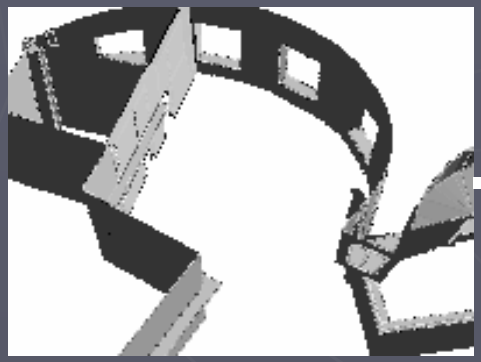
Wide-Area Localization from Omnidirectional Video and Known 3D Structure

Olivier Koch
koch@mit.edu

Seth Teller
teller@csail.mit.edu

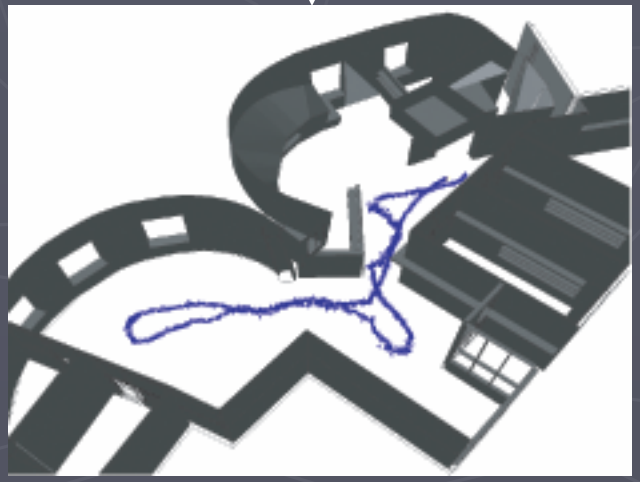
http://rvsn.csail.mit.edu/omni3d

Problem Statement



3D MAP

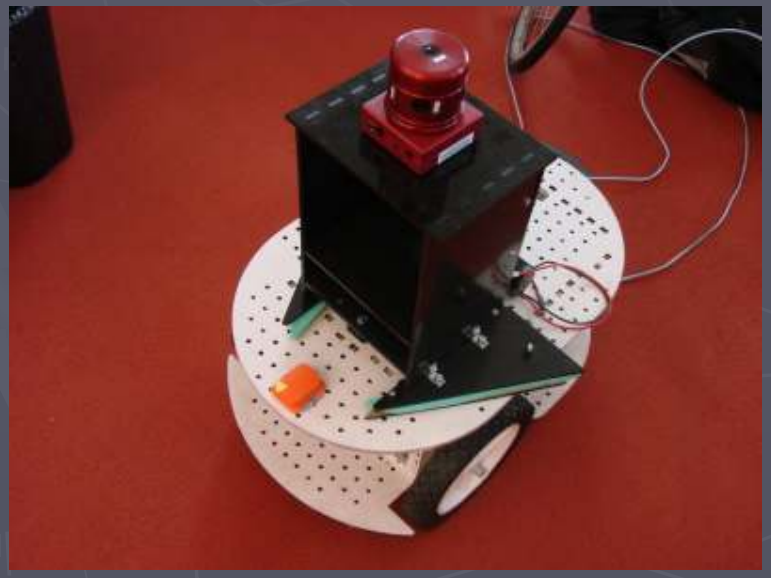
OMNIDIRECTIONAL VIDEO



ACCURATE 6-DOF LOCALIZATION

The device

- PointGrey Research Ladybug Camera
- Field of view: ~75% of full sphere
- 6 x 1024 x 768 8-bit JPG images @ 15Hz



Approach

- Match 3D model line *segments* with 2D image *edges* (model-image *correspondences*).

Approach

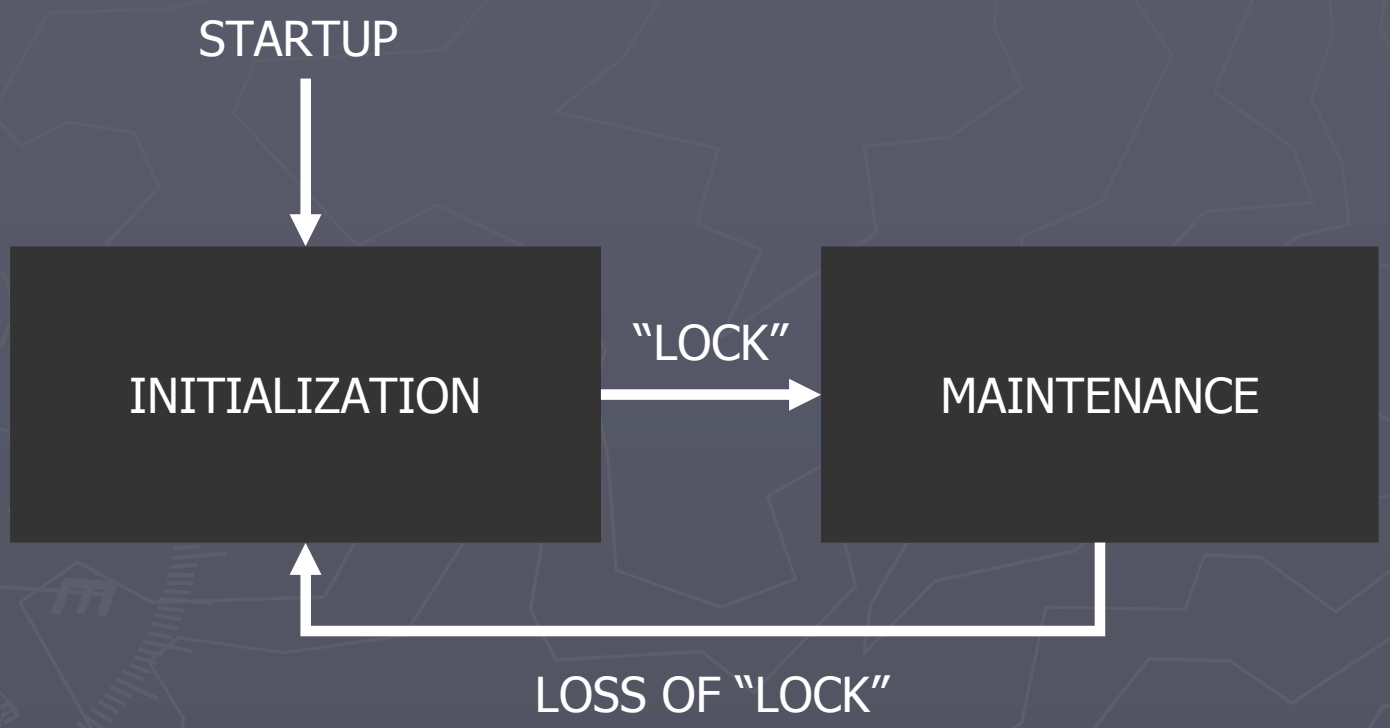
ACCURATE 6-DOF LOCALIZATION

SET OF *N* CORRECT CORRESPONDENCES

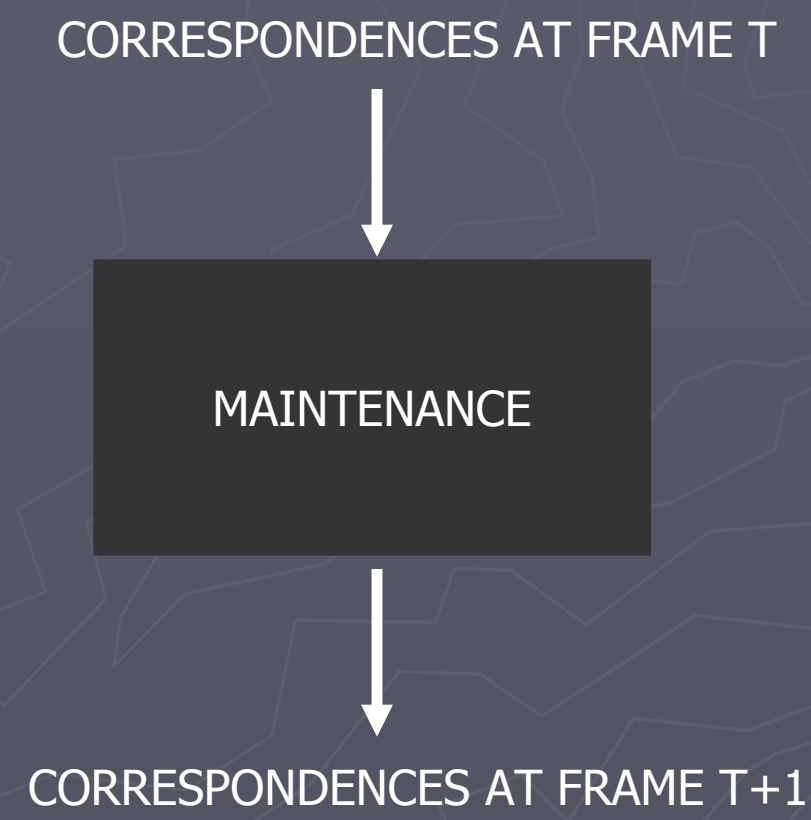
Achieve alignment by minimizing error function:

$$\xi(R,T) = \frac{1}{n} \cdot \sum_{i=1}^{n} \alpha(e_i, R, T, l_i)^2$$

Approach

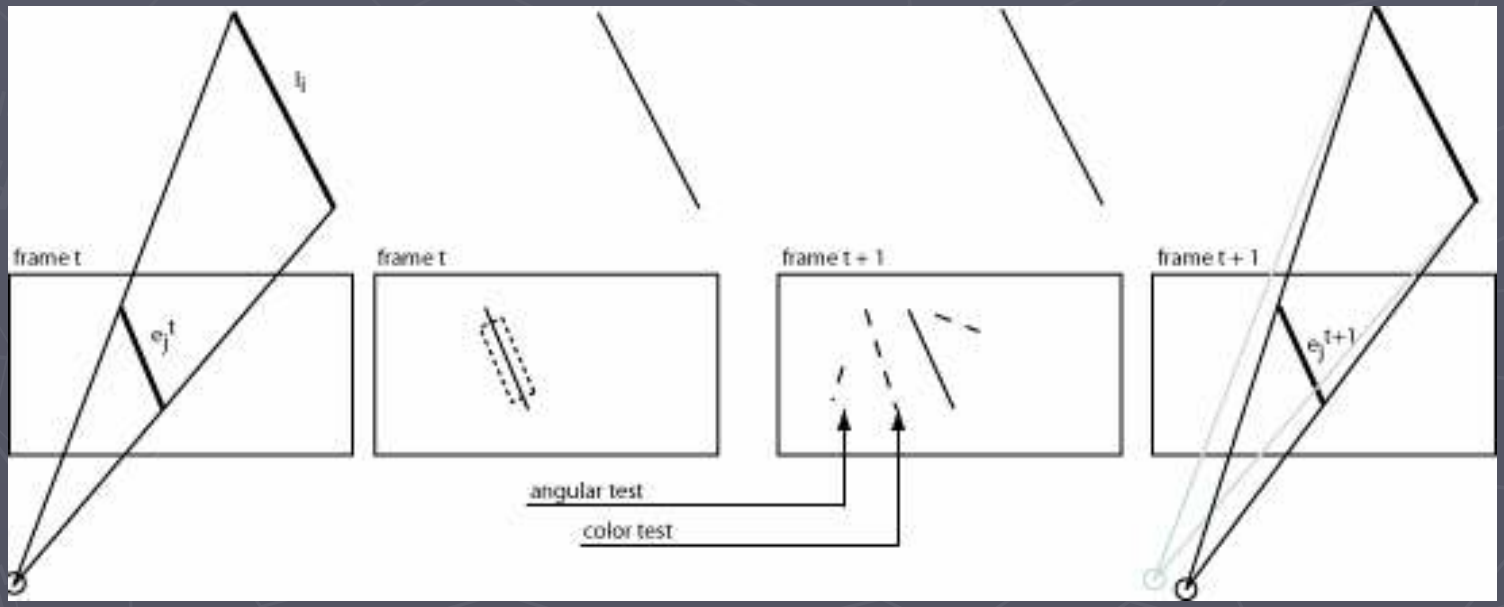


Maintenance



Approach

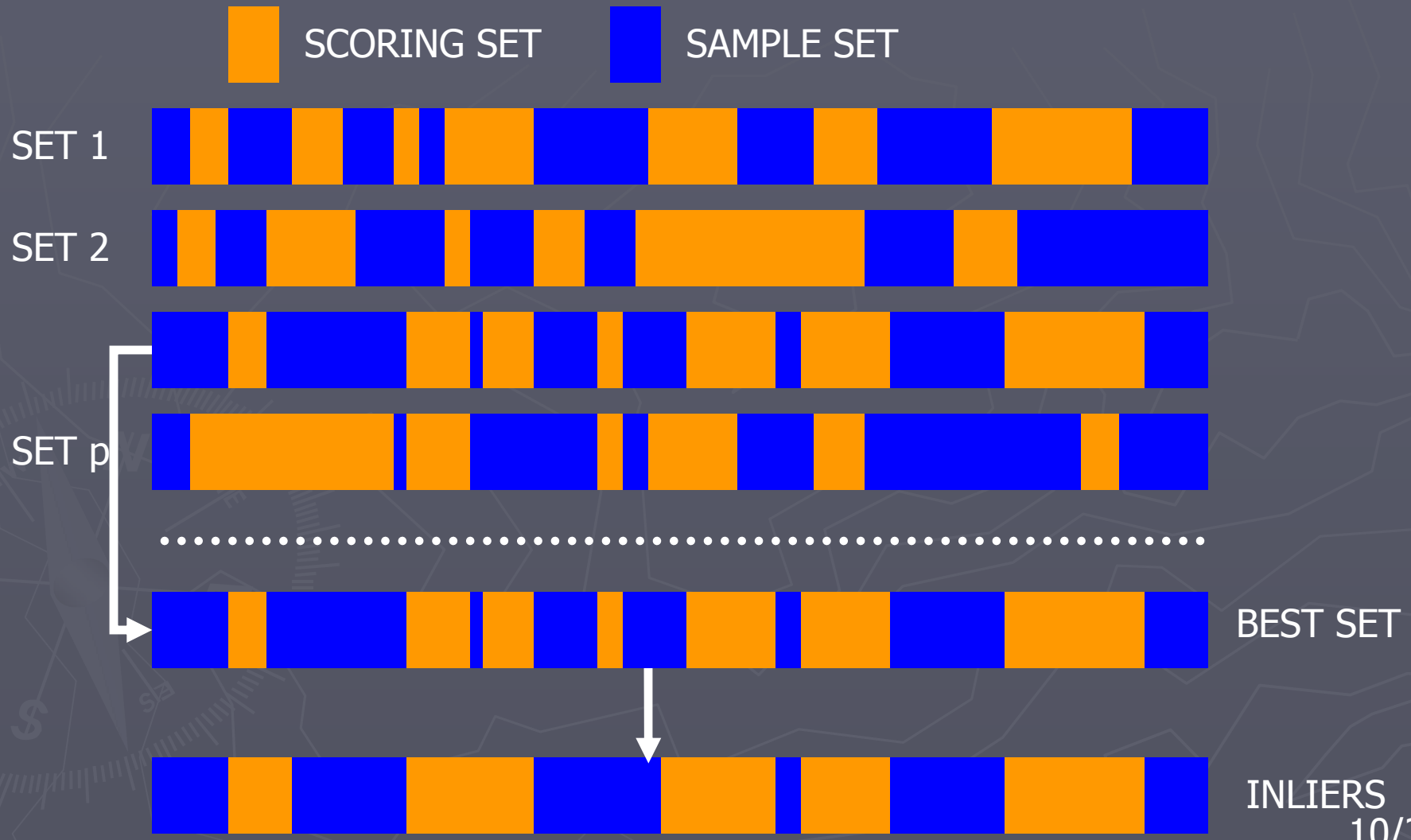
- Each *correspondence* is updated using a hue-based filter (color appearance) and an angle filter (geometric appearance).



Maintenance

- After update, *correspondences* may have :
  - No match (occlusion/matching error)
  - One or more matches (correct/incorrect)
- After random sample consensus, *correspondences* have :
  - No match (occlusion)
  - One correct match

Maintenance



Maintenance

CORRESPONDENCES AT T=0

MAINTENANCE

CORRESPONDENCES AT T=1

MAINTENANCE

CORRESPONDENCES AT T=2

MAINTENANCE

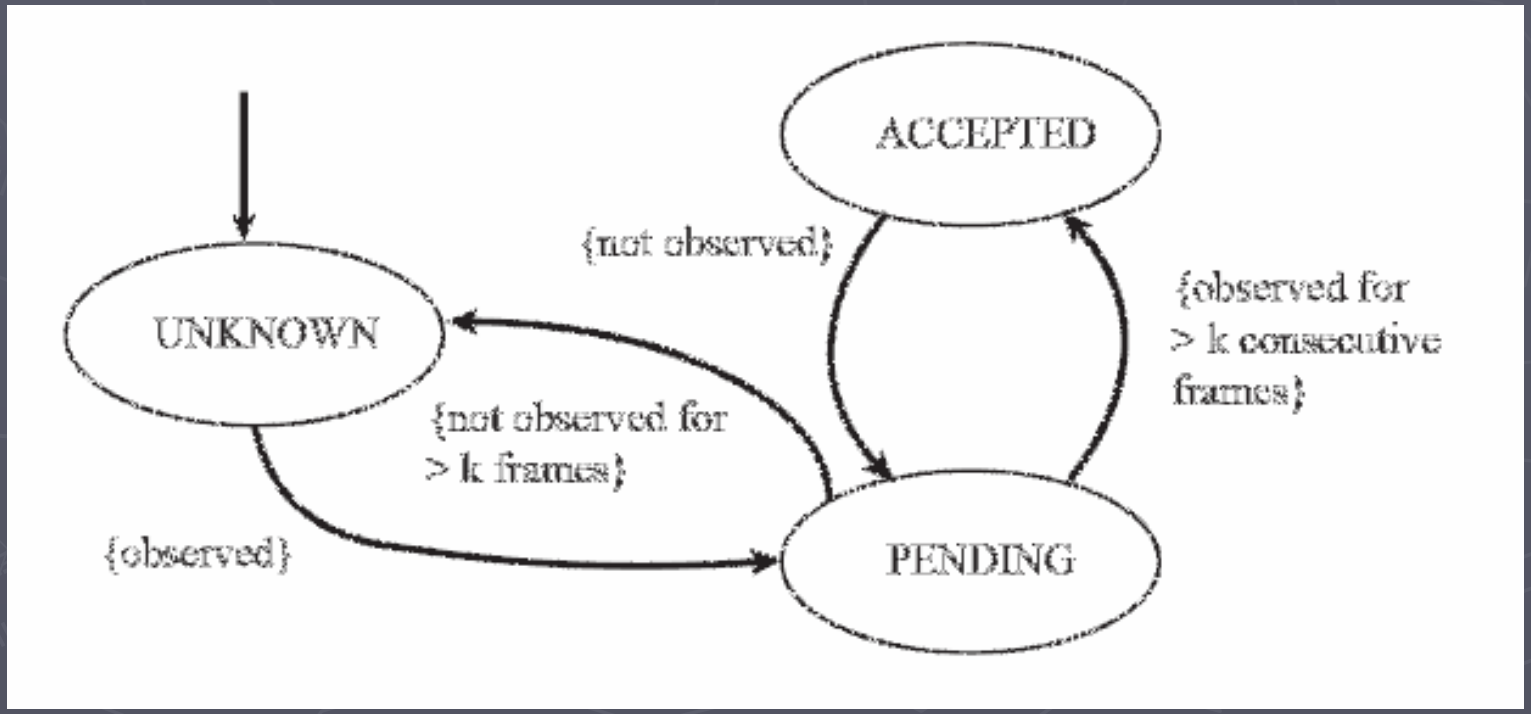
CORRESPONDENCES AT T=N

MAINTENANCE

- Possible “drift” to wrong localization

Maintenance

- Assign a *state* to each correspondence.
  - Only use mature *correspondences* for localization.

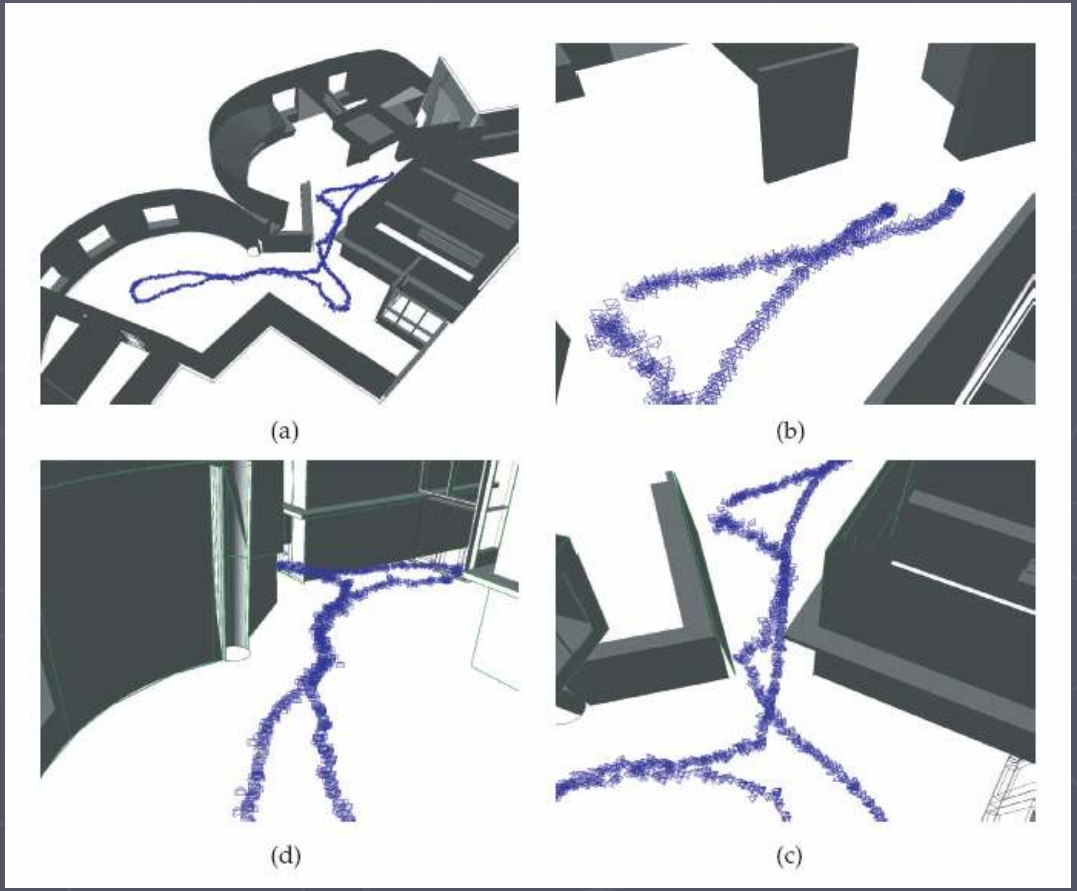


Maintenance

- Robust tracking of correspondences

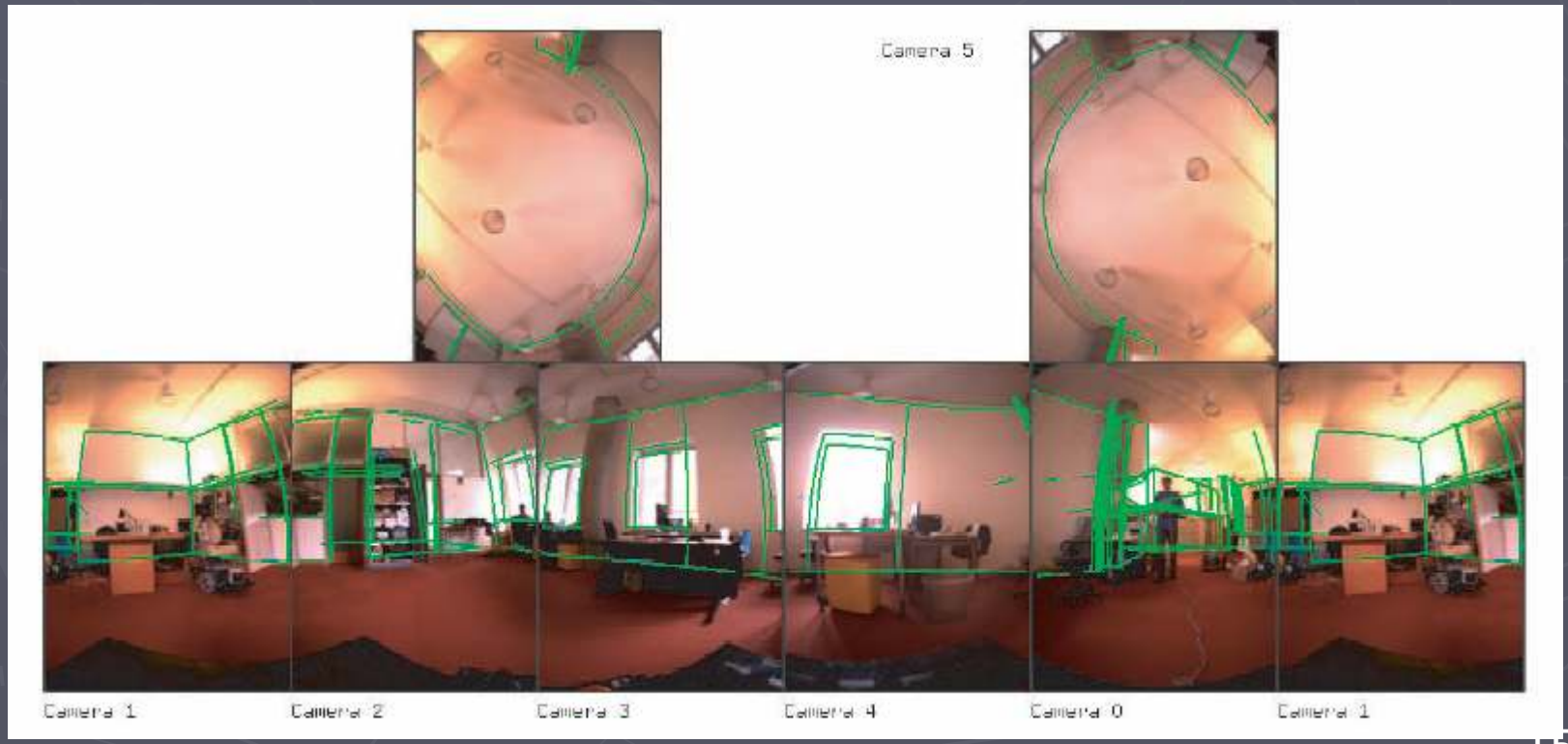
Maintenance Results

- LAB dataset: 1,500 frames @ 5Hz; 5min; 120m

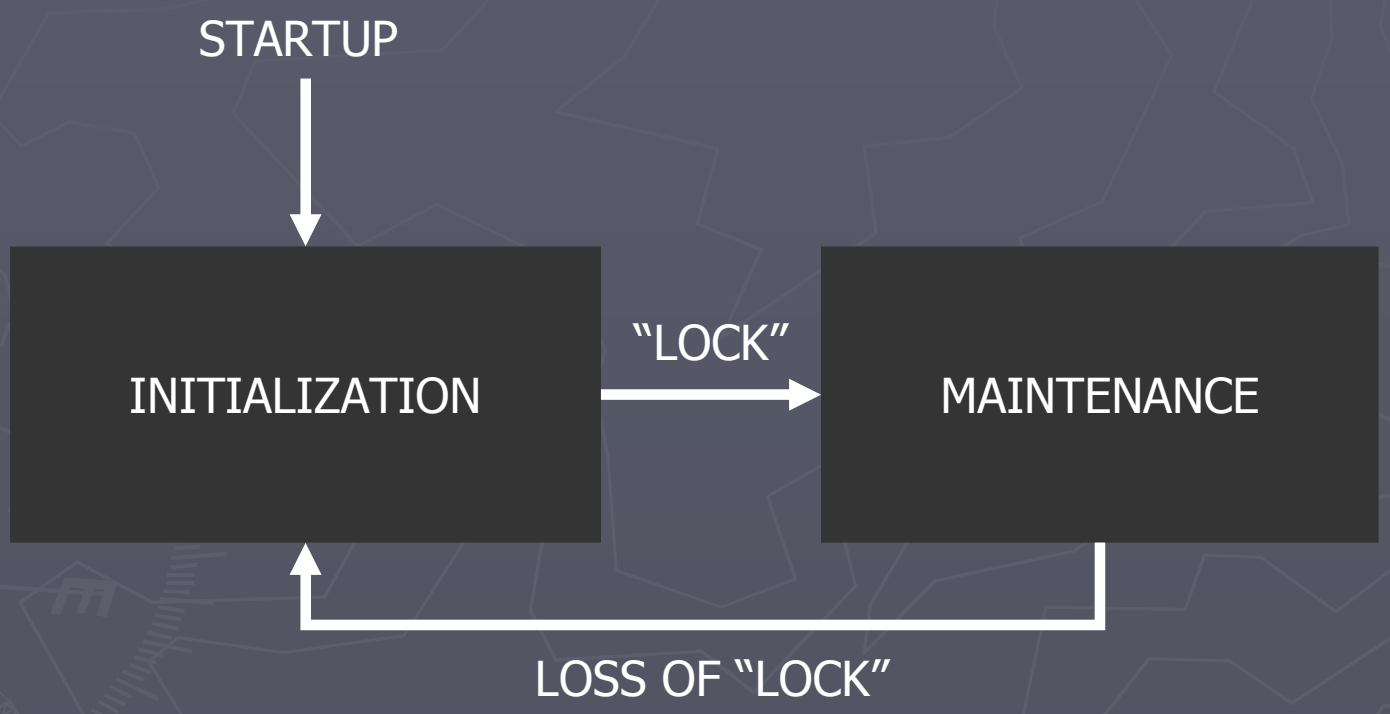


Maintenance Results

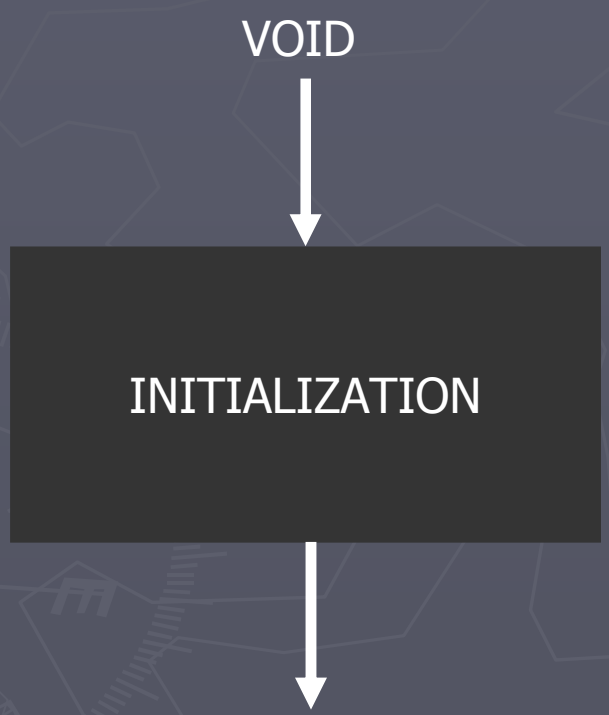
- LAB dataset: 1,500 frames @ 5Hz; 5min; 120m



Approach

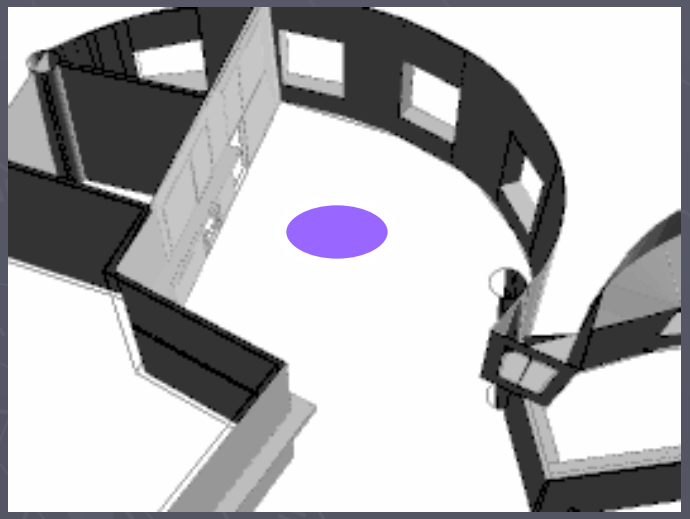


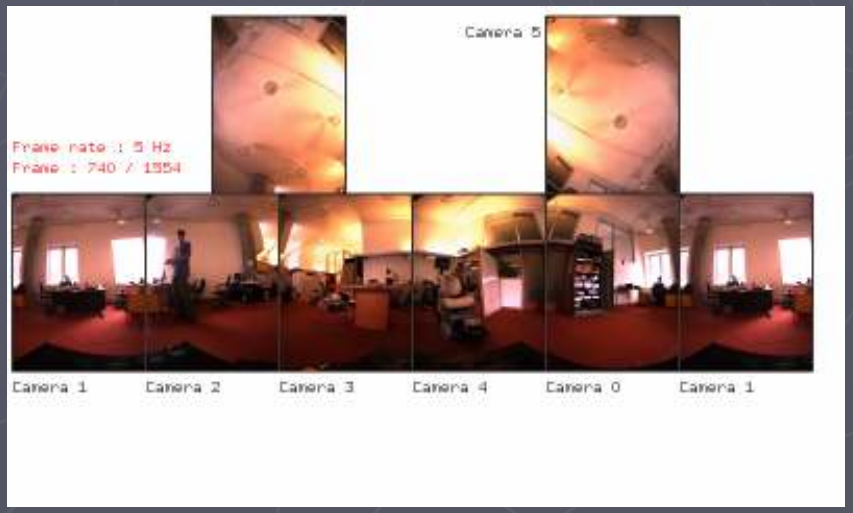
Initialization



Initialization

- Init from a *single* omnidirectional image and a 3-meter diameter position "seed."

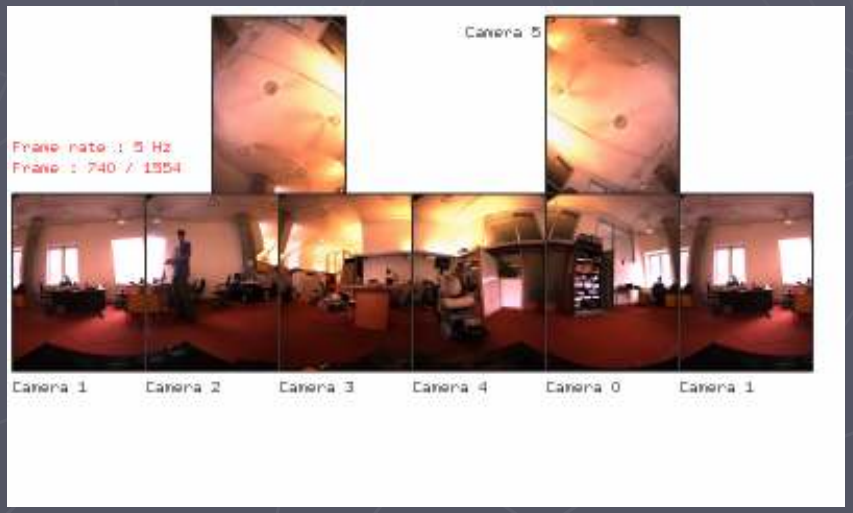




- Data association problem

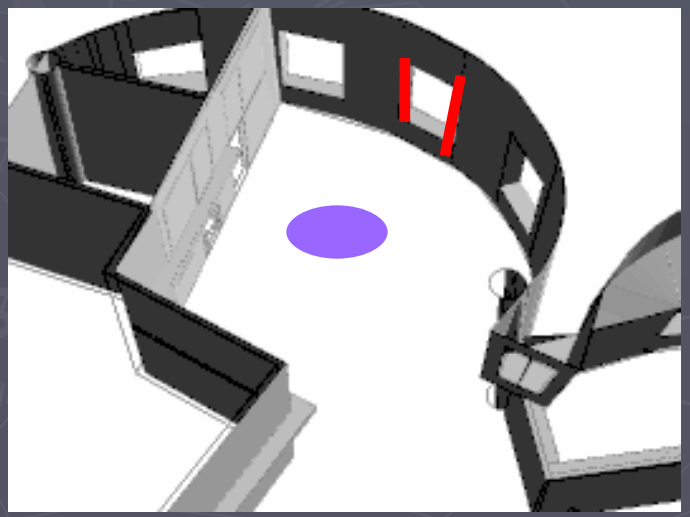
Initialization

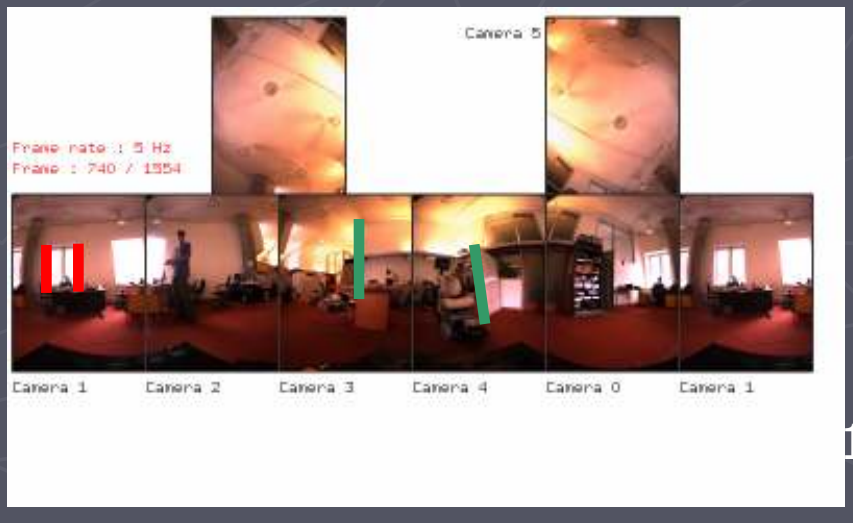
- RANSAC… takes forever!
- Geometric constraints
  - "Smart" RANSAC



Initialization

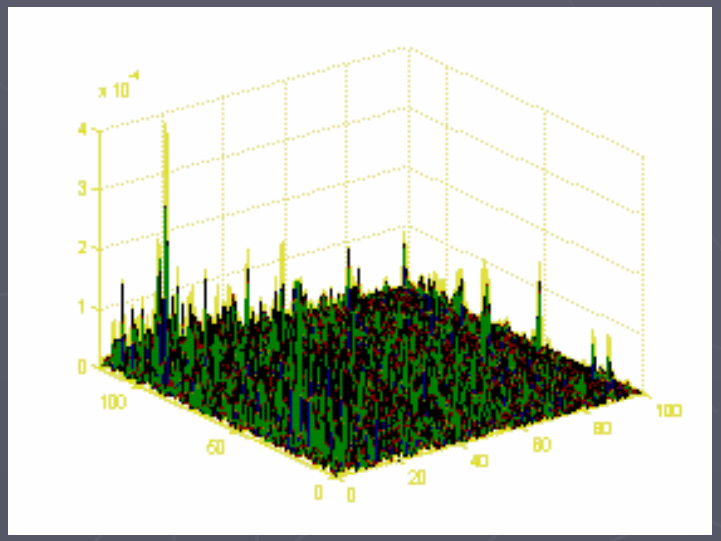
- Start with a pair of image edges and a pair of 3D model segments
- Score [0,1] whenever they could match
- Aggregate score over all possible pairs





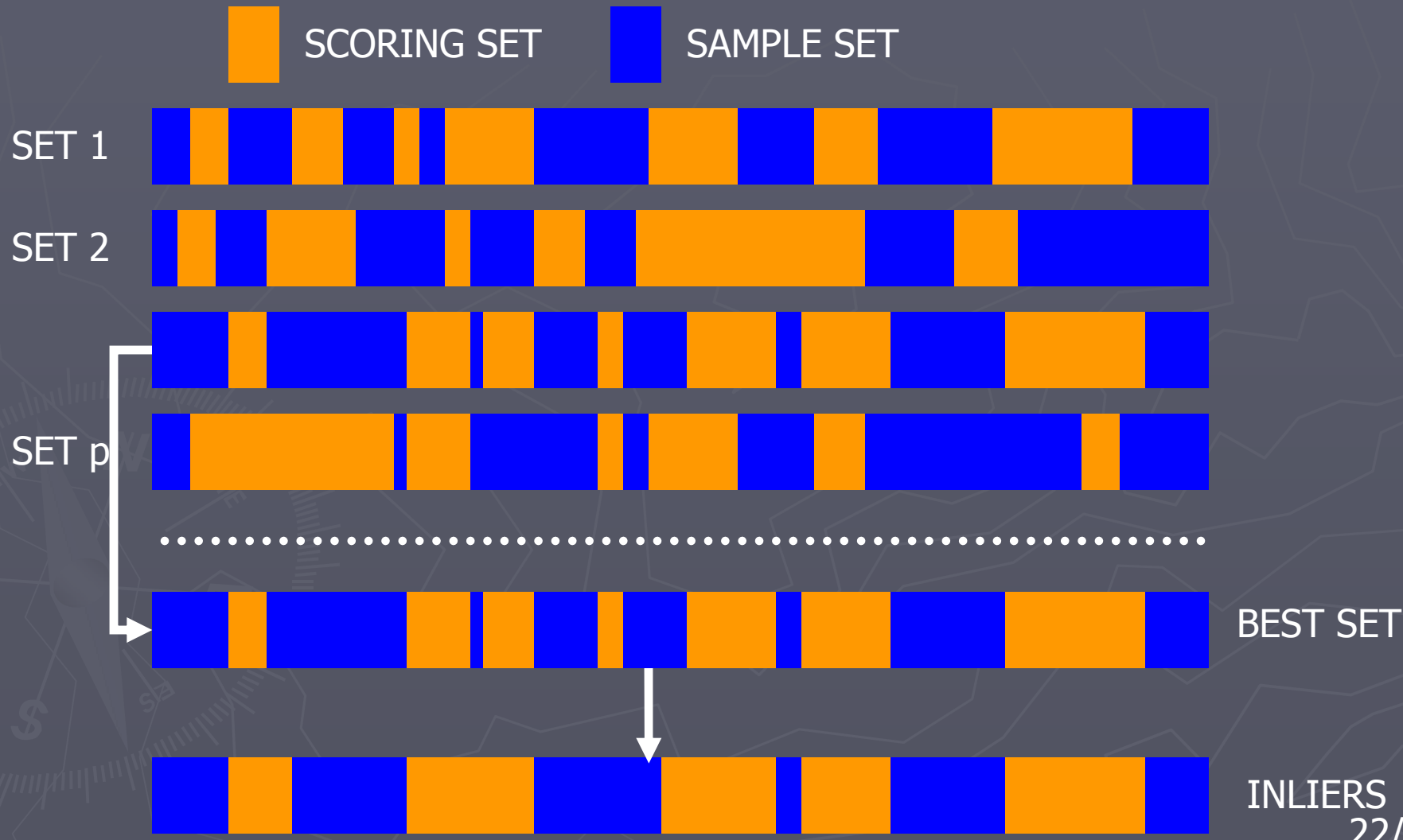
Initialization

- Scoring table:
  (M model lines x N image edges)
- Data is noisy but good enough to extract putative correspondences.
- Generate sets of correspondences between model lines and image edges.



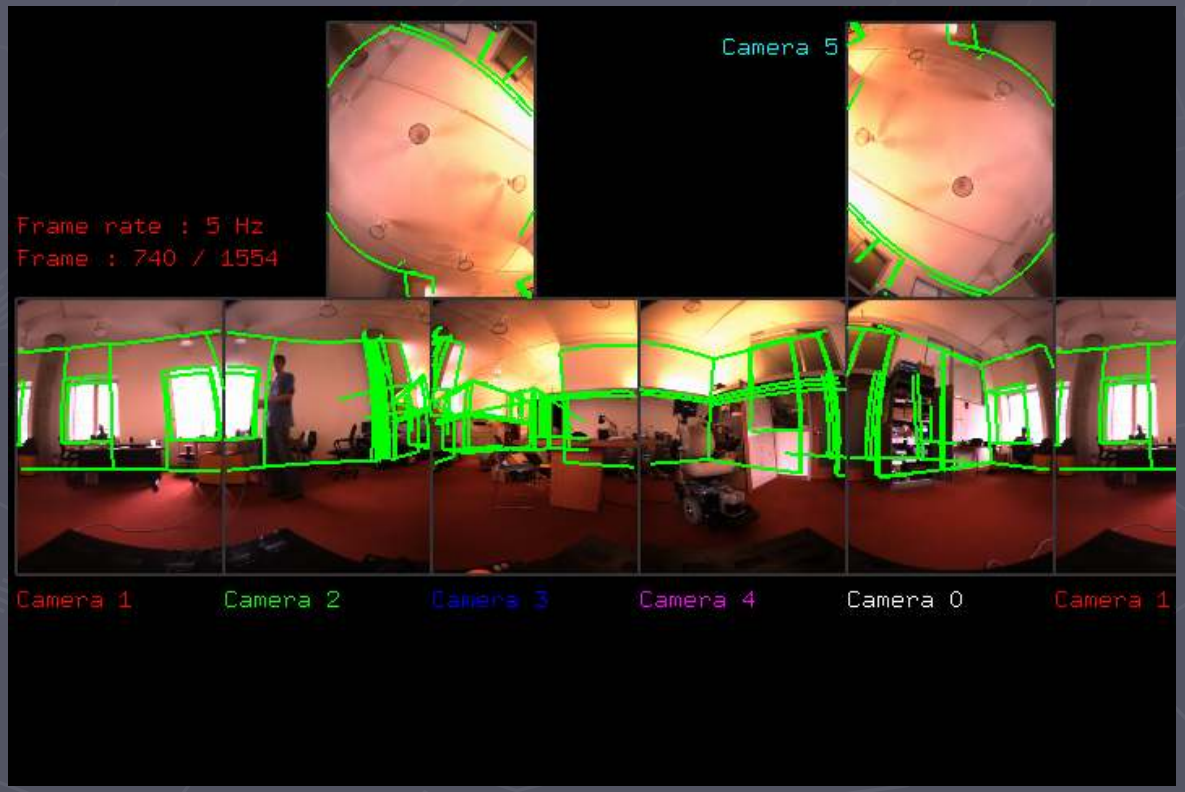
Scoring table between 3D model lines and 2D image edges

Initialization

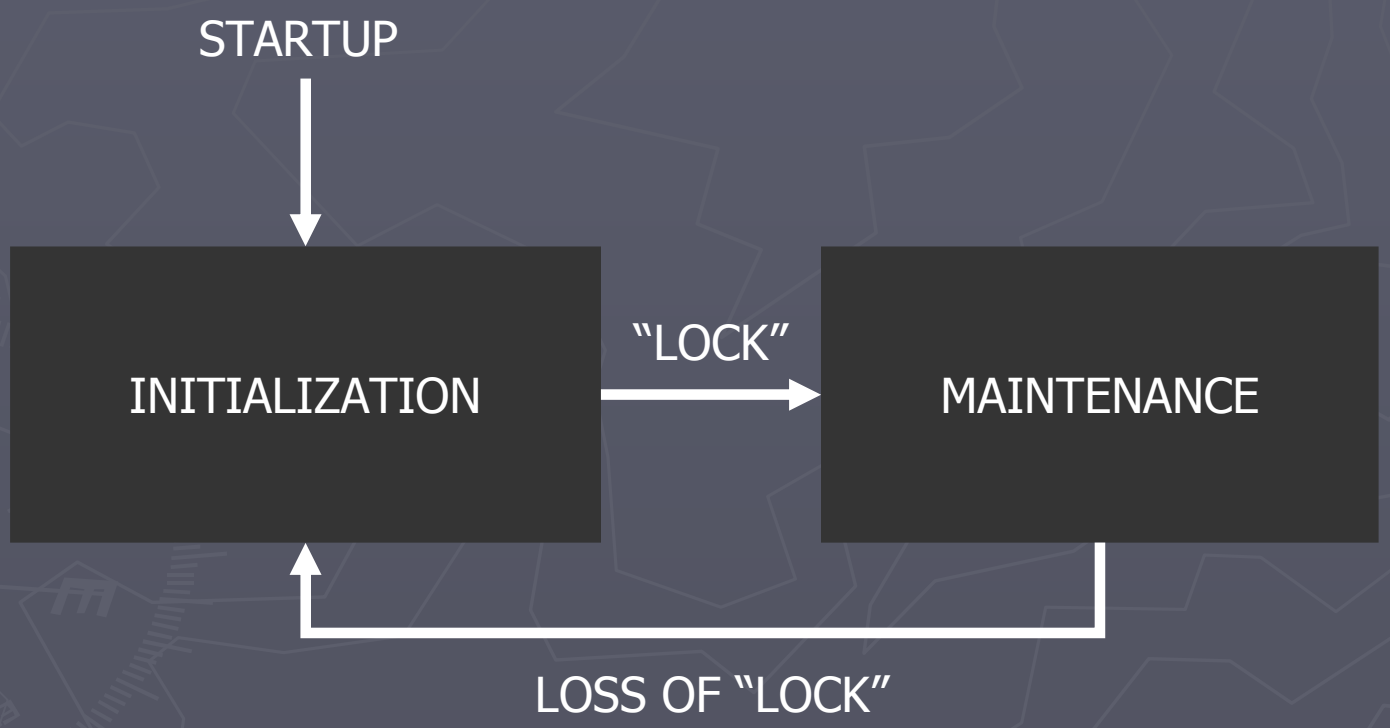


Initialization

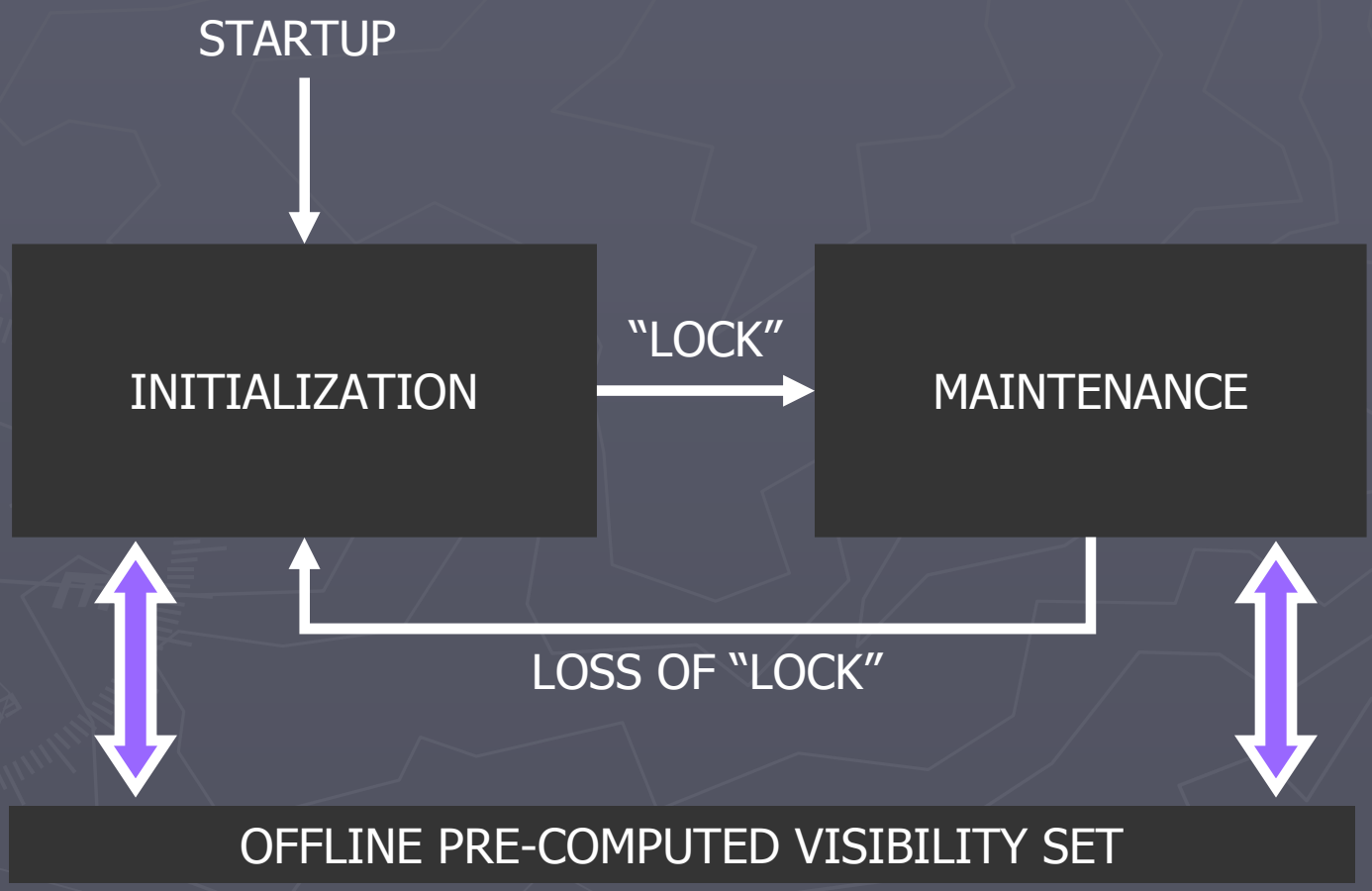
- After initialization…



Approach

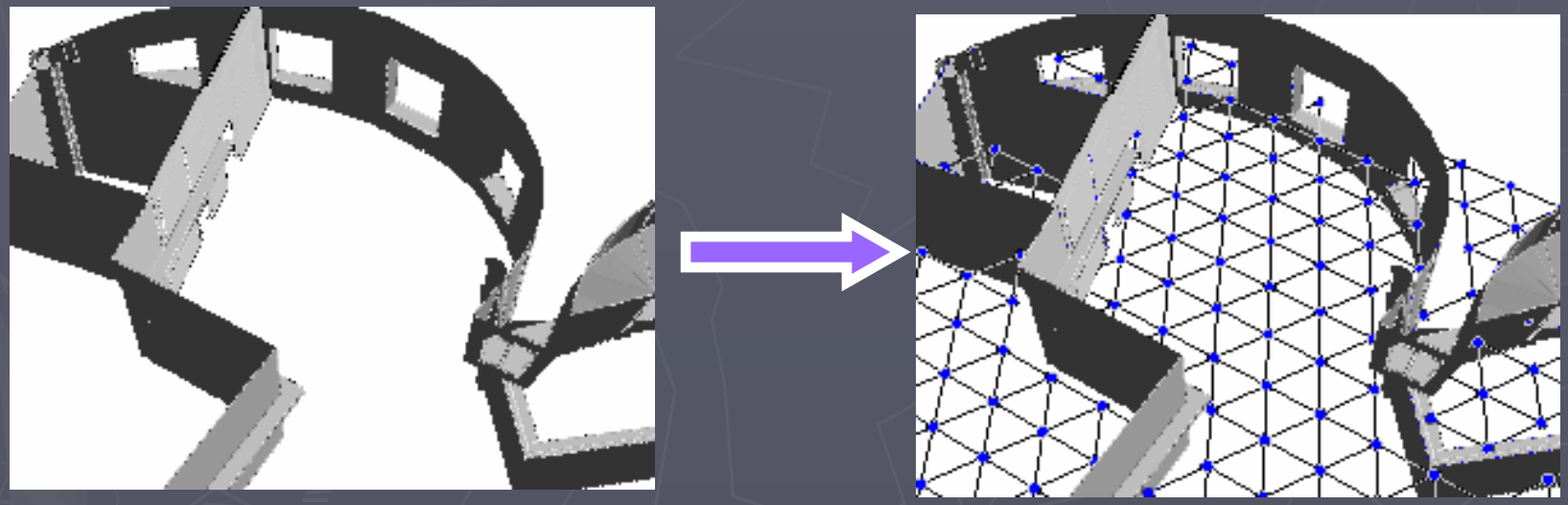


Visibility Set Computation



Visibility Set Computation

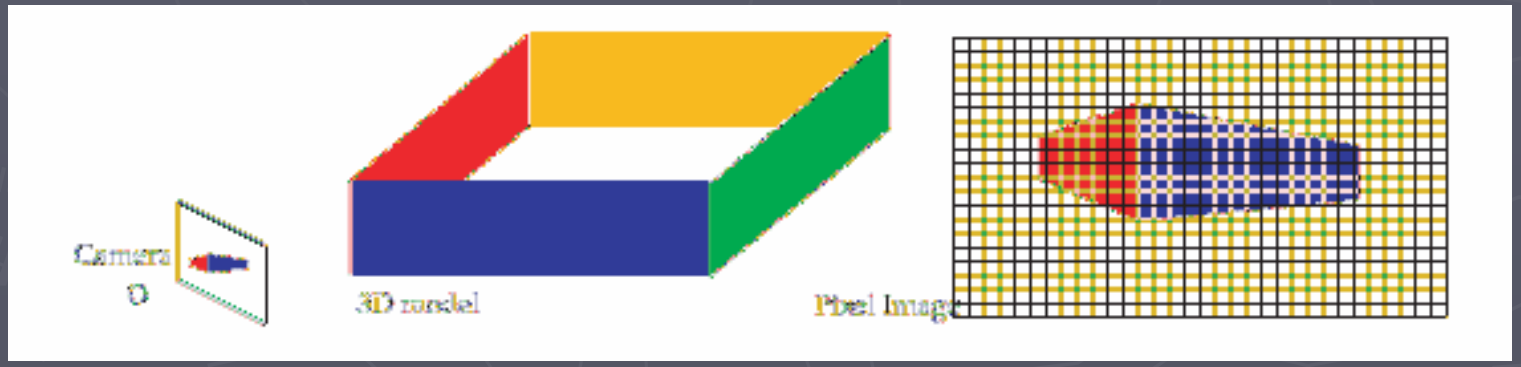
- Space is subdivided into *nodes*.



- At each node, the set of visible faces and 3D line segments is computed and stored in a database.

Visibility Set Computation

- OpenGL-based computation
  - Fast, cheap, easy.

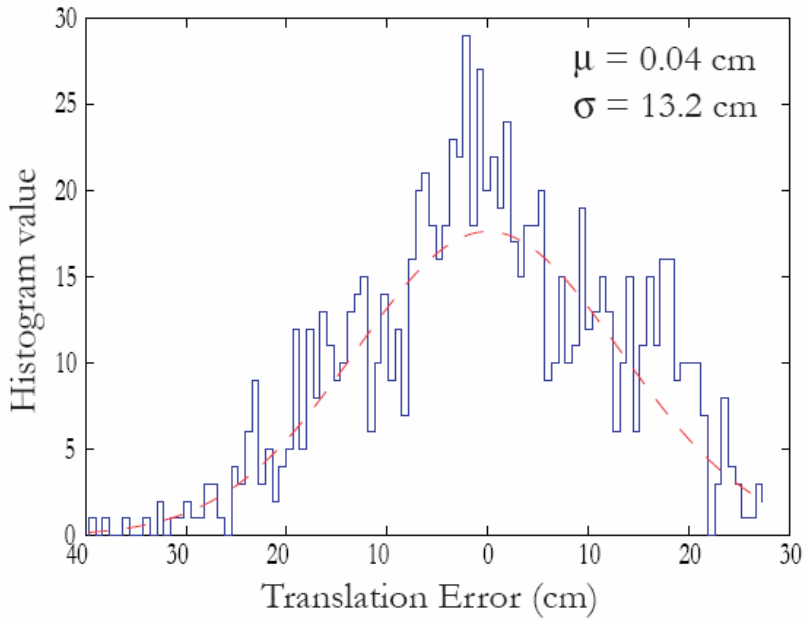


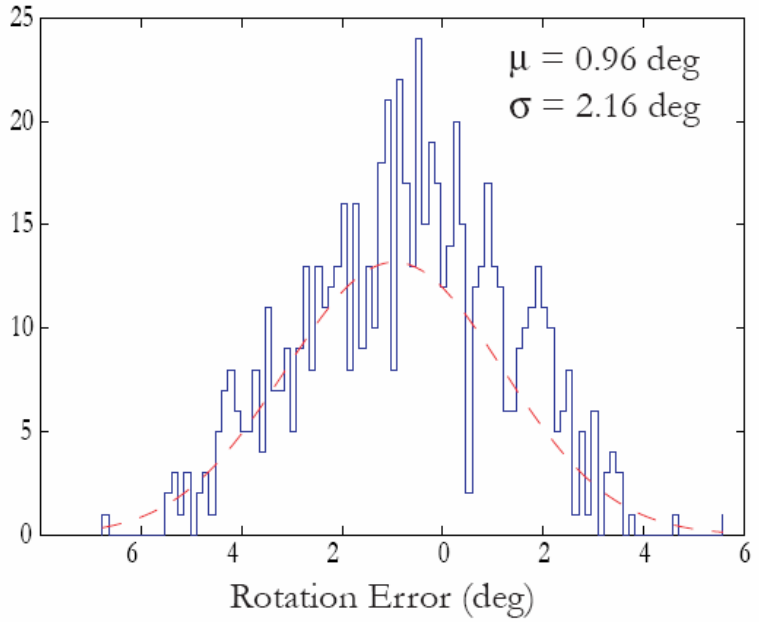
- Standard building = 100MB, 20min CPU.

Visibility Set Computation

Results

- Localization Accuracy (histogrammed):

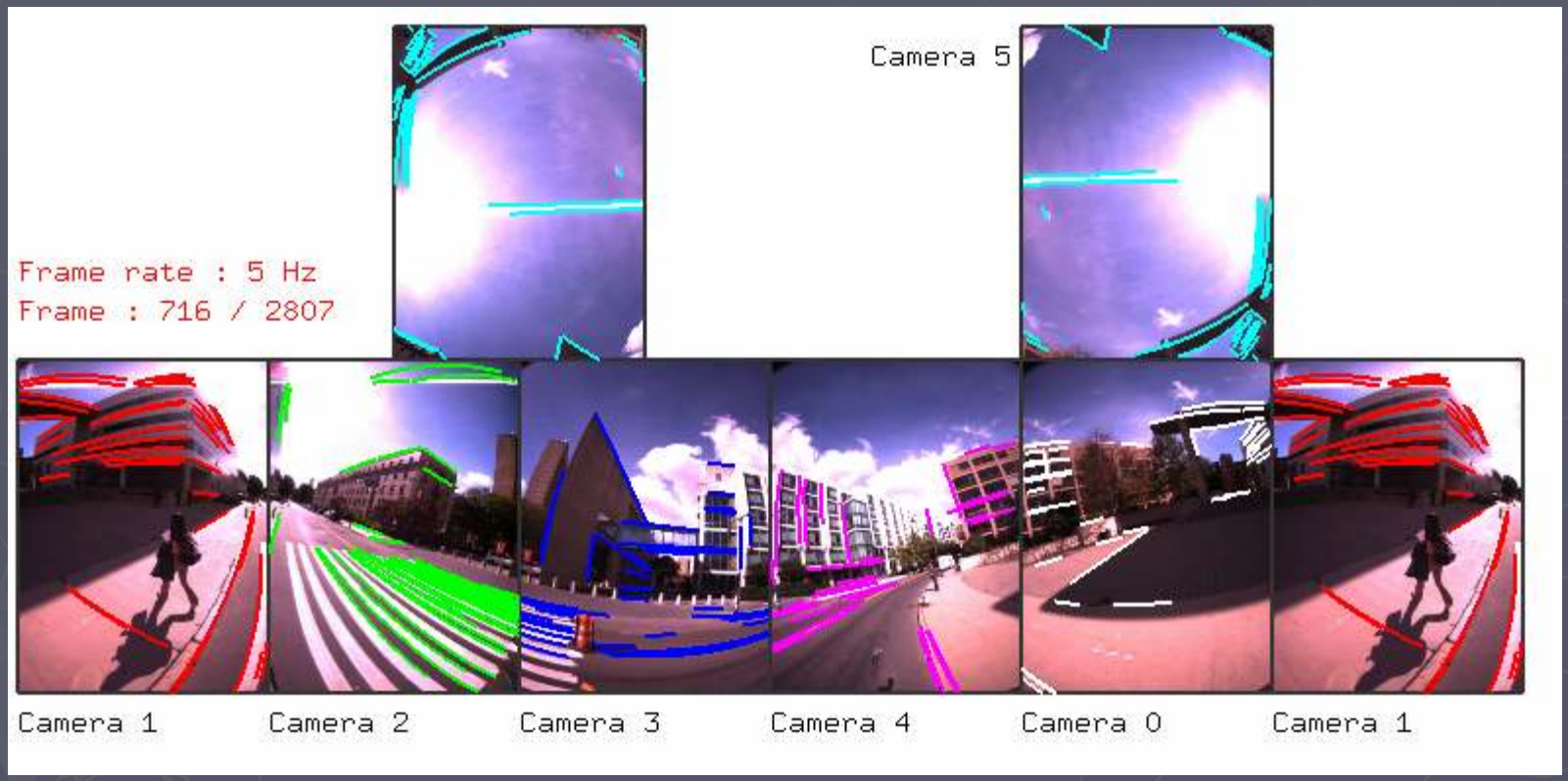




Limitations

- Computationally intensive (achieves 1Hz)
- Initialization requires accurate location hint
- Localization accuracy compromised by:
  - Errors in 3D model
  - Image noise (in edge estimation)
  - Feature matching errors
- Sensor light-sensitivity
  - Challenged by low-light (indoor) scenes

Outdoor imagery



Continuing work ...

- Signature-based initialization
- Integration of inertial sensors
- Fusion of points and segments
- Online update of the 3D model