Constructing spherical panoramas of a bladder phantom from endoscopic video using bundle adjustment

Timothy D. Soper, John E. Chandler, Michael P. Porter, Eric J. Seibel

SPIE Medical Imaging 2011,
Visualization, Image-Guided Procedures, and Modeling
Tuesday February 15th, 2011

UNIVERSITY OF WASHINGTON

HPL Human Photonics Laboratory

Bladder Surveillance

- Bladder Cancer:
  - 5th most prevalent in U.S.[1]
  - 70,530 new cases yearly
  - Recurrence 50%
- Flexible Cystoscopy
  - Annual, semi-annual, or quarterly surveillance
  - ~5mm diameter
  - Local anesthesia
- Rigid Cystoscopy
  - Biopsy or resection
  - General anesthesia

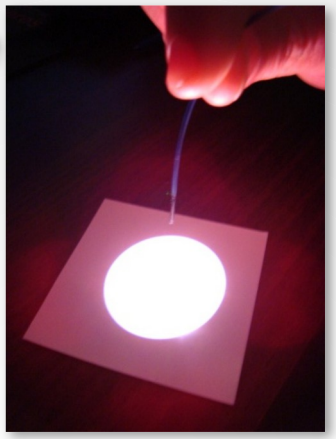
www.nhs.uk

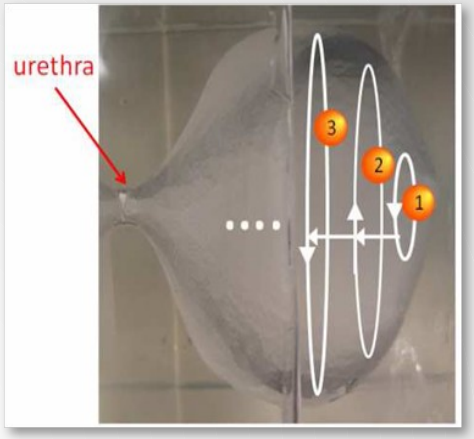
[1] Jemal, A., Siegel, R., Xu, J. *et al.*, "Cancer statistics, 2010," CA Cancer J Clin, 60(5), 277-300 (2010).

Surveillance

Scanning Fiber Endoscope

- Ultrathin (1.5 mm)
- Highly flexible
- High Resolution (500 x 500)
- Full-Color
- 30 Hz video frame rate





Yoon, Sangtae, Reinhall, *et al.*, 2009

Advancement

- Patient comfort
- Laser diagnostics/therapies
- Automation

Proposed Model

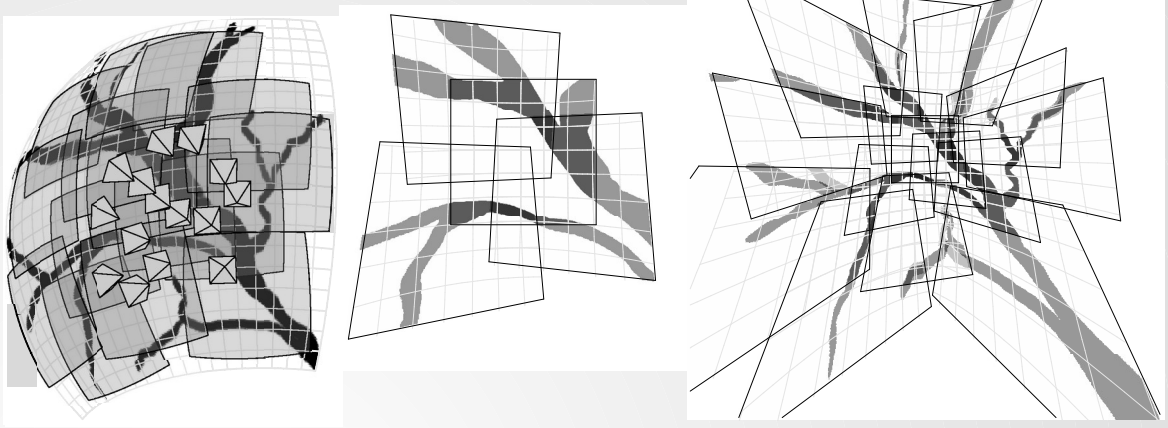
- Replace urologist with nurse/technician
- Implement automated scanning device
- **Urologist performs post-procedural review**

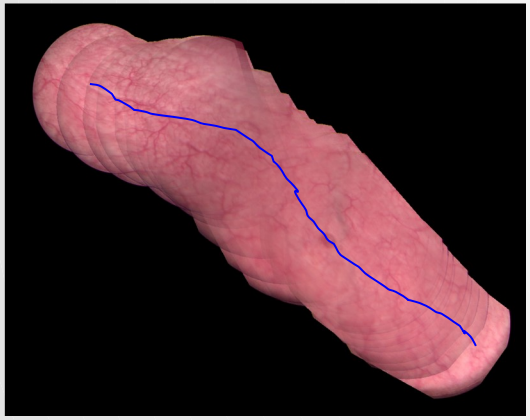
Panorama

Image Alignment

Challenges:

- Presumed planar geometry
- Global alignment
- Precalibration



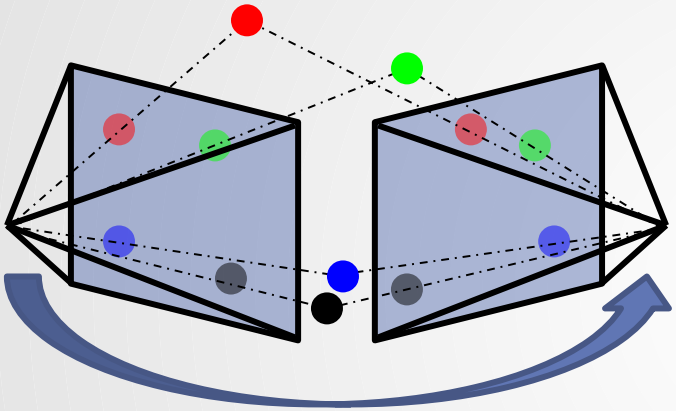


Requirements for full panoramic reconstruction :

- Nonplanar
- Free moving endoscope
- 360° field of view
- Global alignment
- Self calibrating
- Not real time

Structure from Motion

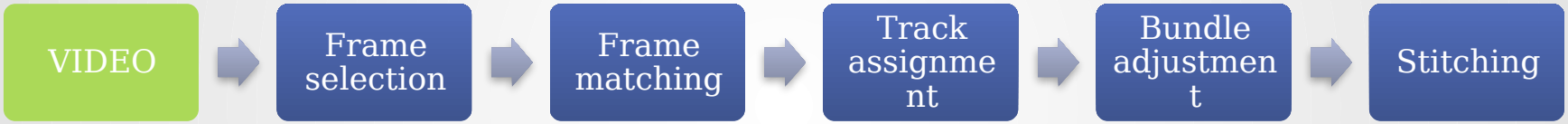
Simultaneous recovery of scene geometry and camera motion



Virtual Tourism[1]

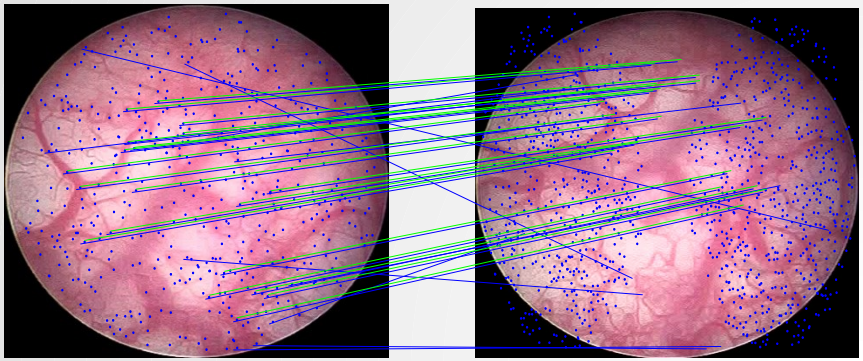
[1]Snavely, N., Seitz, S. M., and Szeliski, R., ACM Transactions on Graphics, *2006*.

Software

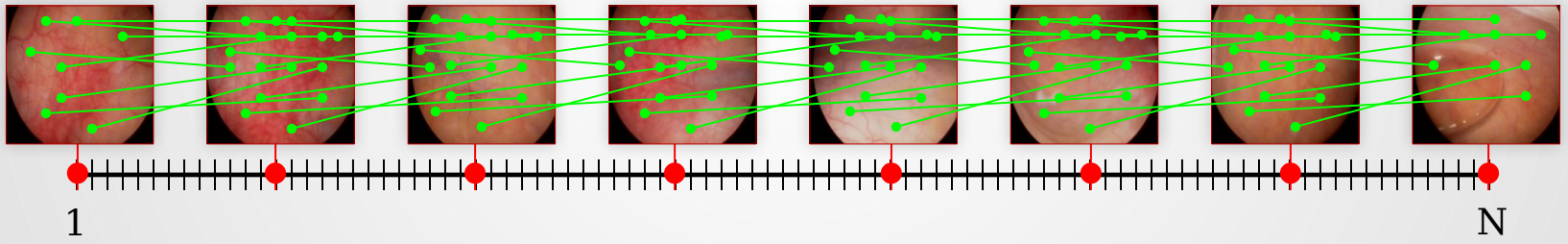

VIDEO
Frame selection
Frame matching
Track assignment
Bundle adjustment
Stitching

Frame Selection

VIDEO → Frame selection → Frame matching → Track assignment → Bundle adjustment → Stitching



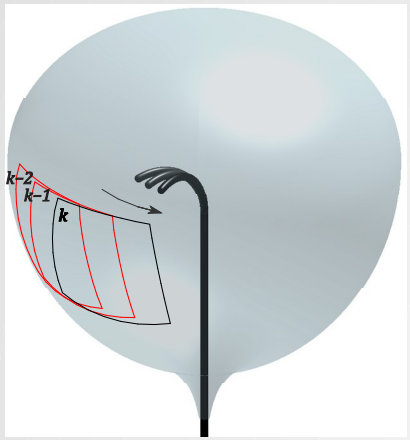
1. Scale Invariant Feature Transform[1]
2. Feature matching
3. Random sample concensus (RANSAC)
4. Sparse selection



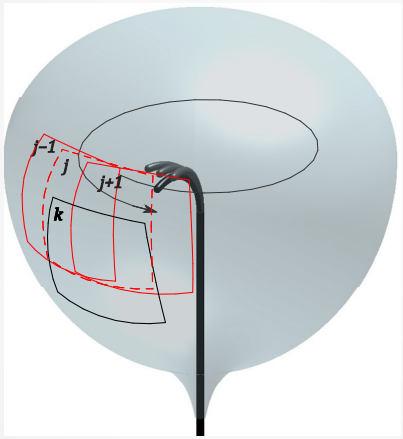
[1]Lowe, International Journal of Computer Vision, 2004

Frame Matching

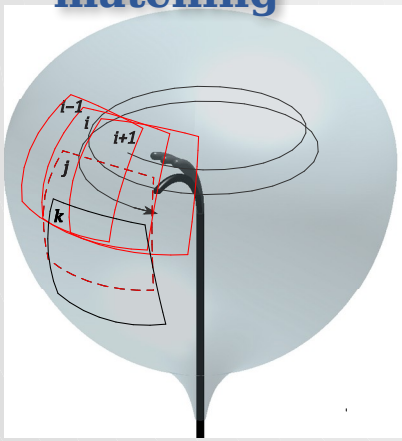
VIDEO → Frame selection → Frame matching → Track assignment → Bundle adjustment → Stitching



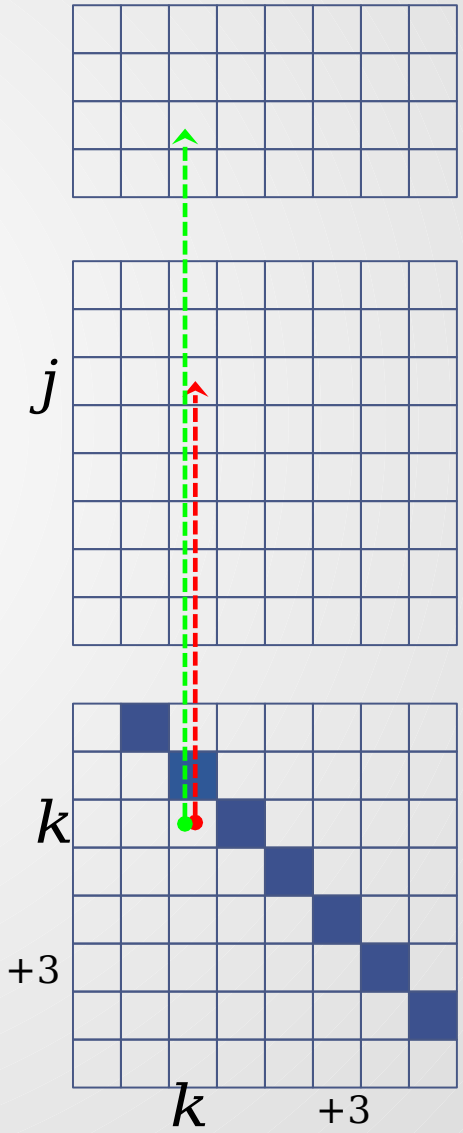
Sequential matching



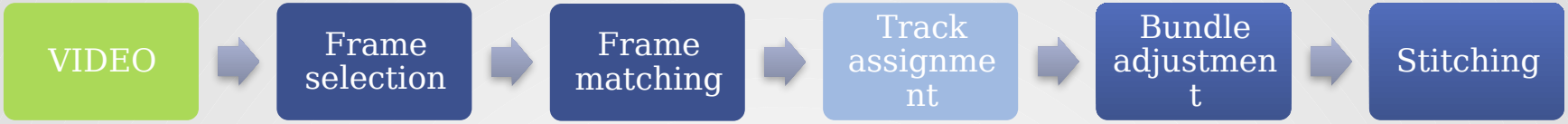
Nonsequential matching



Associated matching



Track Assignment



For all features consistently matched between 3+ frames:

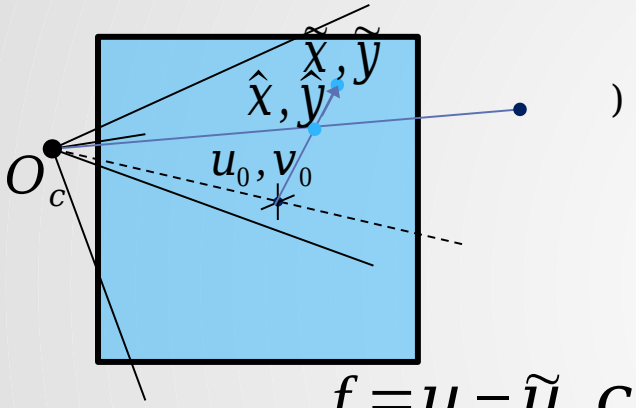
assign as track

$$X = [X\ Y\ Z]$$

Bundle Adjustment

VIDEO → Frame selection → Frame matching → Track assignment → Bundle adjustment → Stitching

Non-linear least squares minimization between the *observed* and *predicted* pixel positions:



)

Camera Equations:

$$[X_c\ Y_c\ Z_c]^T = R_c[X\ Y\ Z]^T + t_c \quad \text{(3D rigid body)}$$

(projective)

$$\tilde{x} = x\left(1+\kappa_1 r^2+\kappa_2 r^4\right)$$
$$\tilde{y} = y\left(1+\kappa_1 r^2+\kappa_2 r^4\right) \quad \text{(radial distortion)}$$

$$f = u - \tilde{u}, c = f^T f$$

$$\tilde{u} = f_x \tilde{x} + u_0, \tilde{v} = f_y \tilde{y} + v_0 \quad \text{(camera equations)}$$

$3M$ $6N$ 6

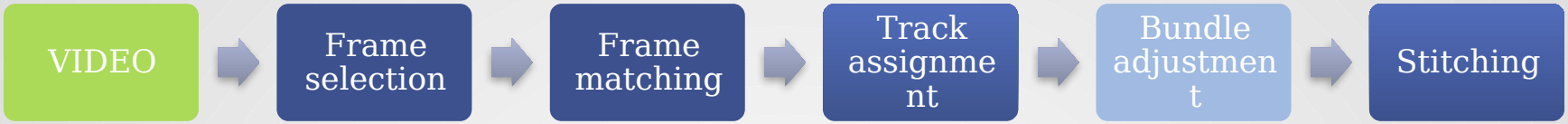
State variable:

Update: Levenberg-Marquardt

Large parameter sets solved by exploiting sparsity!

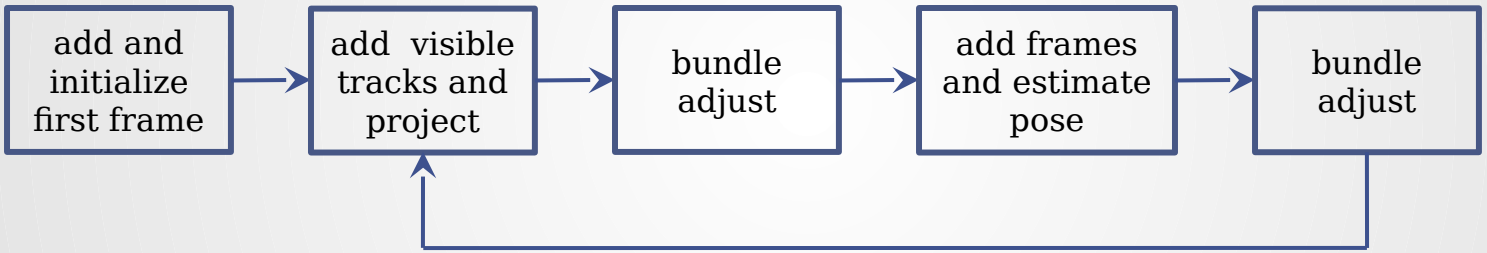
- 3D point parameters and camera parameters do not combine
- Each camera sees only a subset of points

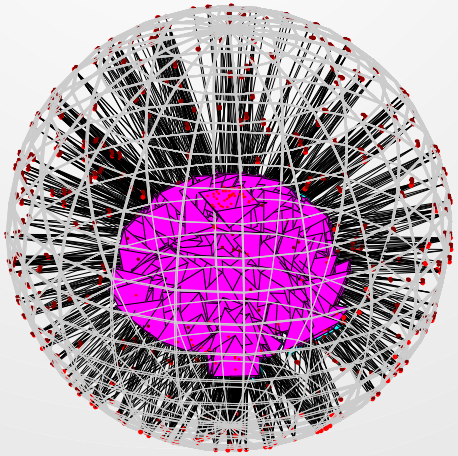
Bundle Adjustment



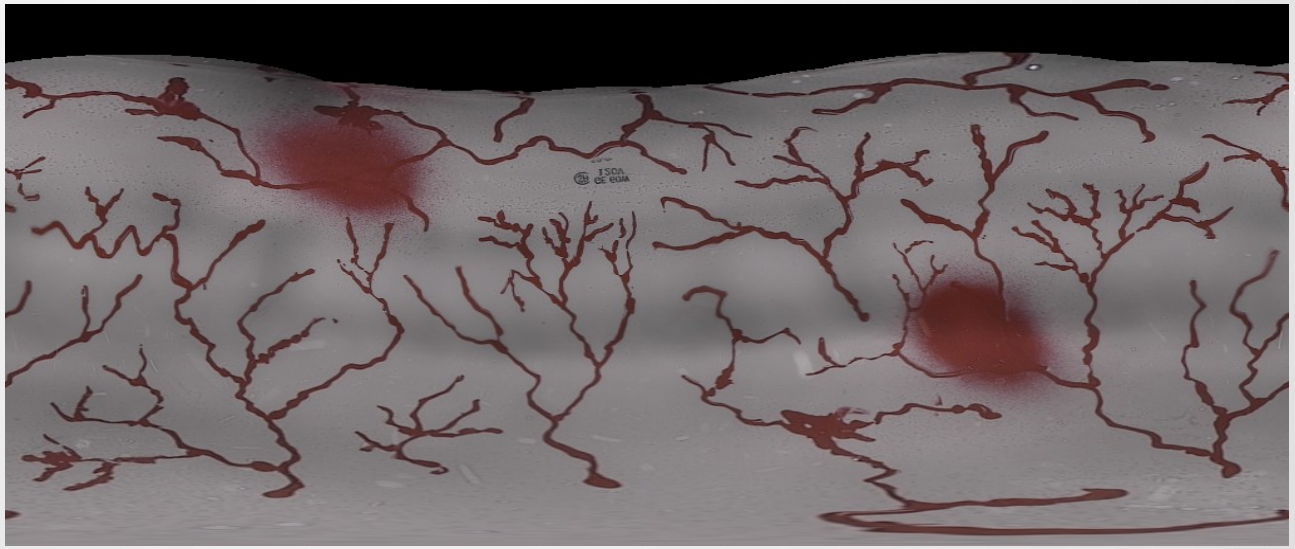
Implementation:

- Constrain mosaic to spherical surface $(X,Y,Z) \rightarrow (\theta,\phi,1)$
- Constrain frames to lie within sphere.
- Incremental adjustment



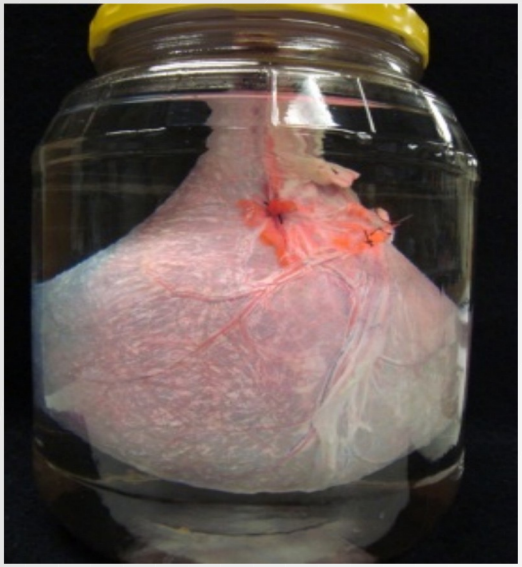


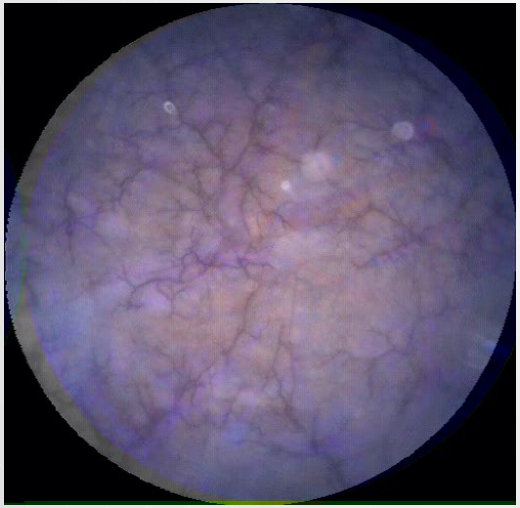
Experiment 1: Bladder Phantom

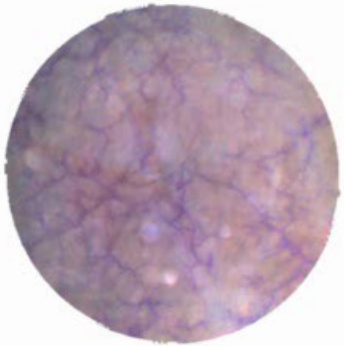


- 240 frames
- Projection error = 0.47 pixels

Bladder

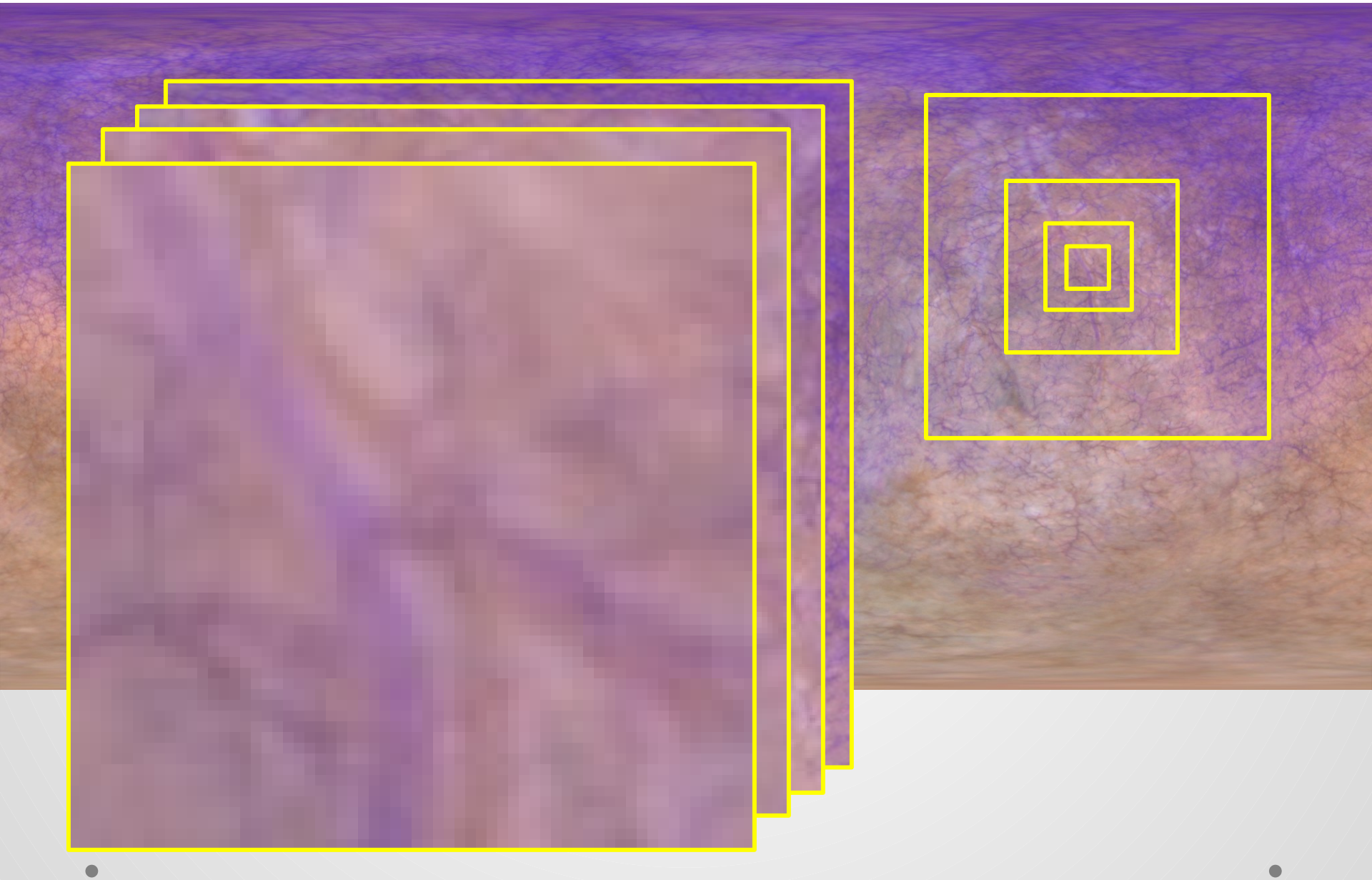






- 857 frames
- Projection error = 1.41 pixels

Experiment 2: Reconstruction 2D



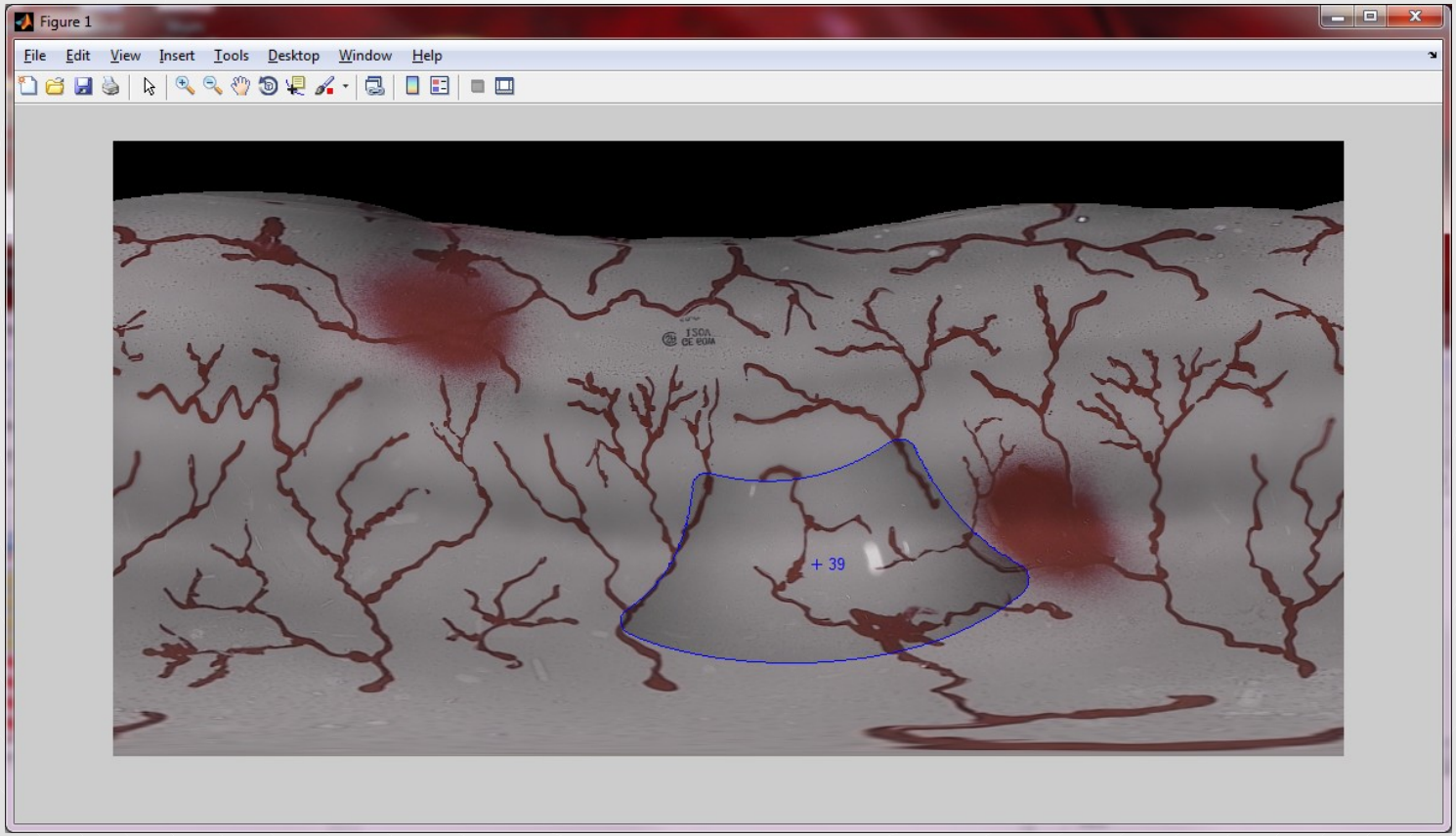
Conclusions

- Several hundred frames aligned with pixel level accuracy.
- No calibration necessary
- Spherical model proved adequate

- Sufficient overlap is essential!

Future Work

- Validate automated scanning mechanism
- Improve blending, deghosting
- Develop interface



Acknowledgements

Human Photonics Laboratory

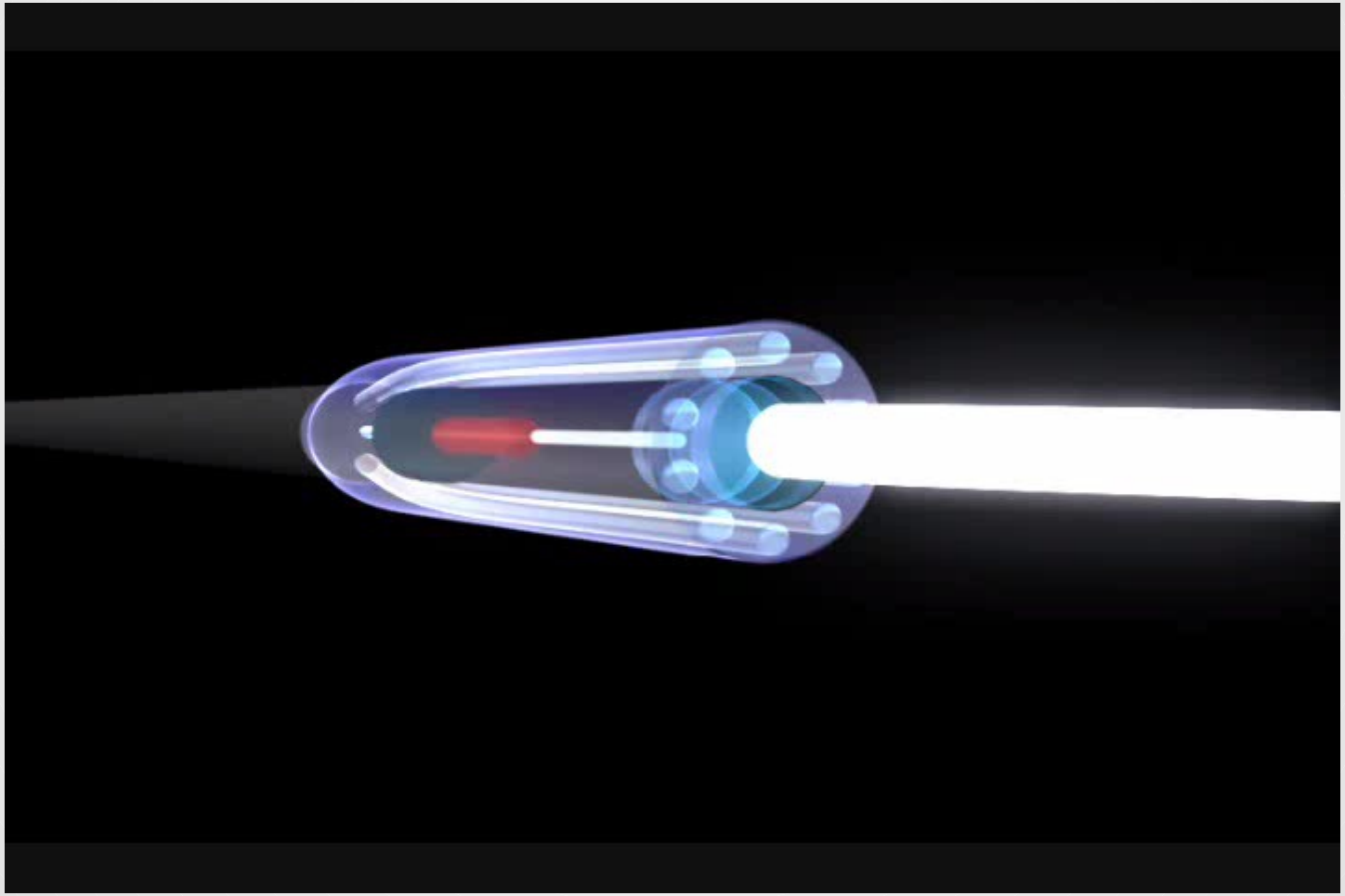
- Rich Johnston
- Charles Melville
- Cameron Lee

Funding

- Coulter Foundation & University of Washington Department of Bioengineering
- University of Washington Center for Commercialization

email: tsoper@uw.edu

Scanning Fiber Endoscope (SFE)



mation by Mr. Duff Hendrickson, Seattle, WA, copyright University of Washingt

Full Panoramic Reconstruction

Reconstruction Metrics:

- Single mosaic of entire bladder
- No presumed shape
- Global alignment
- No Calibration

Structure from Motion:

Simultaneously recover scene geometry and camera motional

Virtual Tourism[1]

$$M = \underset{N}{number\ of\ points}$$

$$number\ of\ parameters = 3M + 6N + 6$$

$$(X, Y, Z)$$

$$(\theta, \phi, \rho, t_x, t_y, t_z)$$

$$(f_x, f_y, u_0, v_0, K_1, K_2)$$

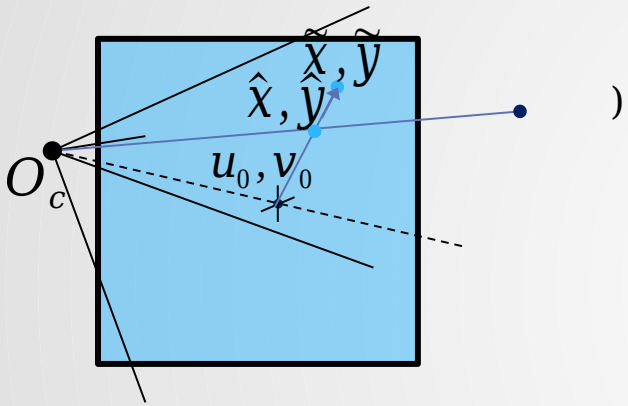
[1]Snavely, N., Seitz, S. M., and Szeliski, R., ACM Transactions on Graphics, *2006*.

Bundle Adjustment

VIDEO → Feature detection → Frame matching → Track assignment → Bundle adjustment → Stitching

Non-linear least squares minimization between the observed and predicted pixel positions:

$$f = u - \tilde{u}, c = f^T f$$



)

Camera Equations:

$$[X_c\ Y_c\ Z_c]^T = R_c [X\ Y\ Z]^T + t_c$$ (3D rigid body)

(projective)

$$\tilde{x} = x\left(1 + \kappa_1 r^2 + \kappa_2 r^4\right)$$
$$\tilde{y} = y\left(1 + \kappa_1 r^2 + \kappa_2 r^4\right)$$ (radial distortion)

$$\tilde{u} = f_x \tilde{x} + u_0, \tilde{v} = f_y \tilde{y} + v_0$$ (camera equations)

State variable:

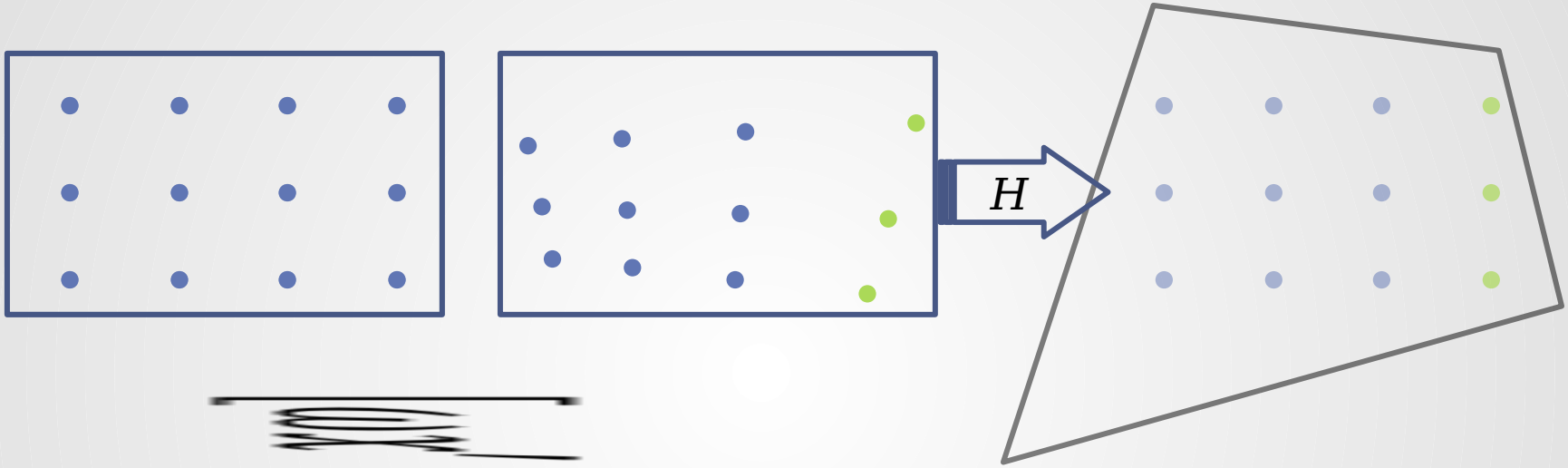
$$\Delta \mathbf{x} = -\left(H + \lambda \cdot diag(H)\right)^{-1} \nabla c(\mathbf{x})$$ Levenberg-Marquardt

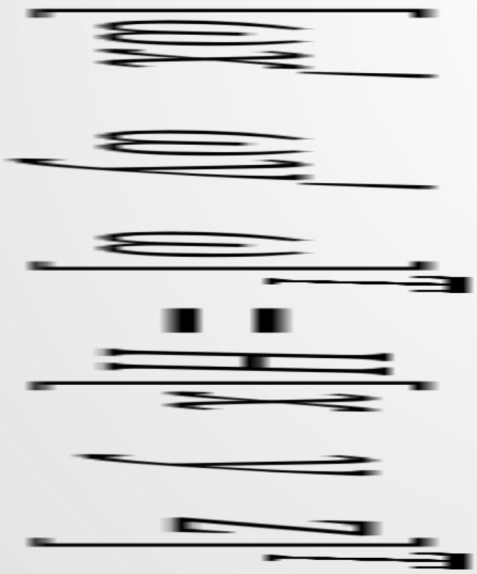
Large parameter sets solved by exploiting sparsity!

- 3D point parameters and camera parameters do not combine
- Each camera sees only a subset of points

Bladder

Planar Homography: $\begin{bmatrix} \omega x' & \omega y' & \omega \end{bmatrix}^T = H \begin{bmatrix} x & y & z \end{bmatrix}^T$



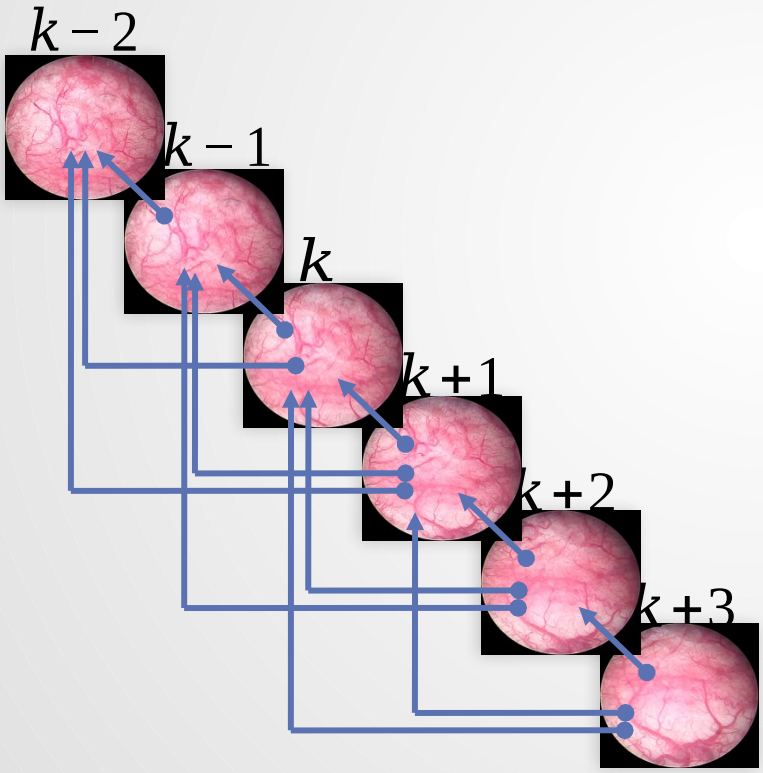


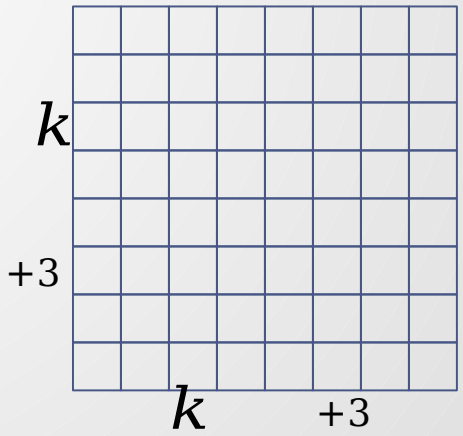
Limitations:

- Assumes planar scene
  - misalignment
  - pixel stretching
- Limited FOV
- Lacks global alignment
- Requires precalibration

Frame Matching

VIDEO → Feature detection → Frame matching → Track assignment → Bundle adjustment → Stitching

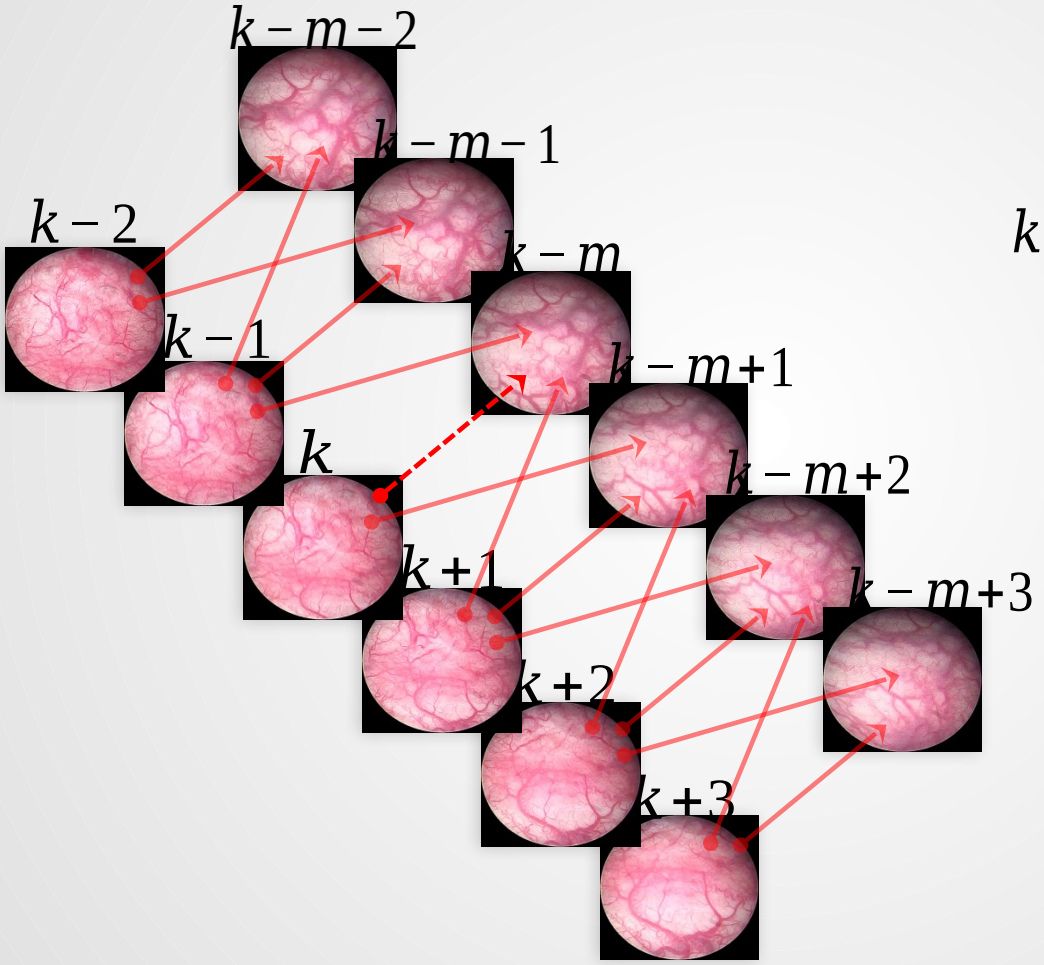


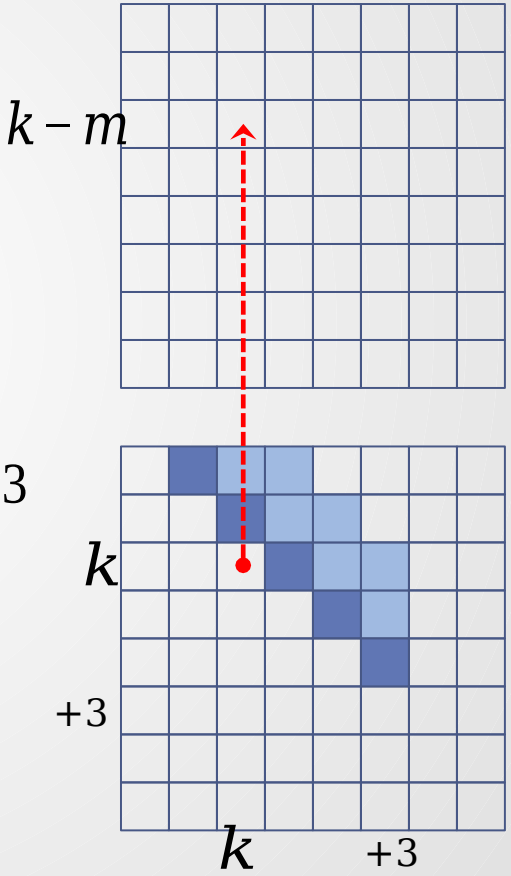


Sequential frame matching

Frame Matching

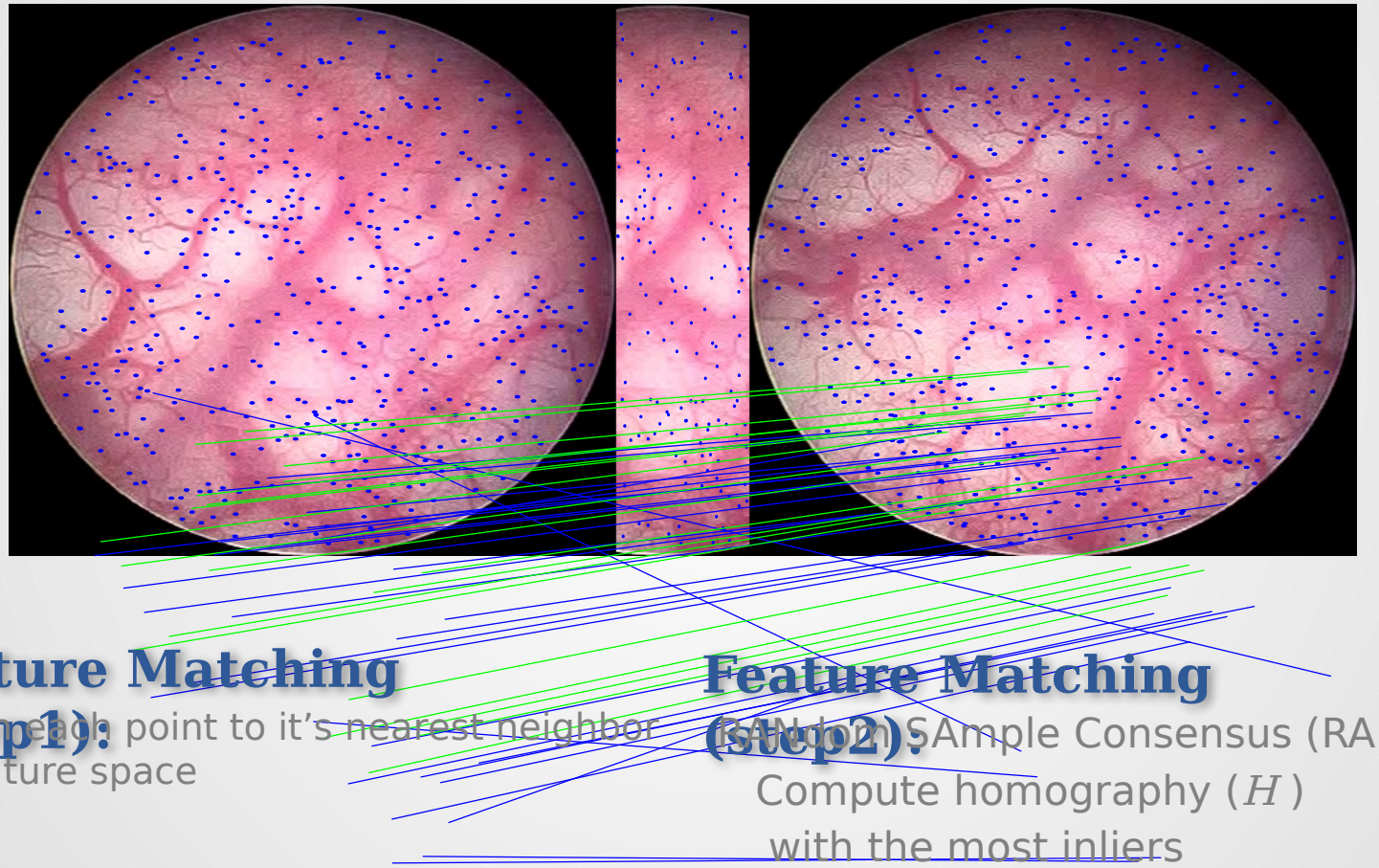
VIDEO → Feature detection → Frame matching → Track assignment → Bundle adjustment → Stitching





Nonsequential frame matching

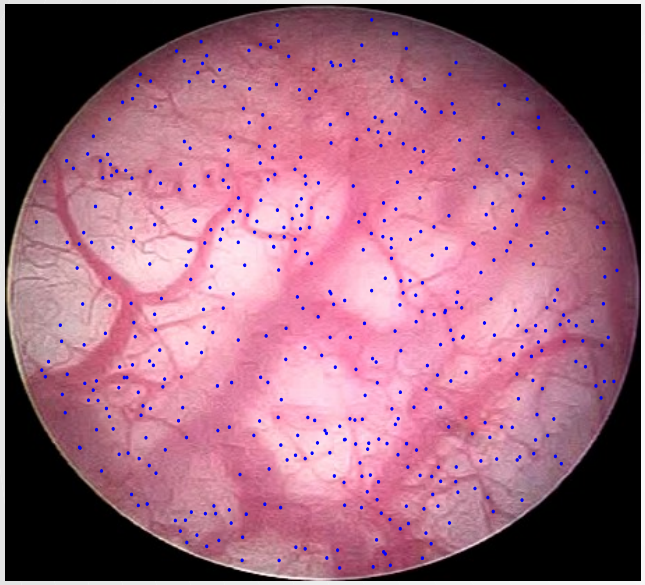
Frame Matching



Feature Matching (step1):
Match each point to it's nearest neighbor in feature space

Feature Matching (step2):
RANdom SAmple Consensus (RANSAC):
Compute homography (H)
with the most inliers

Feature Detection

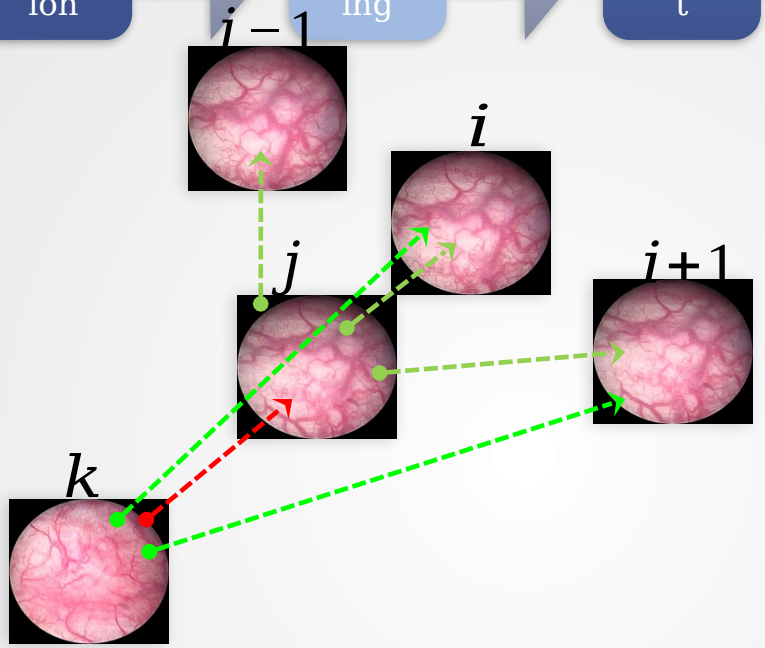


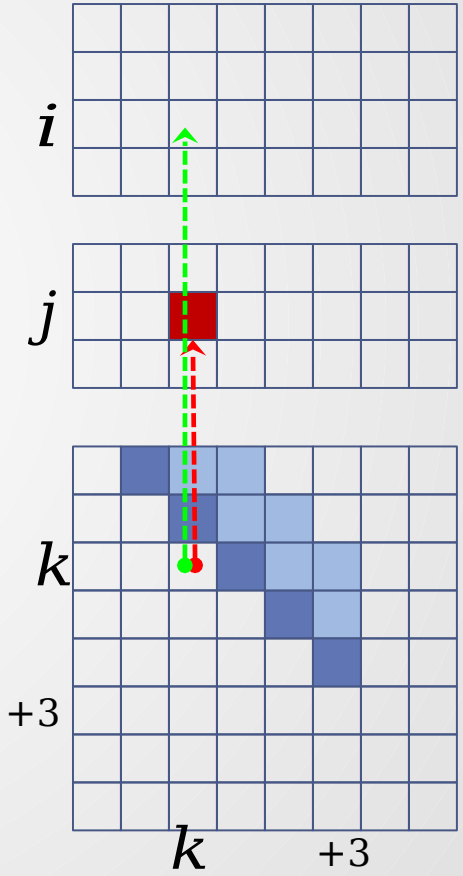
Scale Invariant Feature Transform (SIFT)

- Scale, rotation invariant[1]
- Keypoints
- Descriptors

[1]Lowe, International Journal of Computer Vision, 2004

Frame Matching





Associated frame matching