Computational Photography

Prof. Feng Liu

Spring 2022

http://www.cs.pdx.edu/~fliu/courses/cs510/

05/12/2022

Last Time

- Compositing and Matting

Today

- Video Stabilization
  - Video stabilization pipeline

A Tracking Shot

Orson Welles, *Touch of Evil*, 1958

Images courtesy Peter Sand and Flickr user Charles W. Brown

Input Amateur Video

Traditional 2D Video Stabilization Result

3D Video Stabilization Result [Liu et al. 09]

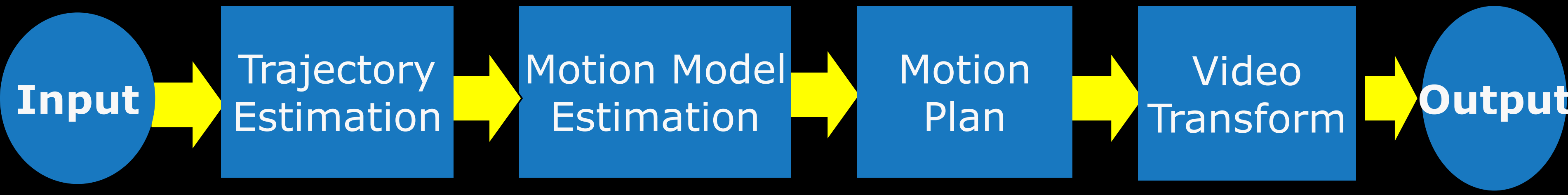
Input

Stabilization result

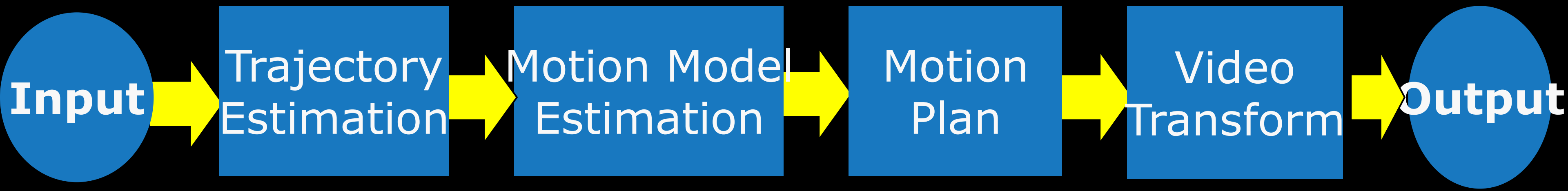
Stabilization: An Old Problem

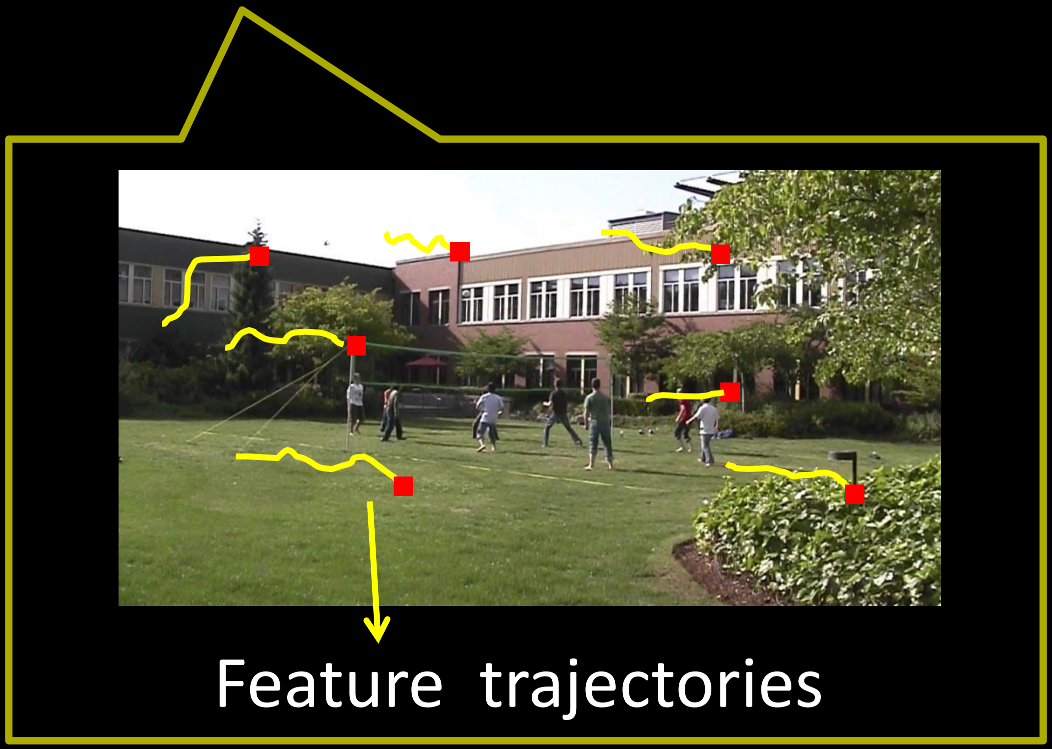
- iMovie from Apple
- De-shaker, a free tool
- Most modern camcorders

Video Stabilization Pipeline

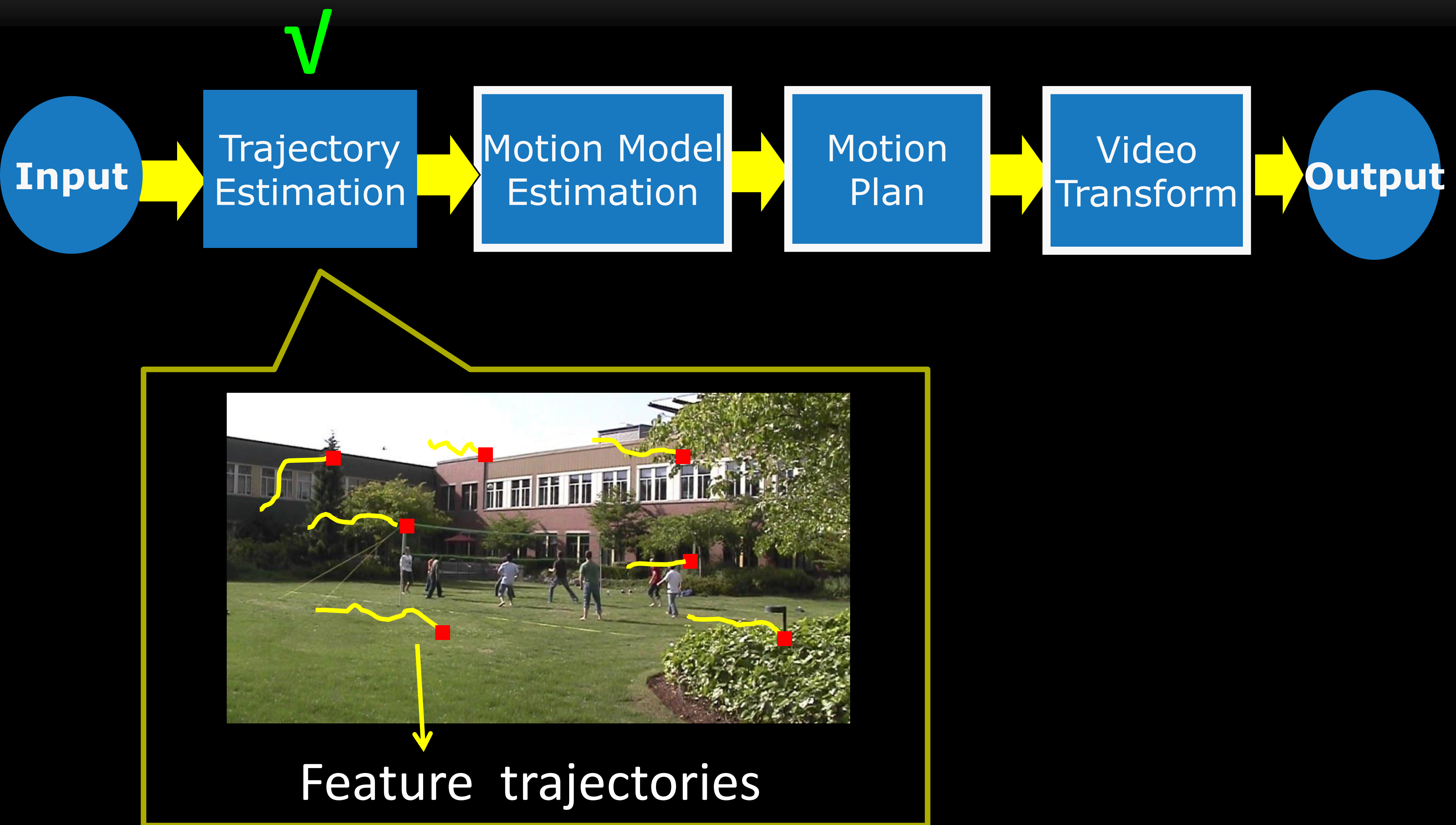


Video Stabilization Pipeline





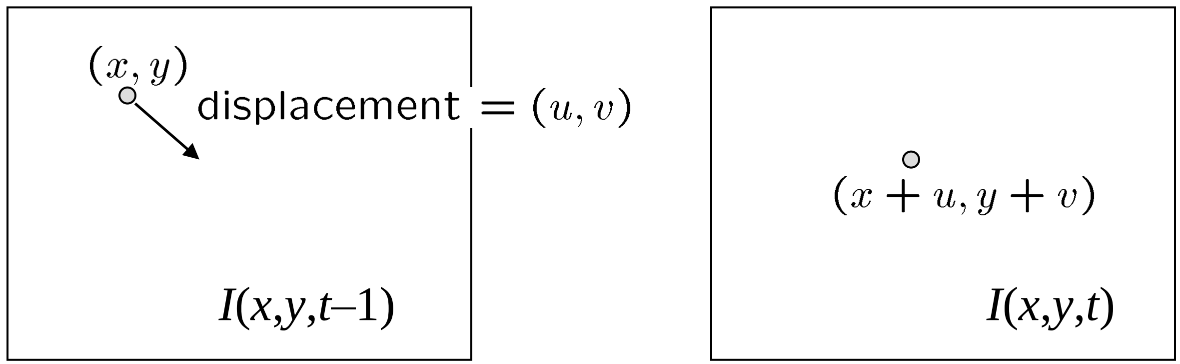
Video Stabilization Pipeline



Trajectory Estimation

- Kanade-Lucas-Tomasi feature tracker (KLT)
  - B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. IJCAI, pp. 674-679, 1981.
  - C. Tomasi and T. Kanade. Detection and Tracking of Point Features. CMU-CS-91-132, 1991.
  - J. Shi and C. Tomasi. Good Features to Track. CVPR, pp. 593-600, 1994.
- Implementations
  - OpenCV
  - http://www.ces.clemson.edu/~stb/klt/
  - …

Feature Tracking



- Brightness Constancy Equation:

$$I(x, y, t-1) = I(x + u(x, y), y + v(x, y), t)$$

- Linearizing the right side using Taylor expansion:

$$I(x, y, t-1) \approx I(x, y, t) + I_x \cdot u(x, y) + I_y \cdot v(x, y)$$

Hence, $$I_x \cdot u + I_y \cdot v + I_t \approx 0$$

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Spatial Coherence Constraint

$$I_x \cdot u + I_y \cdot v + I_t = 0$$

- How many equations and unknowns per pixel?
  - One equation, two unknowns
- How to get more equations for a pixel?
  - Spatial coherence constraint: pretend the pixel's neighbors have the same (u,v)

$$0 = I_t(\mathbf{p_i}) + \nabla I(\mathbf{p_i}) \cdot [u \;\; v]$$

$$\begin{bmatrix} I_x(\mathbf{p_1}) & I_y(\mathbf{p_1}) \\ I_x(\mathbf{p_2}) & I_y(\mathbf{p_2}) \\ \vdots & \vdots \\ I_x(\mathbf{p_{25}}) & I_y(\mathbf{p_{25}}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p_1}) \\ I_t(\mathbf{p_2}) \\ \vdots \\ I_t(\mathbf{p_{25}}) \end{bmatrix}$$

Solving the Tracking Problem

- Least squares problem:

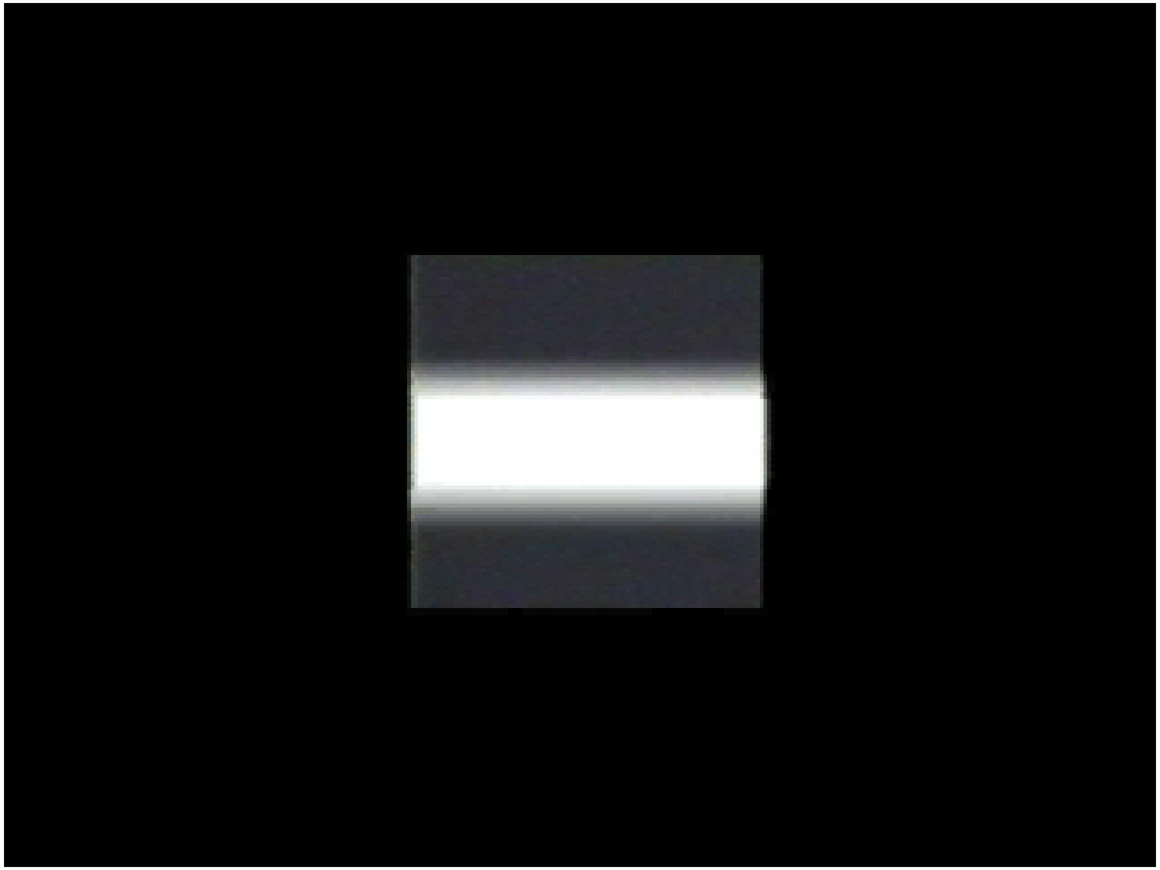
$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\times2}{A} \; \underset{2\times1}{d} = \underset{25\times1}{b}$$

- When is this system solvable?
  - What if the window contains just a single straight edge?

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Conditions for Solvability

- "Bad" case: single straight edge



Lucas-Kanade Flow

- Least squares problem:

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix}$$

$$\underset{25\text{x}2}{A} \quad \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

Solution given by $(A^T A)\, d = A^T b$

$$\underbrace{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

The summations are over all pixels in the window

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

Lucas-Kanade Flow

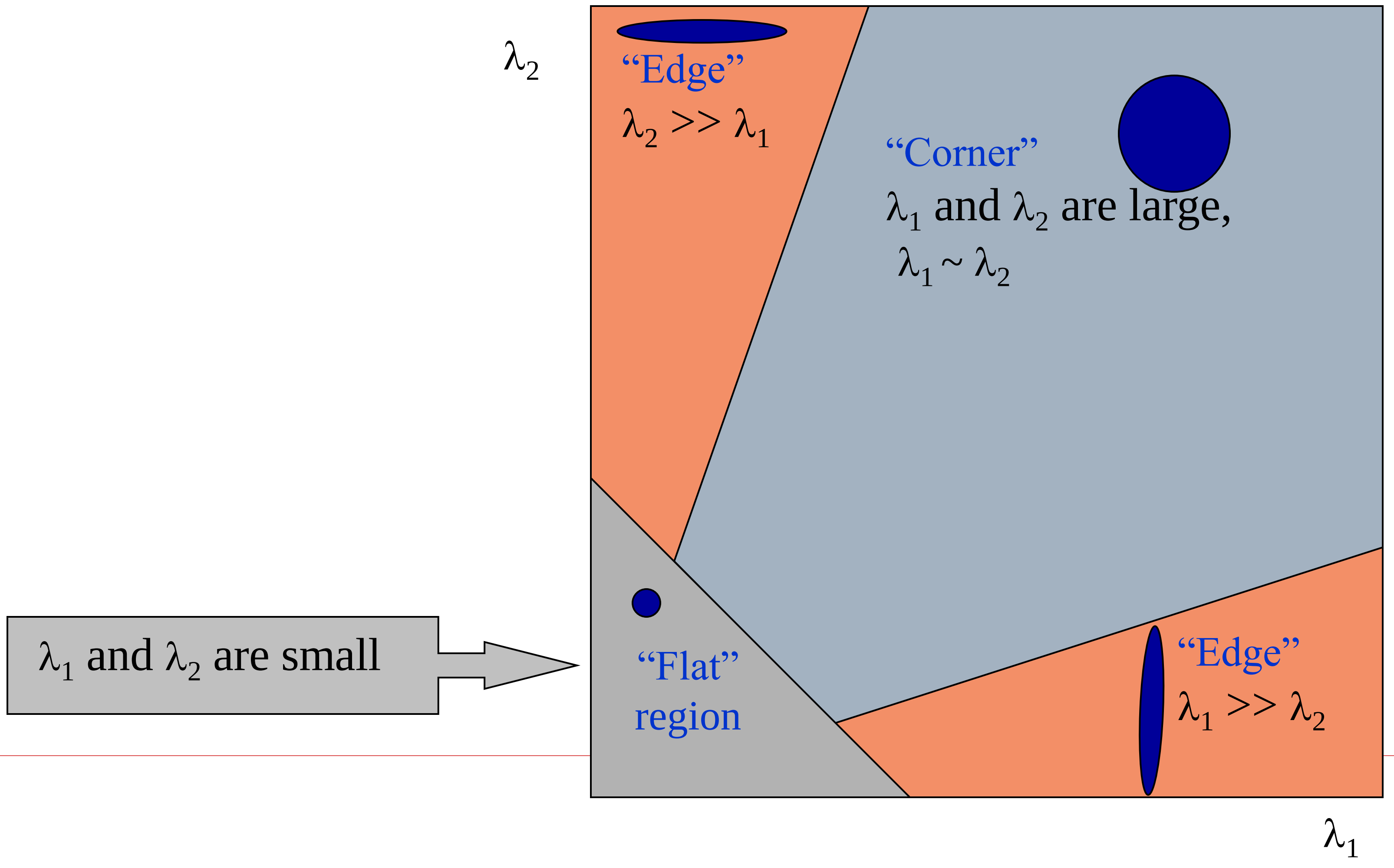
$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$

$$A^T A \qquad\qquad\qquad A^T b$$

- Recall the Harris corner detector: $M = A^T A$ is the *second moment matrix*
- We can figure out whether the system is solvable by looking at the eigenvalues of the second moment matrix
  - The eigenvectors and eigenvalues of M relate to edge direction and magnitude
  - The eigenvector associated with the larger eigenvalue points in the direction of fastest intensity change, and the other eigenvector is orthogonal to it

Interpreting the eigenvalues

Classification of image points using eigenvalues of the second moment matrix:



Uniform Region

– gradients have small magnitude
– small λ_1, small λ_2
– system is ill-conditioned

Edge

– gradients have one dominant direction
– large λ_1, small λ_2
– system is ill-conditioned

High-texture or Corner Region

– gradients have different directions, large magnitudes
– large λ_1, large λ_2
– system is well-conditioned

Feature tracking

- So far, we have only considered feature tracking in a pair of images
- If we have more than two images, we can track feature from each frame to the next
- Given a point in the first image, we can in principle reconstruct its path by simply “following the arrows”

Tracking over Many Frames

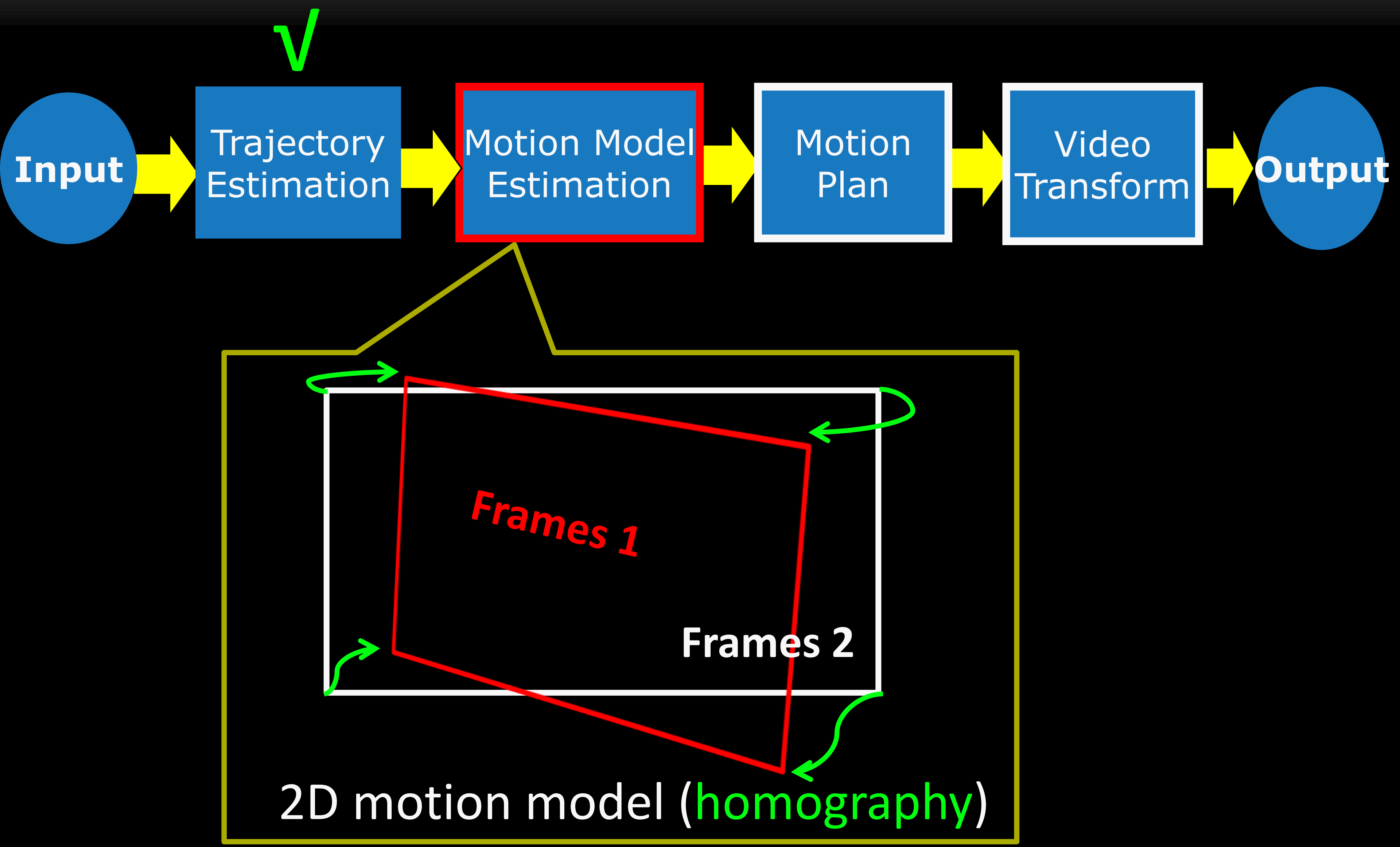
- Select features in first frame
- For each frame:
  - Update positions of tracked features
    - Discrete search or Lucas-Kanade (or a combination of the two)
  - Terminate inconsistent tracks
    - Compute similarity with corresponding feature in the previous frame or in the first frame where it's visible
  - Find more features to track

Shi-Tomasi Feature Tracker

- Find good features using eigenvalues of second-moment matrix
  - Key idea: "good" features to track are the ones whose motion can be estimated reliably
- From frame to frame, track with Lucas-Kanade
  - This amounts to assuming a translation model for frame-to-frame feature movement
- Check consistency of tracks by *affine* registration to the first observed instance of the feature
  - Affine model is more accurate for larger displacements
  - Comparing to the first frame helps to minimize drift

J. Shi and C. Tomasi. Good Features to Track. CVPR 1994.

Traditional 2D Video Stabilization



Homography

$$\lambda \mathbf{x}_i' = \mathbf{T}\mathbf{x}_i$$

expand

$$\lambda \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

Fitting a homography

- Equation for homography:

$$\lambda \begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

$$\lambda \mathbf{x}_i' = \mathbf{T}\mathbf{x}_i$$

$$\mathbf{x}_i' \times \mathbf{T}\mathbf{x}_i = 0$$

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{h}_1^T \mathbf{x}_i \\ \mathbf{h}_2^T \mathbf{x}_i \\ \mathbf{h}_3^T \mathbf{x}_i \end{bmatrix} = \begin{bmatrix} y_i' \mathbf{h}_3^T \mathbf{x}_i - \mathbf{h}_2^T \mathbf{x}_i \\ \mathbf{h}_1^T \mathbf{x}_i - x_i' \mathbf{h}_3^T \mathbf{x}_i \\ x_i' \mathbf{h}_2^T \mathbf{x}_i - y_i' \mathbf{h}_1^T \mathbf{x}_i \end{bmatrix}$$

$$\begin{bmatrix} 0^T & -\mathbf{x}_i^T & y_i' \mathbf{x}_i^T \\ \mathbf{x}_i^T & 0^T & -x_i' \mathbf{x}_i^T \\ -y_i' \mathbf{x}_i^T & x_i' \mathbf{x}_i^T & 0^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = 0$$

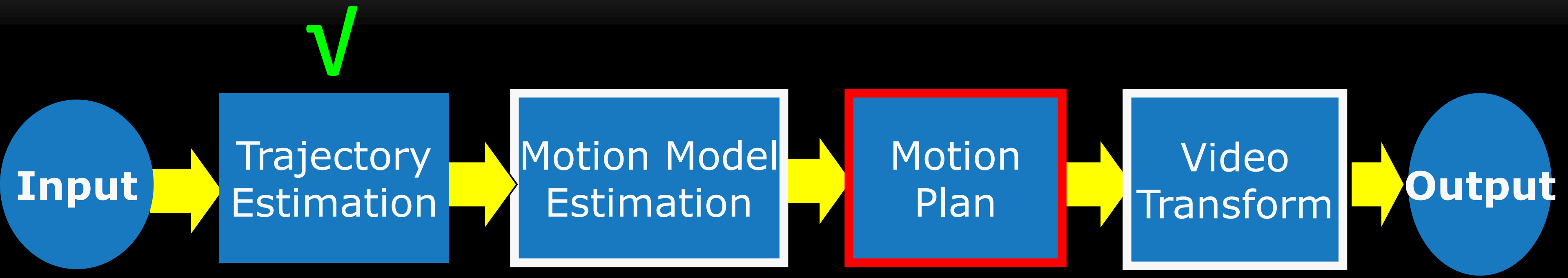
3 equations, only 2 linearly independent

Direct linear transform

$$\begin{bmatrix} 0^T & \mathbf{x}_1^T & -y_1'\,\mathbf{x}_1^T \\ \mathbf{x}_1^T & 0^T & -x_1'\,\mathbf{x}_1^T \\ \cdots & \cdots & \cdots \\ 0^T & \mathbf{x}_n^T & -y_n'\,\mathbf{x}_n^T \\ \mathbf{x}_n^T & 0^T & -x_n'\,\mathbf{x}_n^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = 0 \qquad \mathbf{A}\,\mathbf{h} = 0$$

- H has 8 degrees of freedom (9 parameters, but scale is arbitrary)
- One match gives us two linearly independent equations
- Four matches needed for a minimal solution (null space of 8x9 matrix)
- More than four: homogeneous least squares

Traditional 2D Video Stabilization



Motion Plan

$$\mathbf{S}_t = \sum_{i \in N_t} \mathbf{T}_t^i * G, \text{where } \mathbf{T}_t^i = \prod_{j=i}^{t} \mathbf{T}_j$$

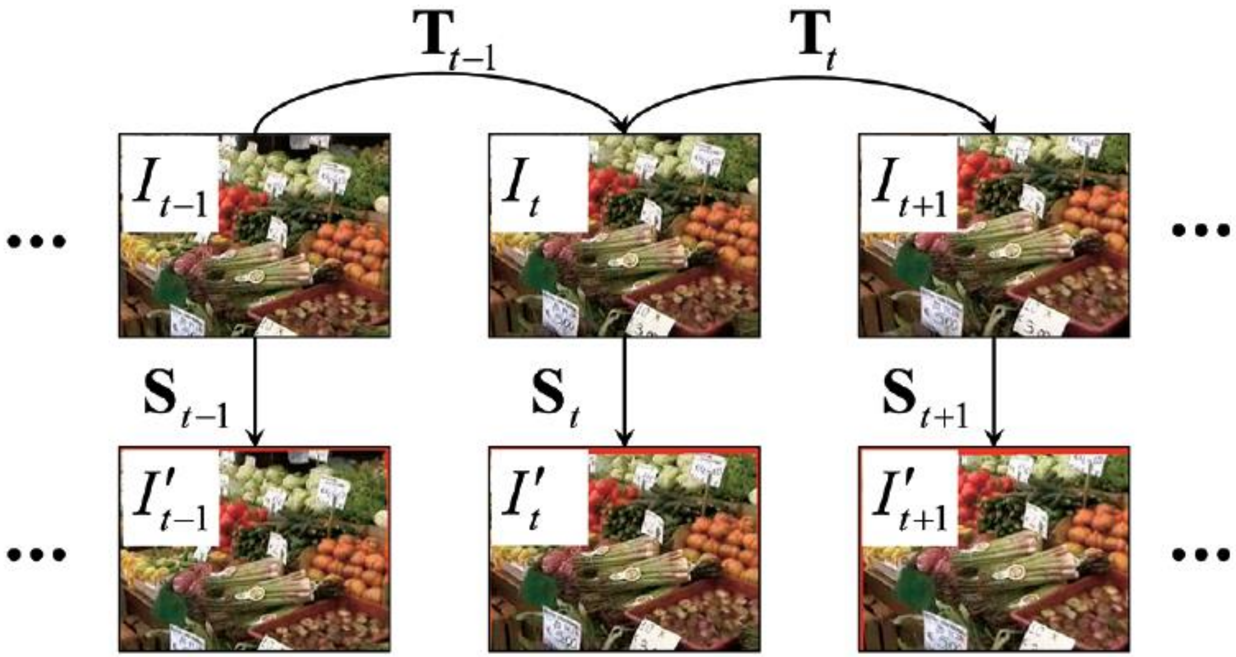


Image courtesy: Matsushita et al. 06

Traditional 2D Video Stabilization Result

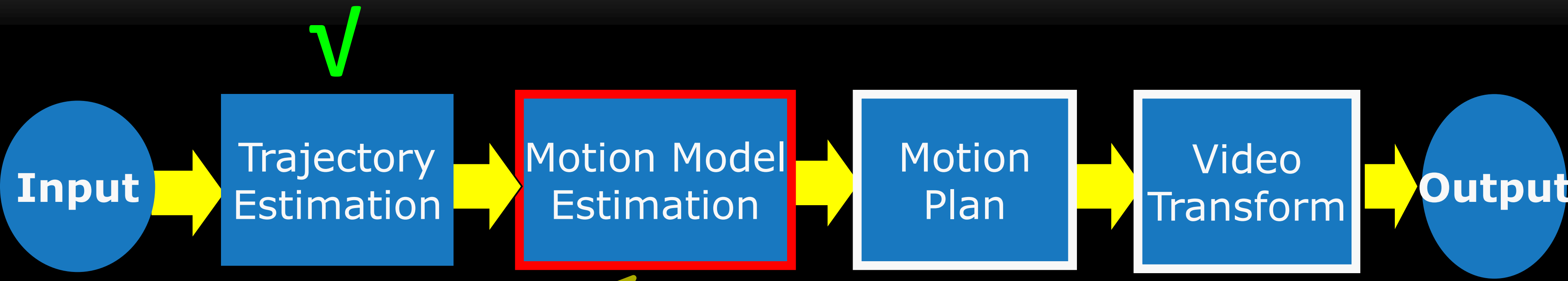
Limitations

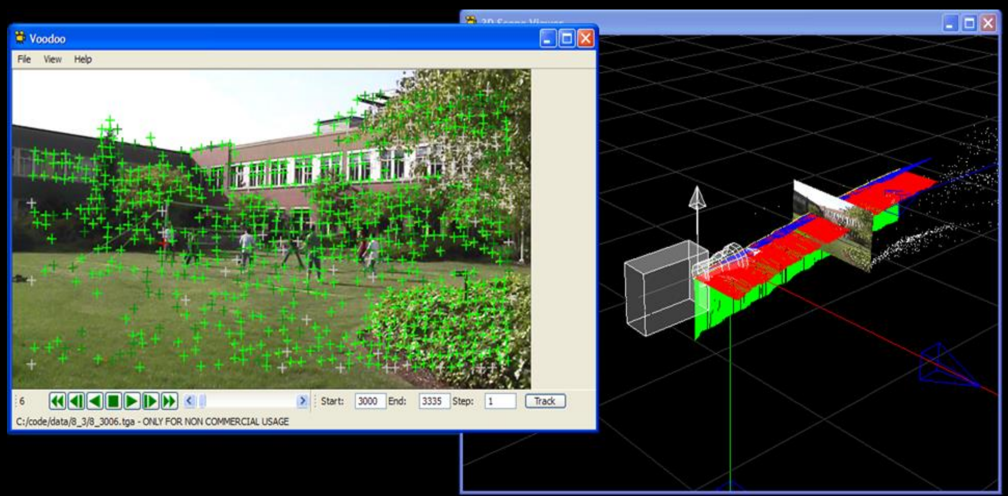
- No knowledge of actual 3D camera path, so cannot control desired motion directly
- Homography cannot model 3D camera motion and scene structure

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

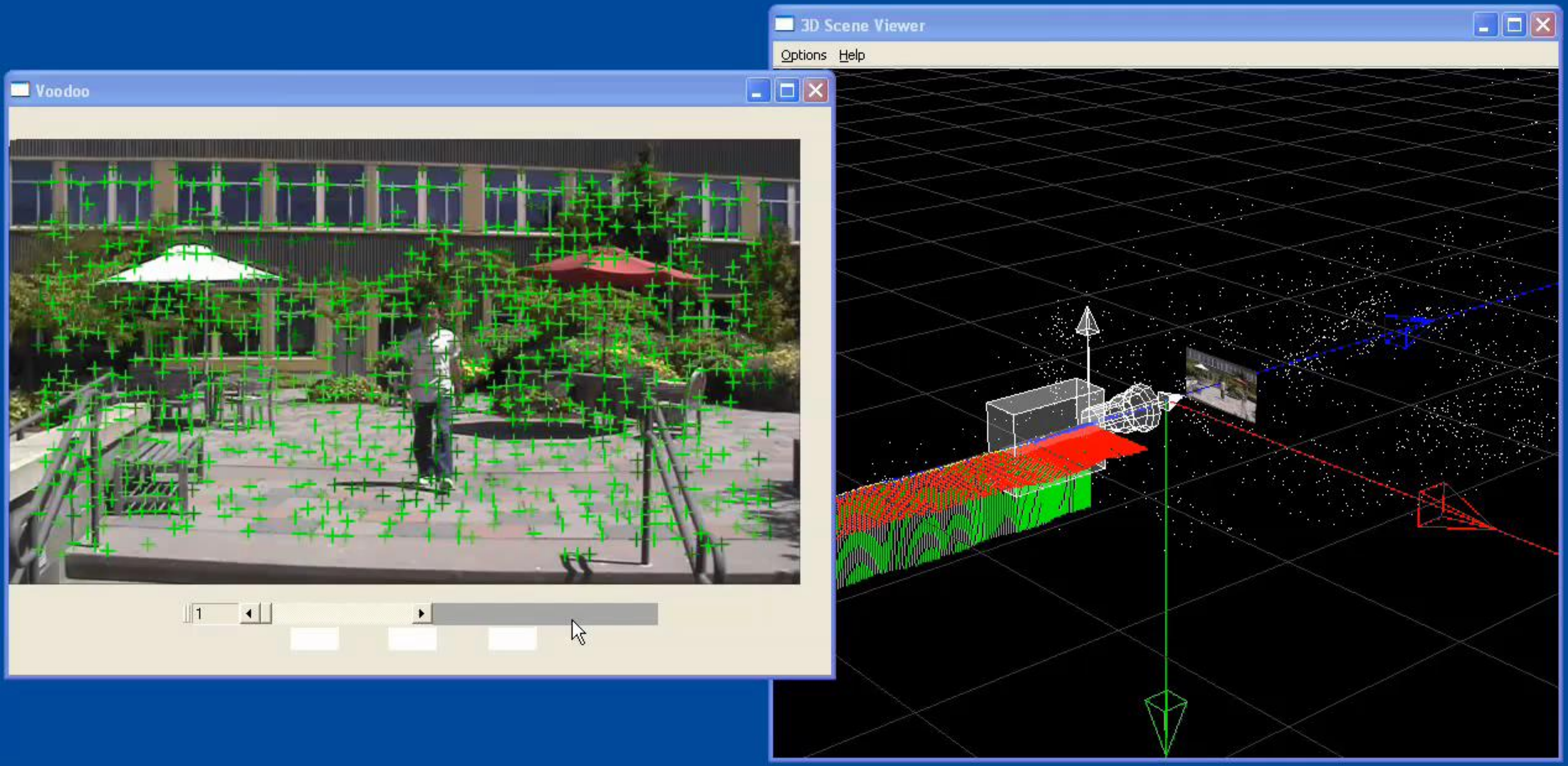




3D reconstruction via
structure from motion

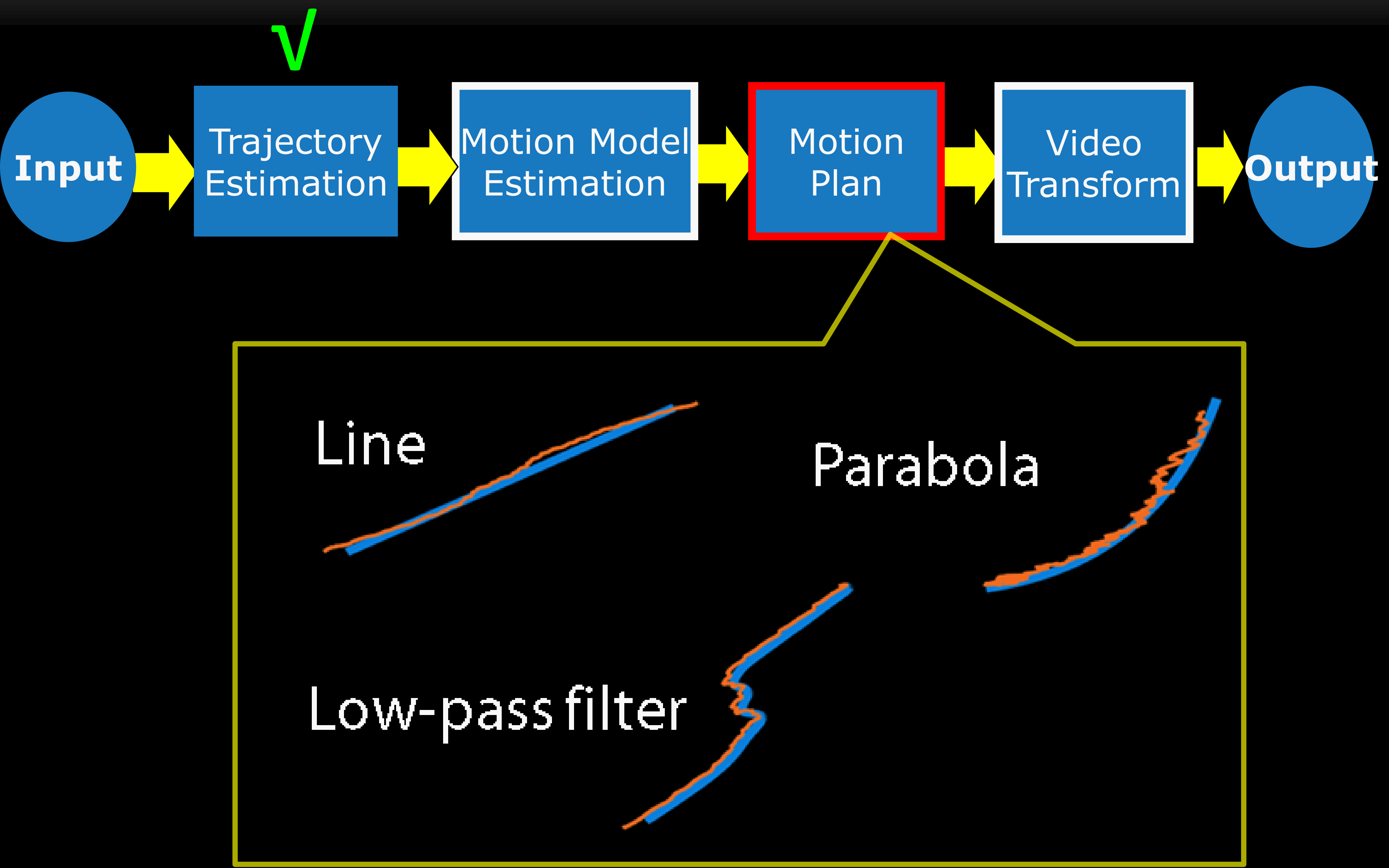
Voodoo Camera Tracker (http://www.digilab.uni-hannover.de)

Structure from Motion

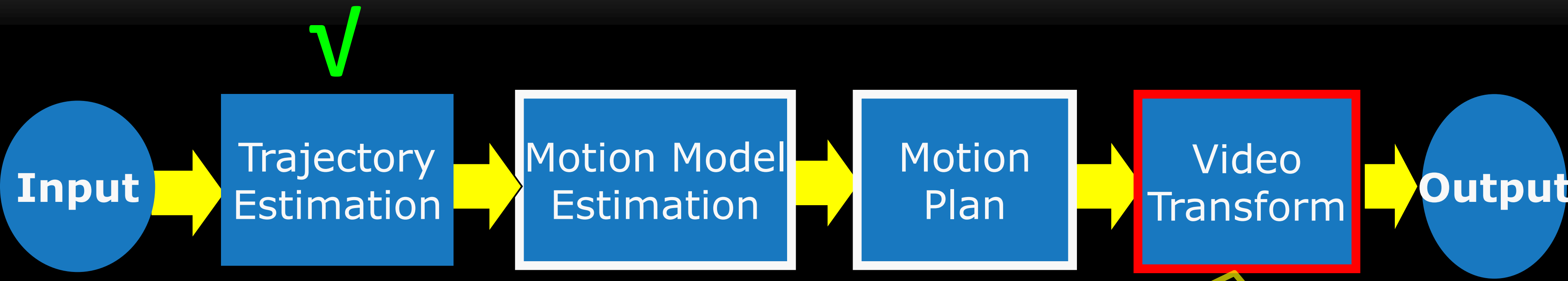


Voodoo Camera Tracker (http://www.digilab.uni-hannover.de)

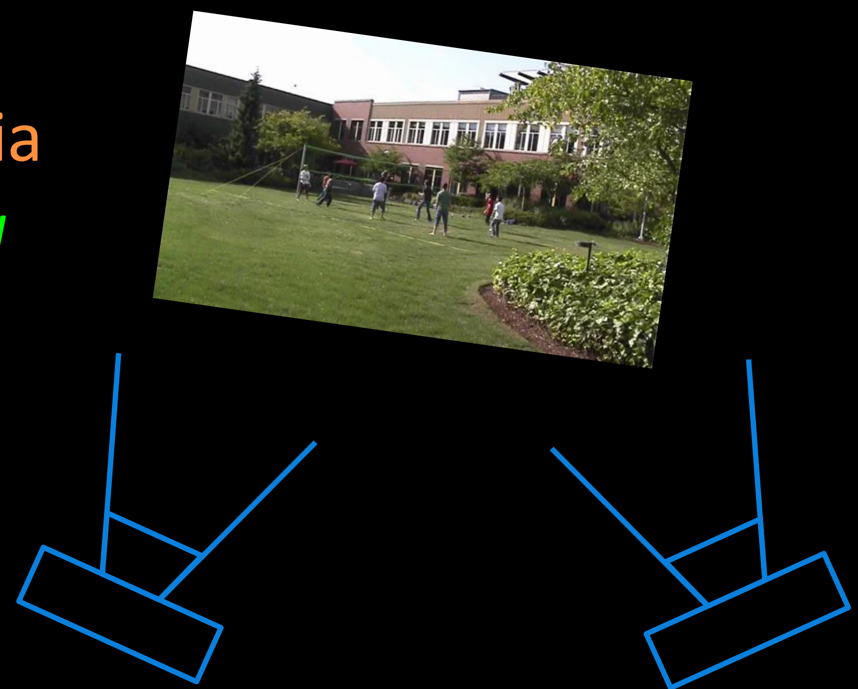
3D Video Stabilization



3D Video Stabilization



Novel view synthesis via *image based rendering*



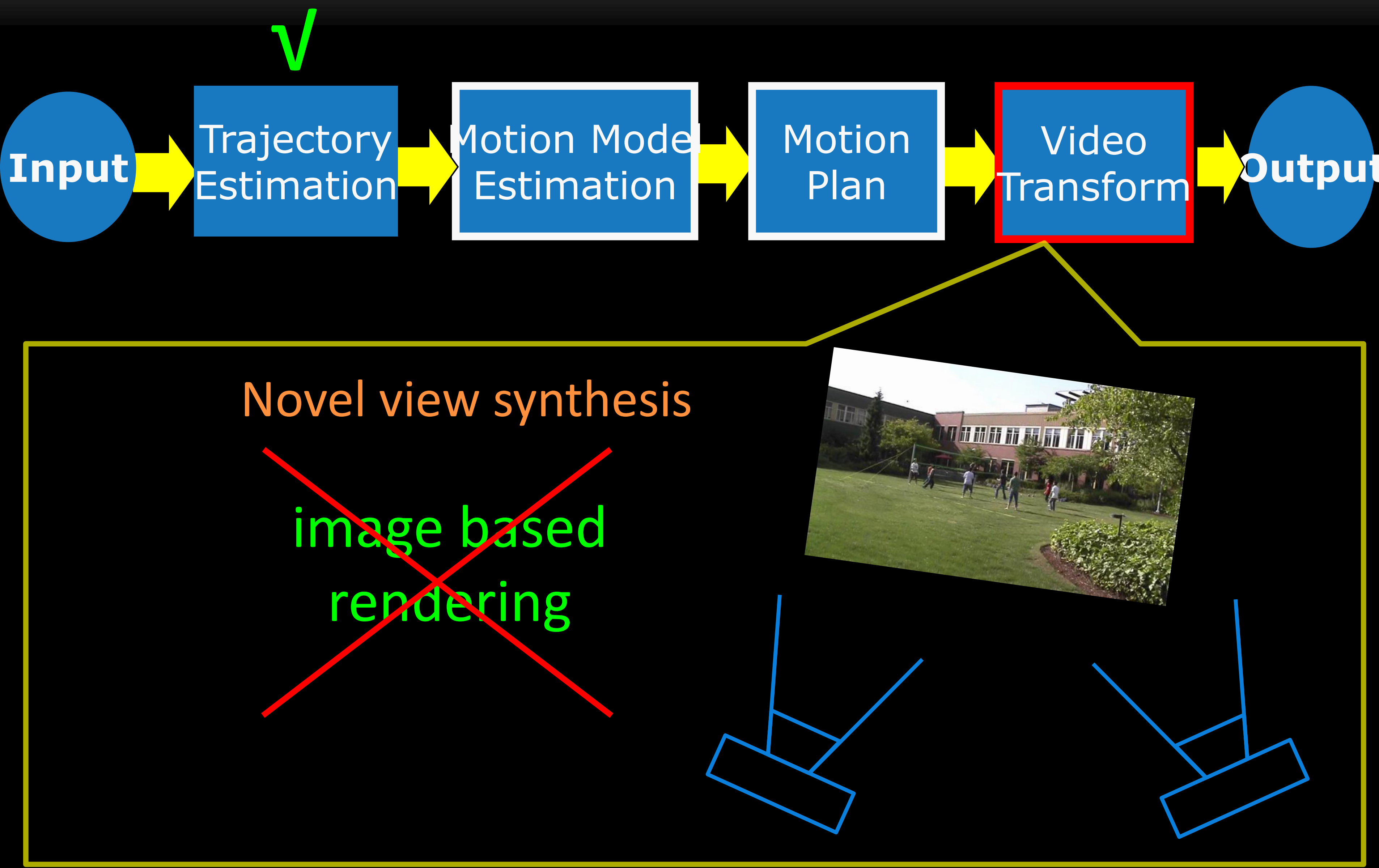
Novel View Synthesis by Image based Rendering

Unstructured lumigraph rendering [Buehler et al. 01]

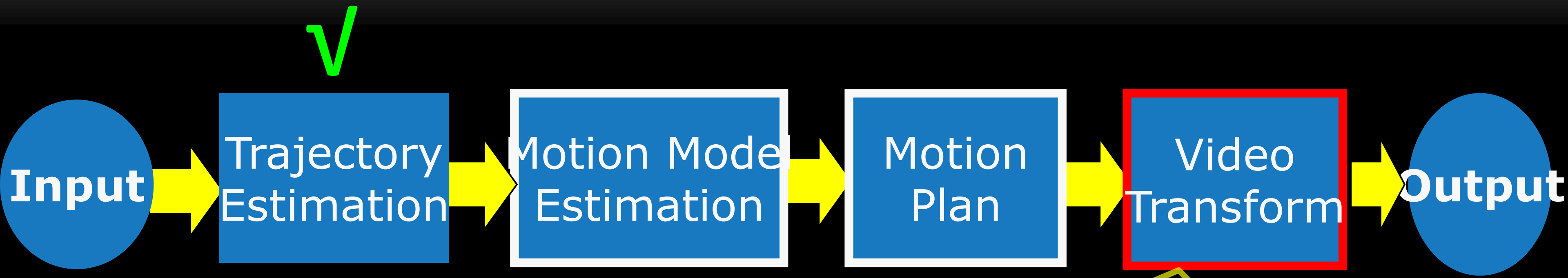
Content-preserving warps based 3D video stabilization

F Liu, M Gleicher, H Jin, A Agarwala. Content-preserving warps for 3D video stabilization, SIGGRAPH 2009

3D Video Stabilization



Temporal Constraint



Our method for novel view synthesis

One input frame

One output frame

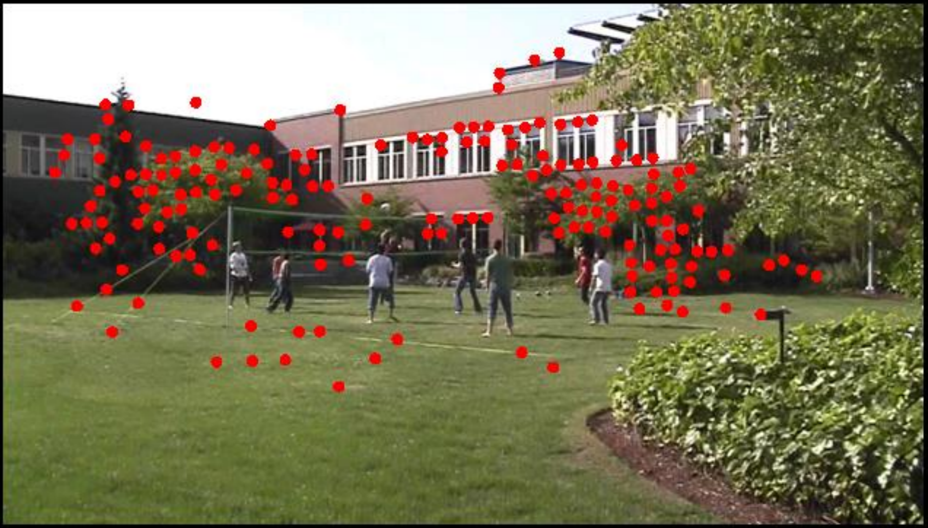
Novel View from One Frame

- 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.

Human Perception

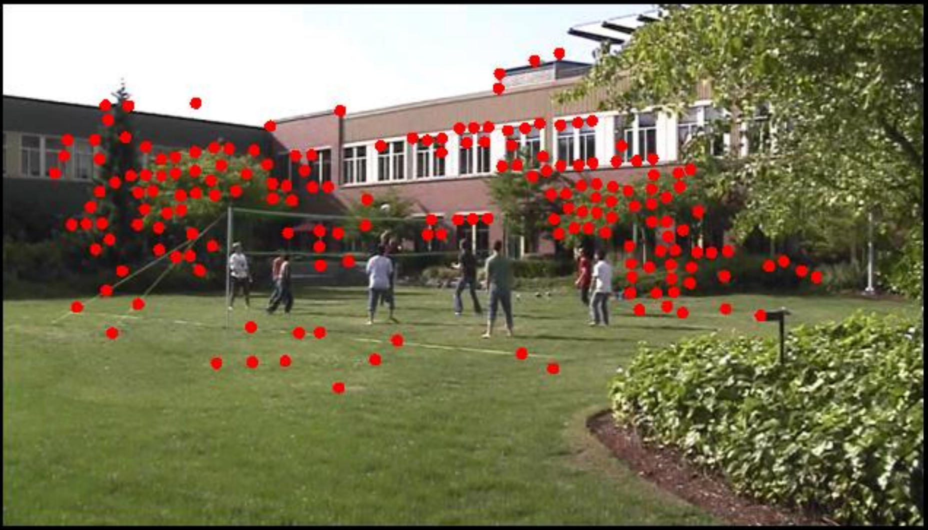
- Viewpoint shifts will be small
- Aim for perceptual plausibility rather than accurate novel view synthesis
  - Move salient content along stabilized paths
  - No noticeable artifacts

Problem Setup

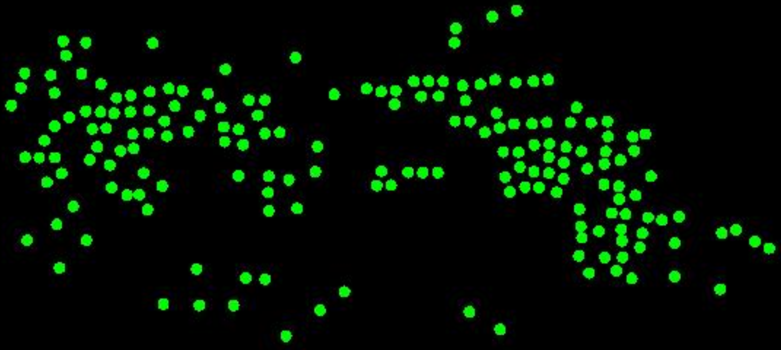


input frame and points

Problem Setup

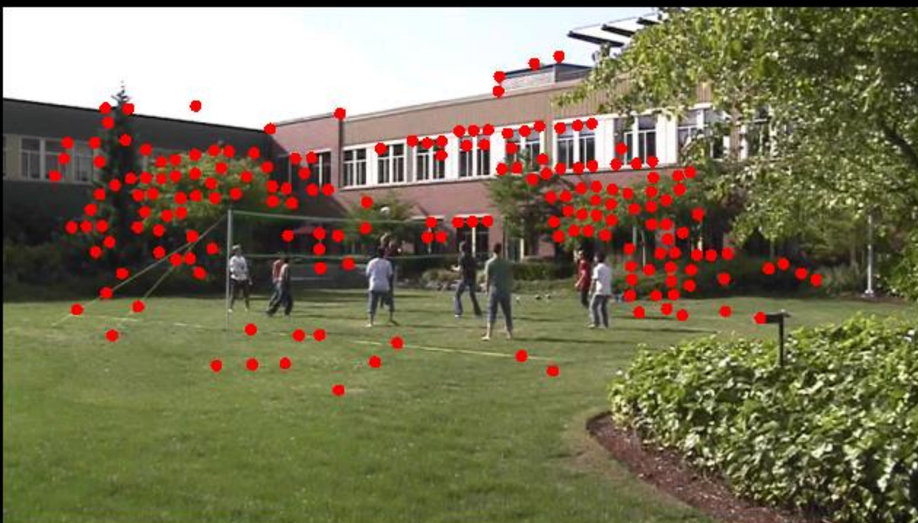


input frame and points



output points

Problem Setup



input frame and points

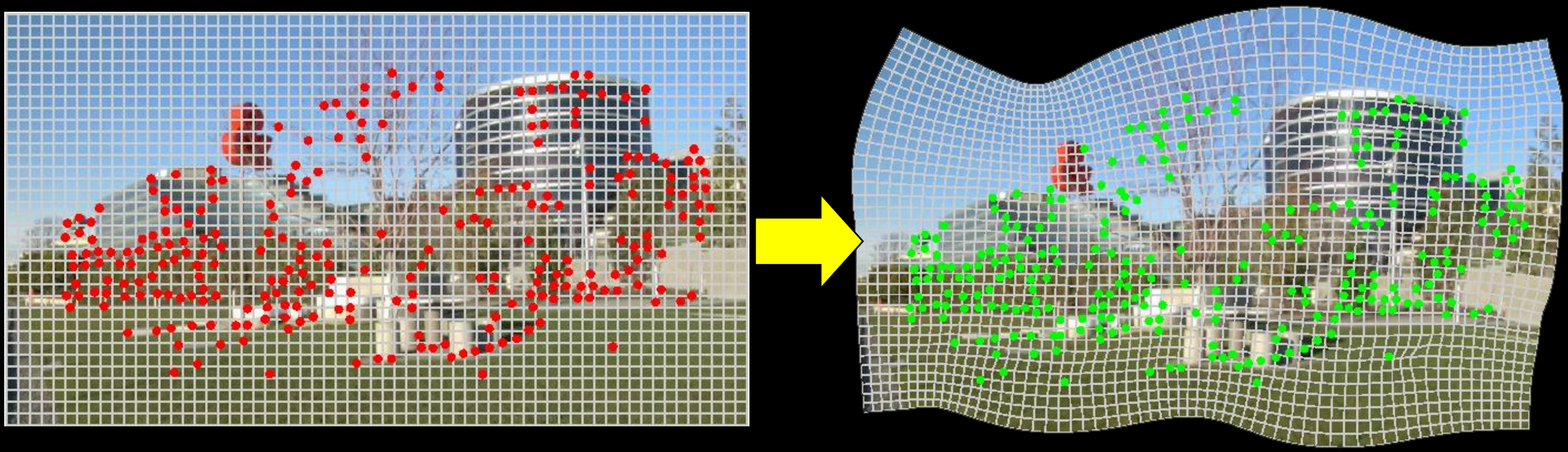
output frame

Option 1: Scattered Data Interpolation

Option 2: Full-frame Warping with Homography

A Less Successful Result

Our Approach: Content-preserving Warping

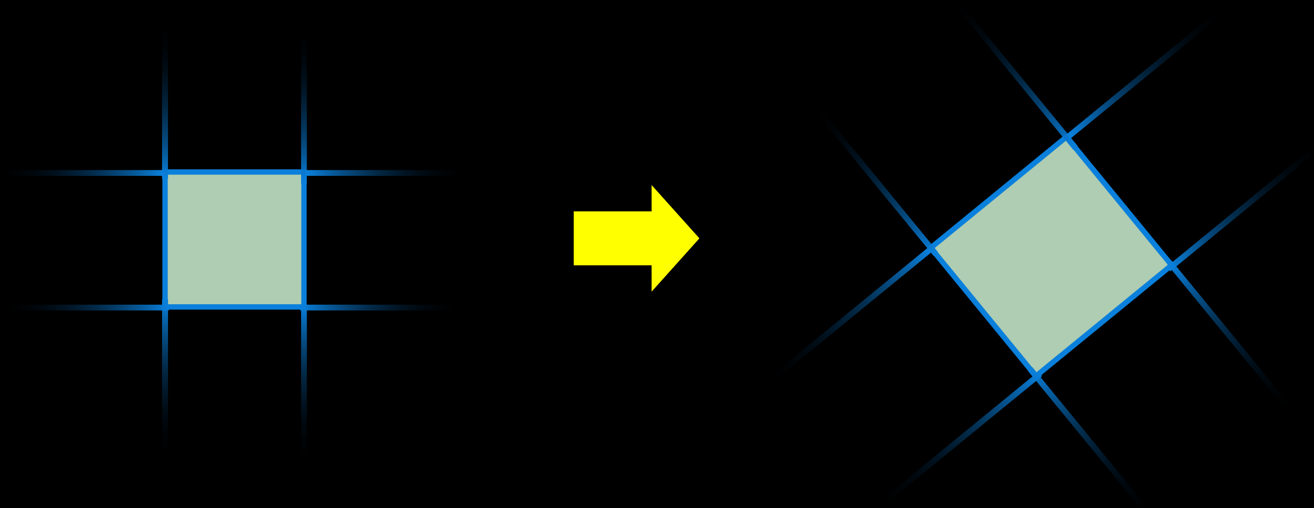


Warp each input frame to create the output frame by least-squares minimization

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

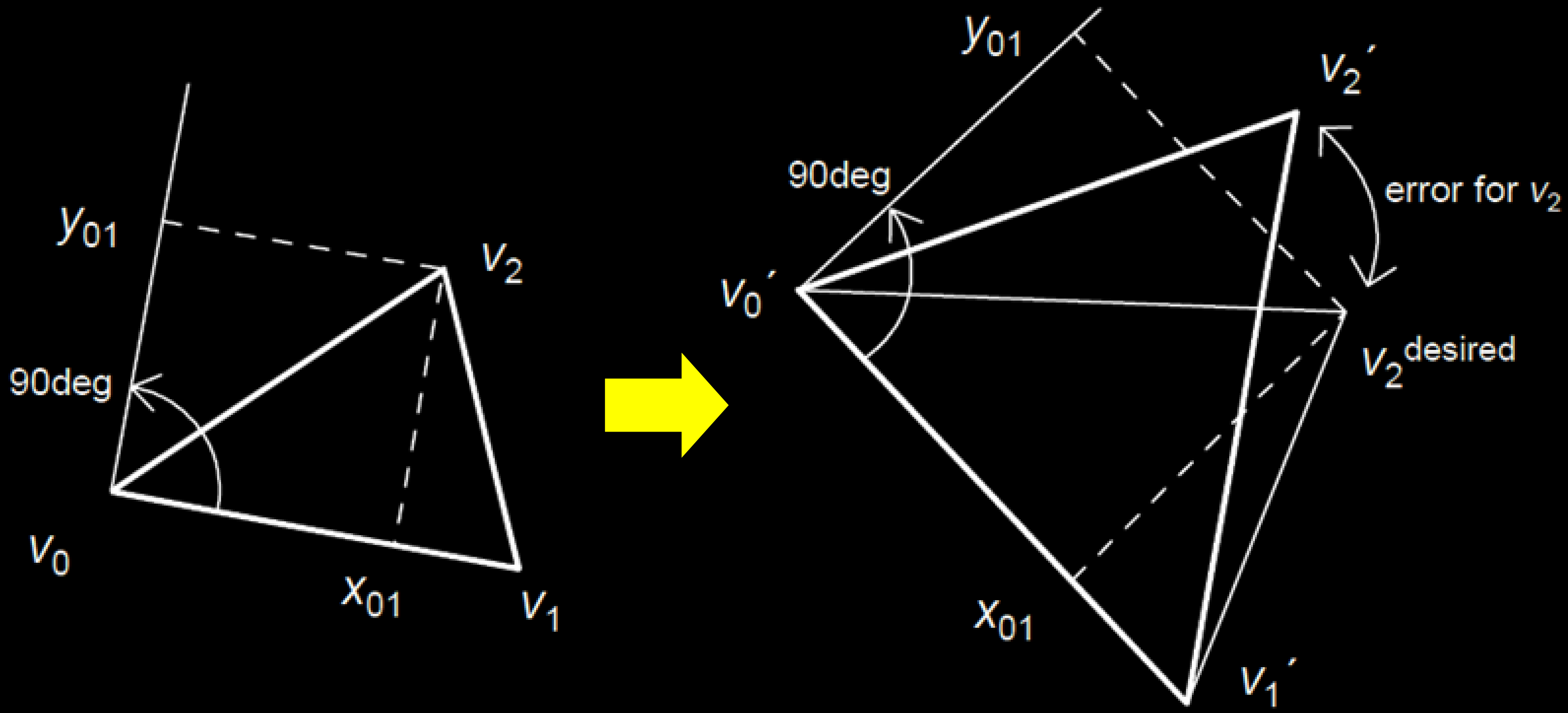
Smoothness Term: Minimize Visual Distortion

Local similarity transformation constraint



Smoothness Term: Minimize Visual Distortion

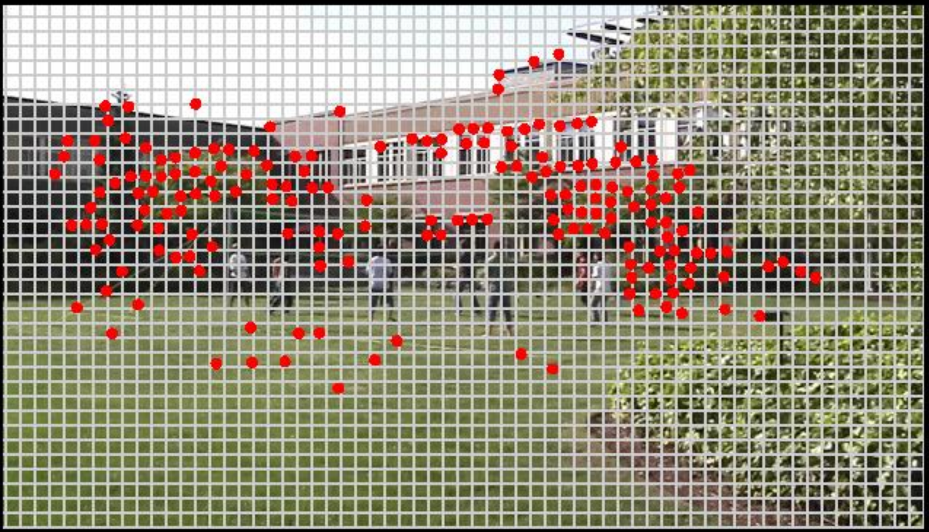
Local similarity transformation constraint



[Igarashi et al. 05]

Saliency Weight

Concentrate distortion to non-salient regions



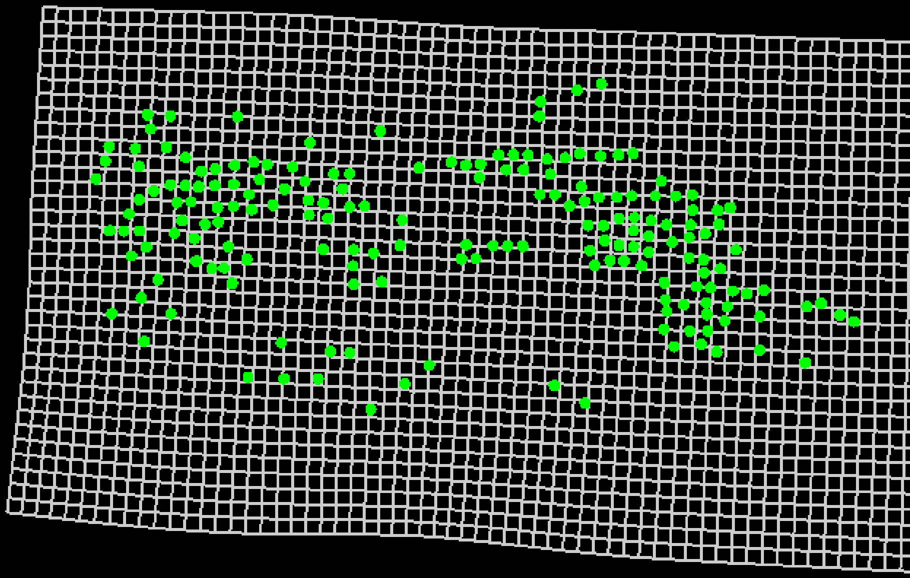
Input

Visual saliency map
[Itti et al. 99]

Visual saliency: "the distinct subjective perceptual quality which makes some items in the world stand out from their neighbors and immediately grab our attention" from [Itti 07]

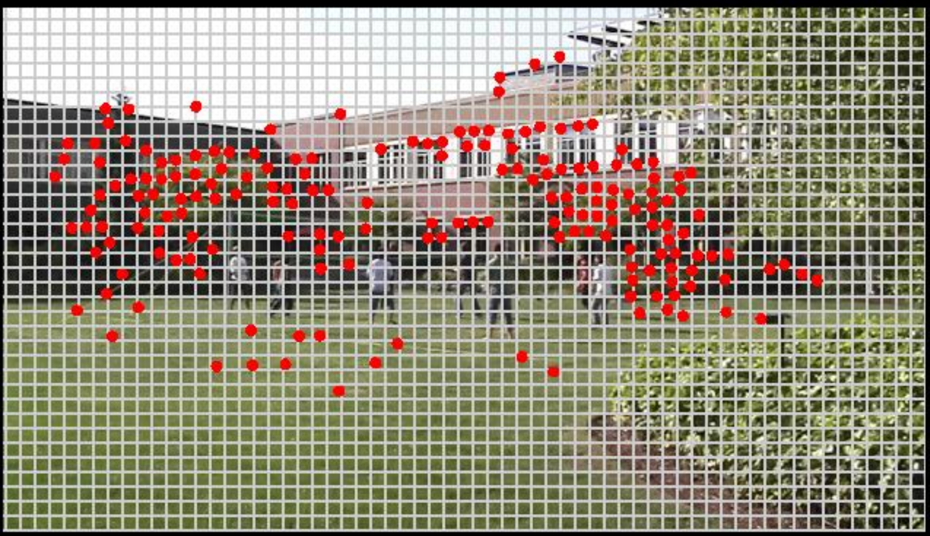
Content-Preserving Warping

Input



Output

Content-Preserving Warping



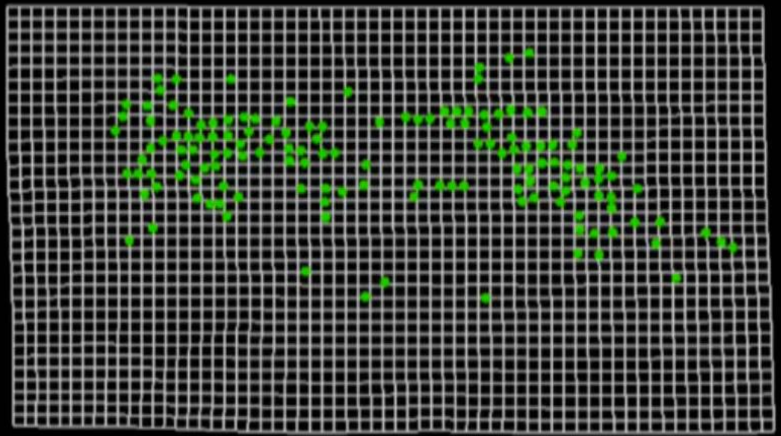
Input

Output

texture mapping [Shirley et al. 2005]

Content-Preserving Warping

Grid mesh
& points



Output

Student paper presentation

Poisson Image Editing

P. Pérez, M. Gangnet, and A. Blake
SIGGRAPH 2003

Presenter: Rojas, Casey

Student paper presentation

Intelligent Scissors for Image Composition

E. Mortensen and W. Barrett
SIGGRAPH 1995

Presenter: Smith, Cassaundra

Next Time

- Video stabilization II
- Student paper presentation
  - 05/17: Wiemholt, Cody
    - Video SnapCut: Robust Video Object Cutout Using Localized Classifiers
      X. Bai, J. Wang, D. Simons, G. Sapiro
      SIGGRAPH 2009
  - 05/17: Zwovic, Kitt
    - A global sampling method for alpha matting
      K. He, C. Rhemann, C. Rother, X. Tang, and J. Sun
      CVPR 2011