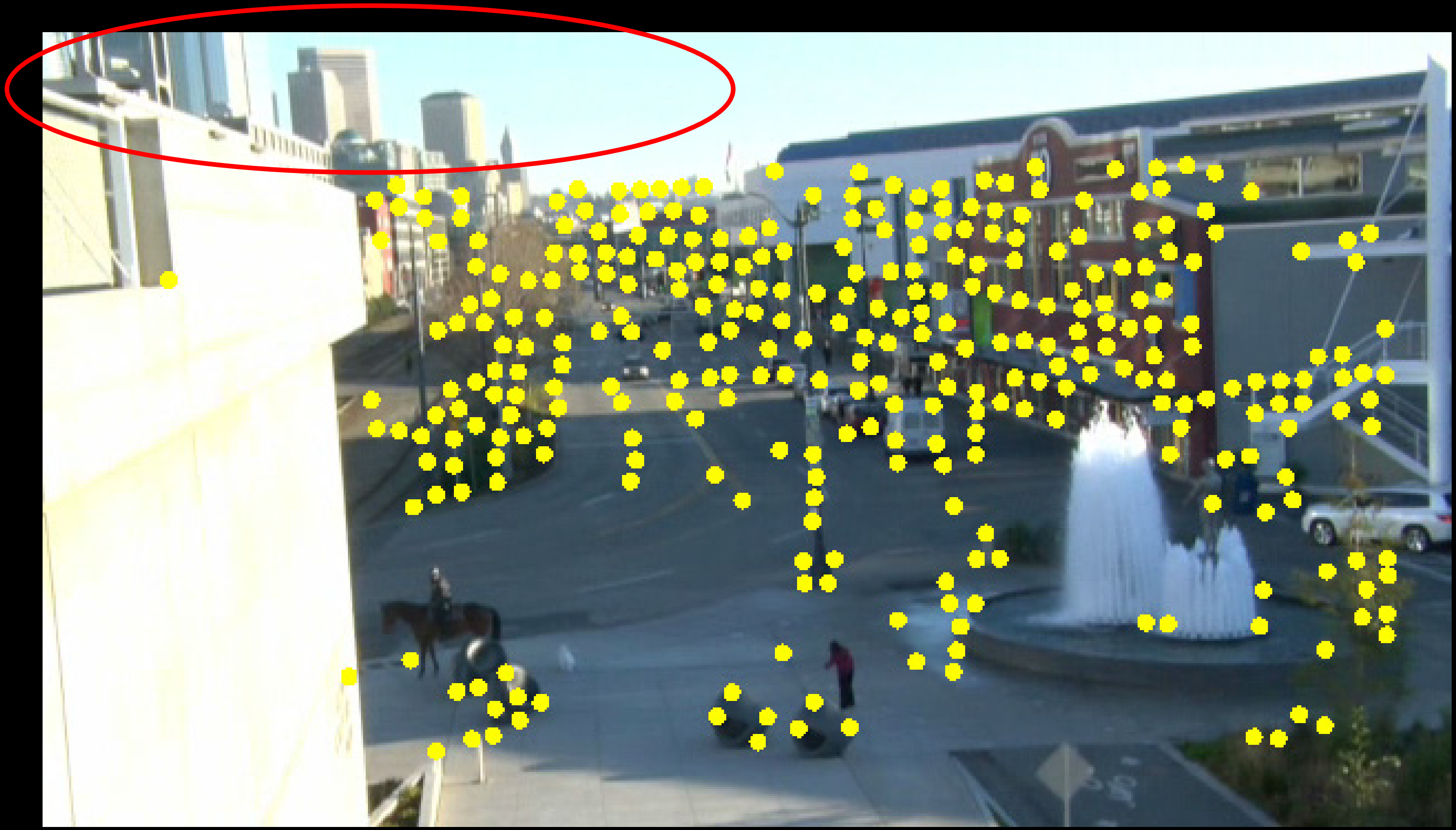
I. Poor distribution of scene points

II. Temporal incoherence from discontinuous scene point tracks

Poor distribution of scene points

Poor distribution of scene points

No feature points



Pre-Warping

Input frame

Pre-Warping

Pre-warping

Pre-Warping

Pre-warping + content-preserving warping

Pre-Warping

Method:

Pre-warp input using a best-fitting homography

Result:

√ Regions with sufficient feature points:
Content-preserving warping dominates

√ Regions without sufficient feature points:
Pre-warping gives a good approximation

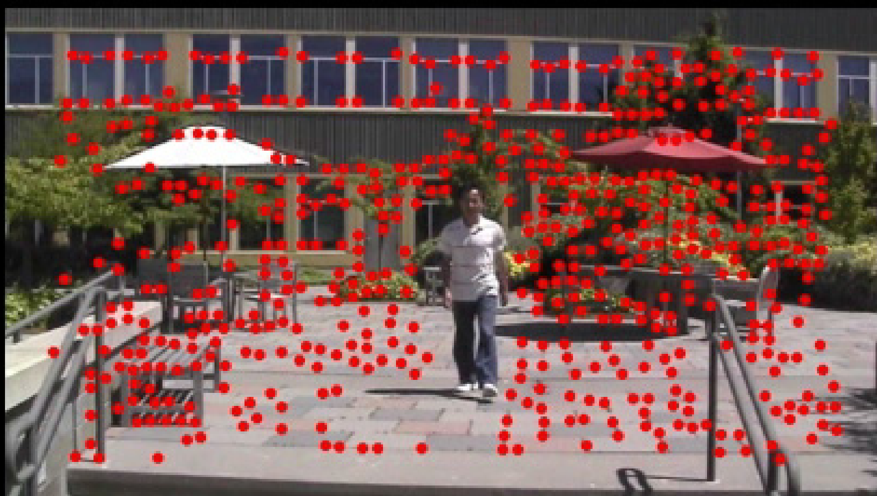
Result of content-preserving warping with pre-warping

More Issues

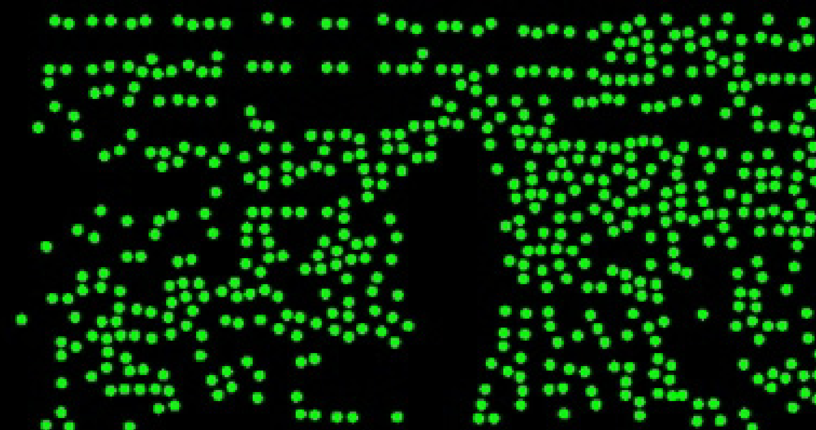
I. Poor distribution of scene points

II. Temporal incoherence from discontinuous scene point tracks

Temporal Coherence



Input video & points

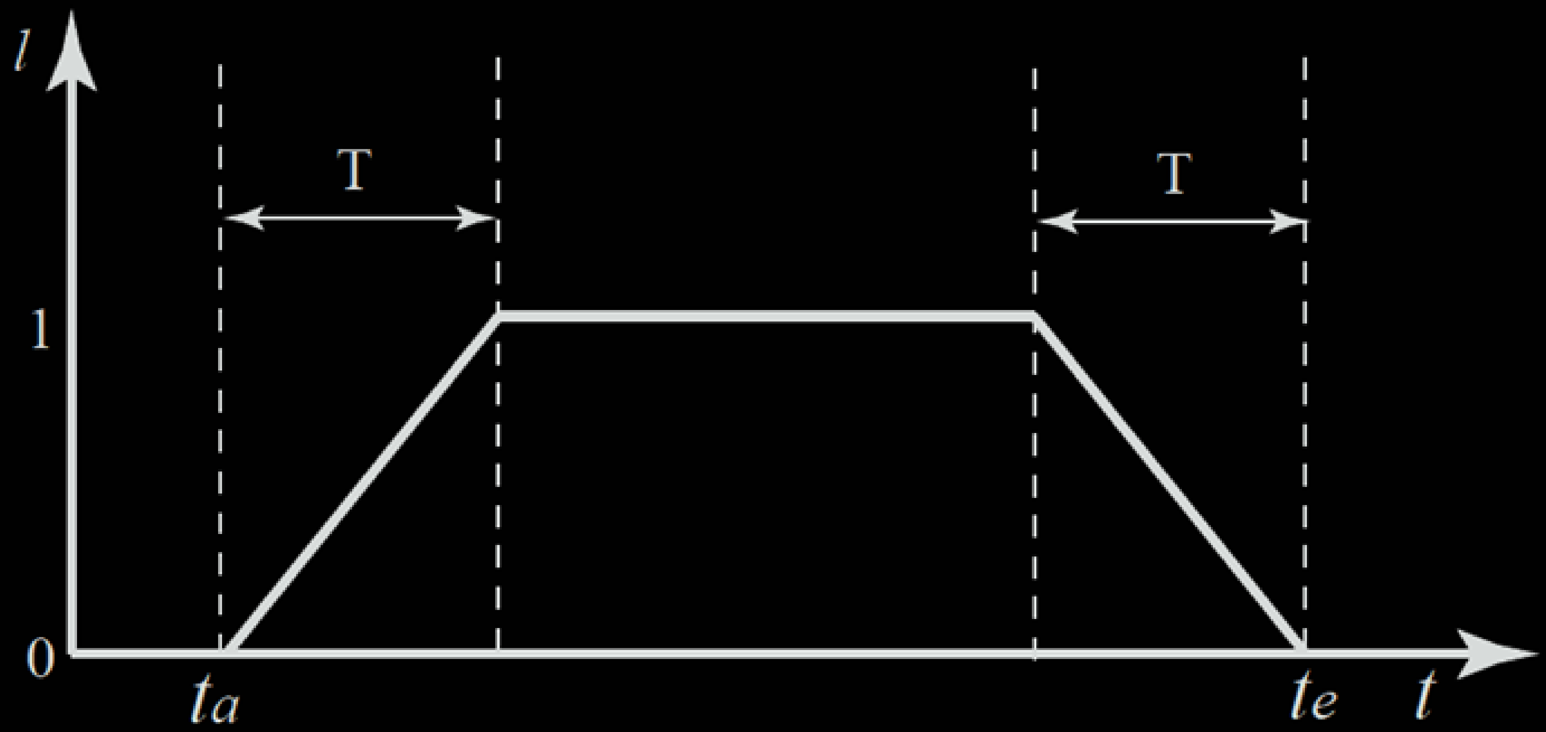


Output points

Temporal Coherence

Temporal Coherence

Temporal Coherence



Fade-in/out the weight of the displacement constraint over time

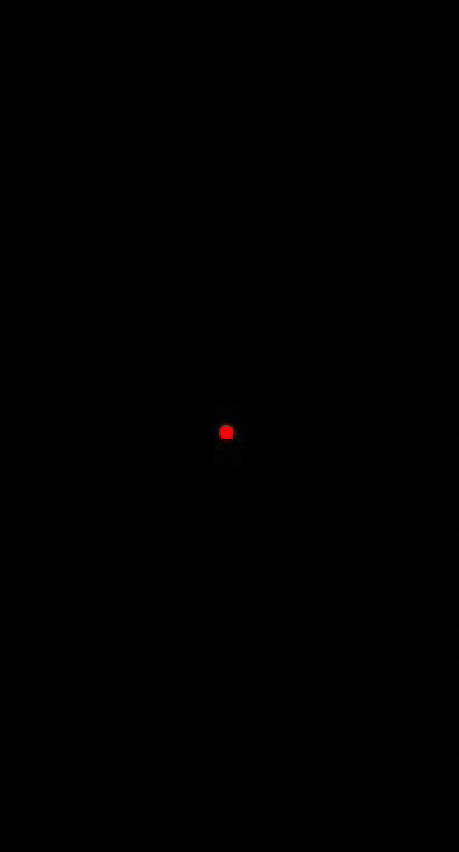
Result of fade-in/out the weight

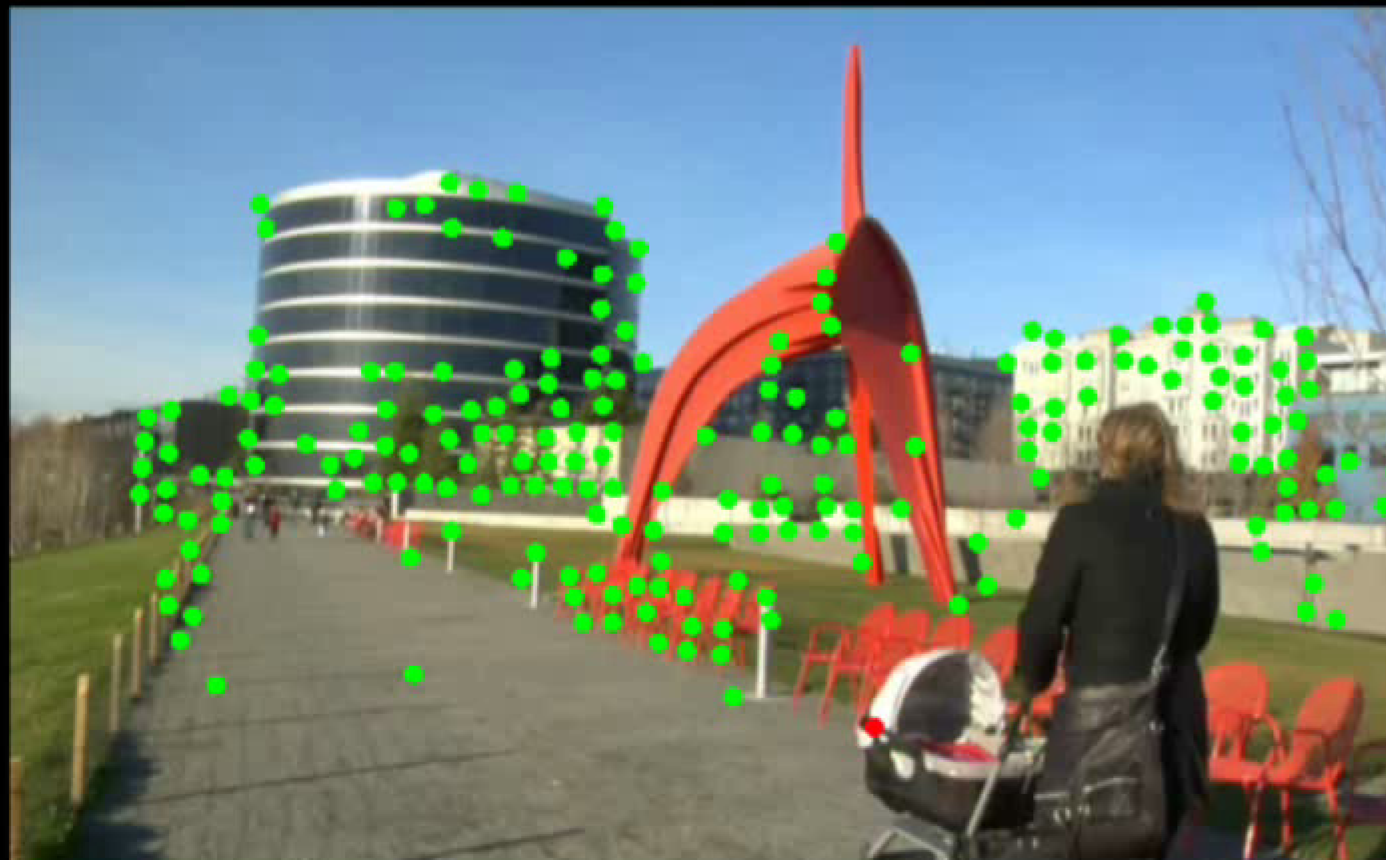
Moving Objects

- No reconstructed 3D points
- Warp follows surrounding background points

- Not correct! But…
  - ✓ Viewpoint shifts are small
  - ✓ Motion clouds the issue
  - ✓ Don’t notice exact occlusion relationships

A Contrived Experiment





Camera position

Output points

Novel View Synthesis

Novel View Synthesis

Novel View Synthesis

Evaluation

- 32 short video clips (5-20 seconds)
  - ✓ completely successful: 14
  - ✓ moderately successful: 15
  - ✓ fail: 3

A Moderately Successful Example

Results & Comparisons

INPUT

OUR OUTPUT

Results & Comparisons

INPUT

OUR OUTPUT

Results & Comparisons

INPUT

OUR OUTPUT

Limitations

- Requires running Structure-from-Motion
  - Slow & memory-intensive
  - More brittle than simple point tracking
  - Requires static regions to lock onto
- Sensitive to poor distribution of recovered scene points
- Aggressive stabilization leads to aggressive cropping

Conclusion

- A technique for simulating the appearance of an idealized camera motion
- Key insight
  - Content-preserving warps can fake small viewpoint shifts
  - Human vision surprisingly tolerant

Thank you

Acknowledgements: We would like to thank the authors of the Voodoo Camera Tracker, and funding from Adobe and NSF grant IIS-0416284.