CSAIL

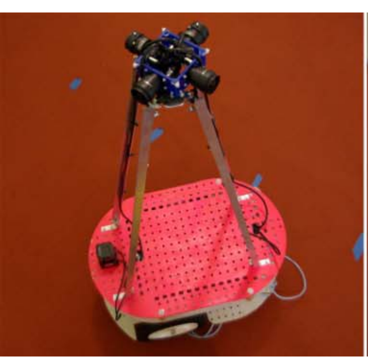
Body-Relative Navigation Guidance using Uncalibrated Cameras

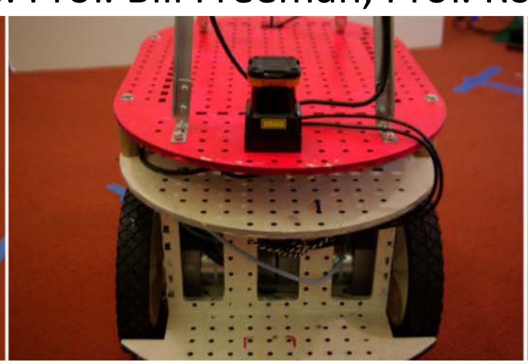
Olivier Koch

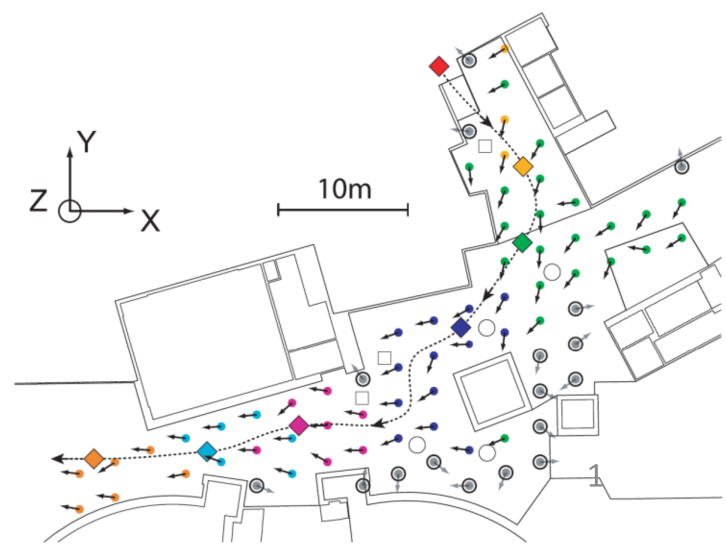
PhD Thesis Defense

21-Jan-2010

Thesis Committee: Prof. Bill Freeman, Prof. Rob Miller, Prof. Seth Teller







Problem Statement

Human-oriented navigation in GPS-denied environments

- Finding your way in **unknown environments** with **no prior map** and **no external source of localization**

Soldiers in the field

Hazmat teams

Visually impaired

Problem Statement

Rely on internal sensing to provide navigation guidance:

- Accurate
- Spatially extended
- Temporally consistent
- Complex, cluttered, dynamic environments

Soldiers in the field

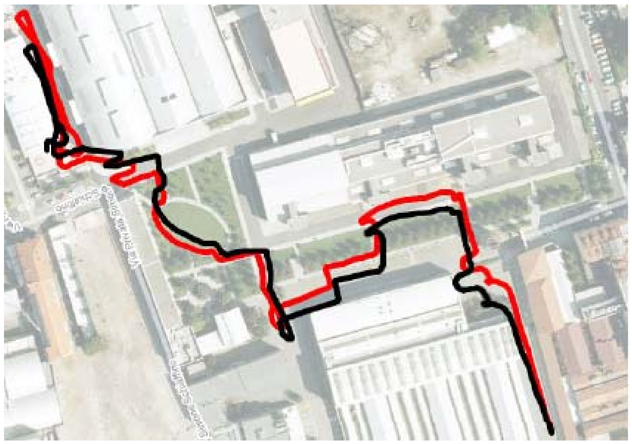
Hazmat teams

Visually impaired

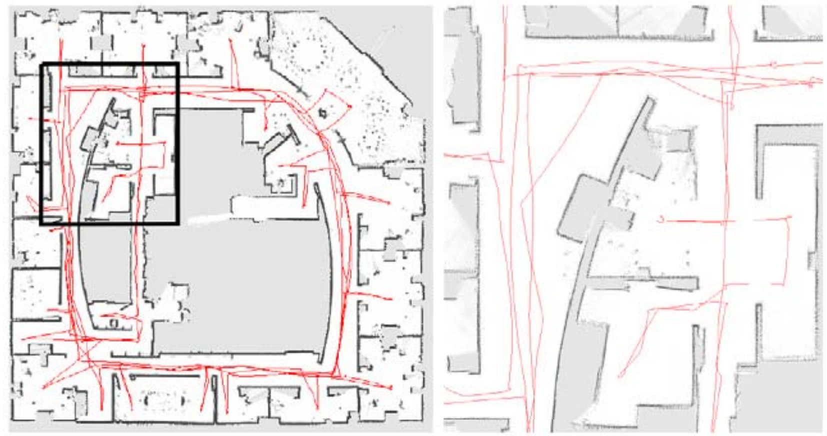
Related Work

Simultaneous Localization and Mapping (SLAM)

- Build map and localize agent in the map at the same time



Visual odometry over 650m trajectory (Civera, 2009)

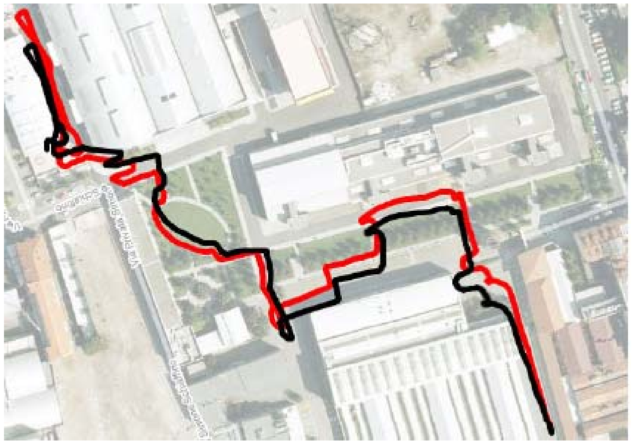


Laser-based SLAM, 30x30m (Grisetti, 2006)

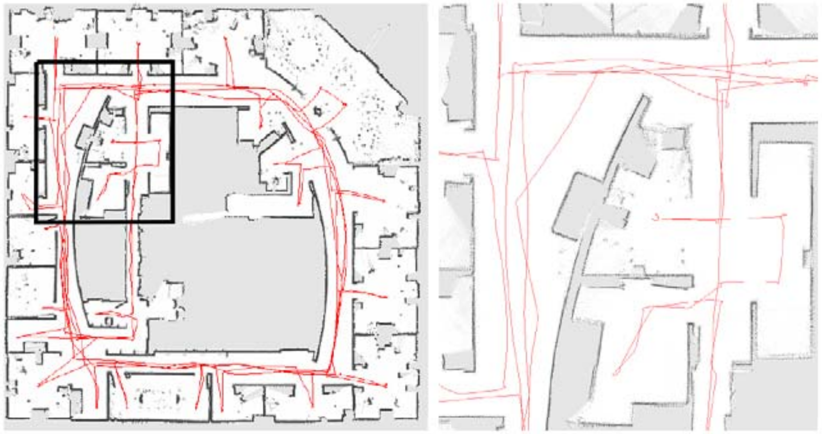
Related Work

Simultaneous Localization and Mapping (SLAM)

- Build map and localize agent in the map at the same time



Visual odometry over 650m trajectory (Civera, 2009)



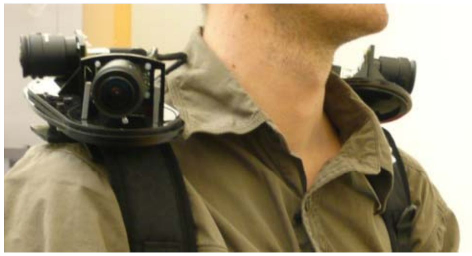
Laser-based SLAM, 30x30m (Grisetti, 2006)

Limitations of SLAM

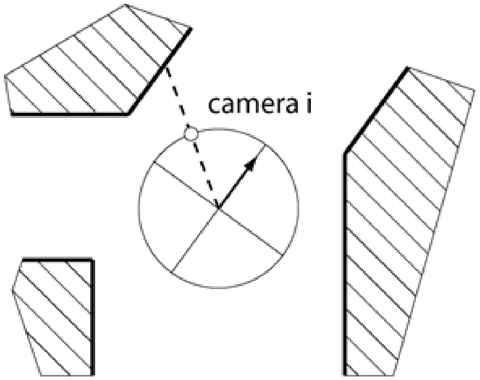
- Sensitive to degenerate user motions
- Sensitive to ambiguous world configurations
- Challenging in dynamic environments
- Requires full sensor calibration

Full metric reconstruction overkill for human navigation?

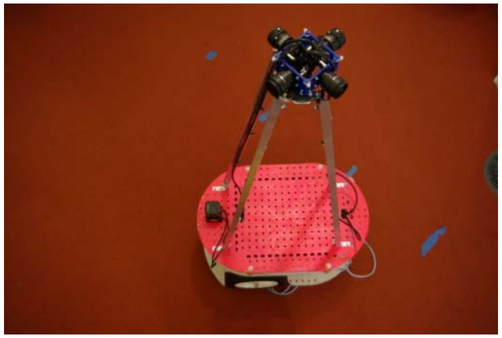
Contributions



- Sensor suite for vision-based navigation

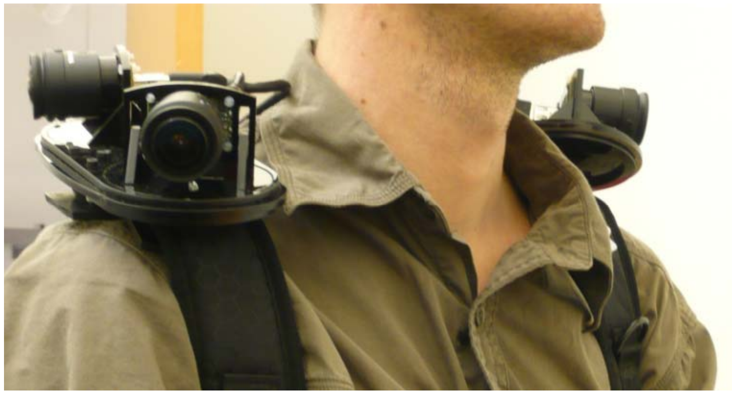


- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance
- Application to ground robot navigation



- User Study

System Overview



Four cameras, 752 x 480 8-bit images
Horizontal FOV: 360°

Sample frame, MIT Bldg 2

Why vision?

- Light, cheap, compact
- Rich information

Why uncalibrated cameras?

- Intrinsic calibration is tedious

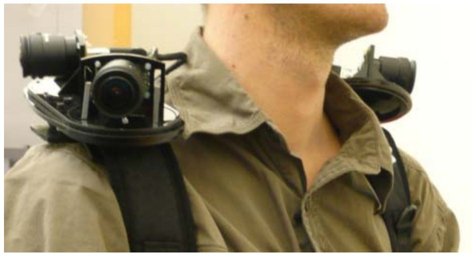
(center of projection, focal length, distortion parameters)

- Extrinsic calibration is hard

(sensor-to-body transformation, 6DOF/camera)

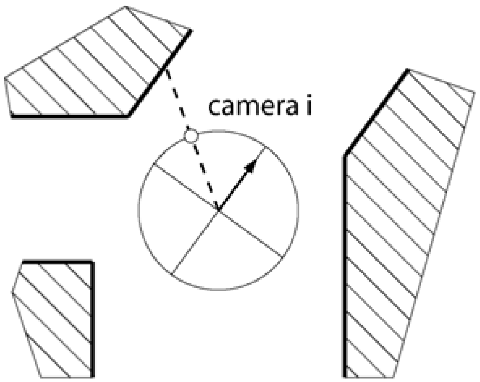
- Calibration subject to change in real conditions

Outline

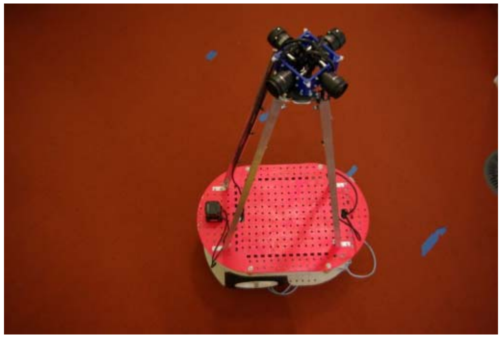


- Sensor suite for vision-based navigation

- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance



- Application to ground robot navigation



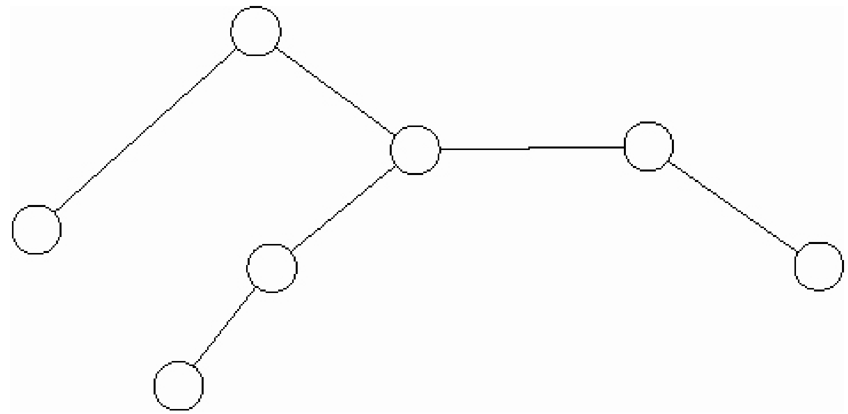
- User Study

Method Overview

Exploration (first visit)

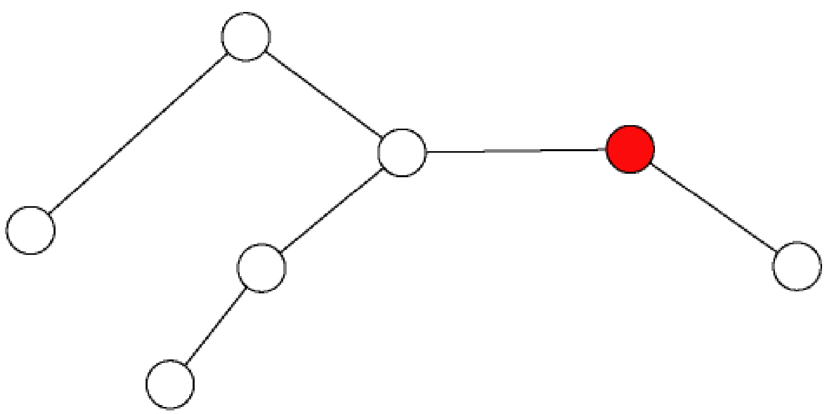
Place graph

- Node = location in the world
- Edge = physical path btw nodes
- Arbitrary 3D environments

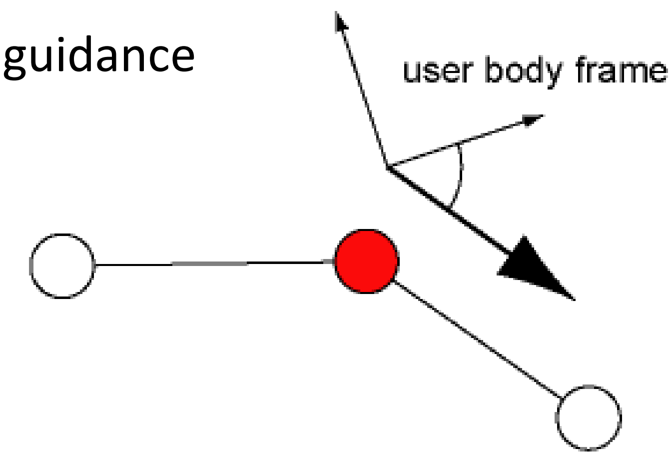


Navigation (successive visit)

Localization (node estimation)

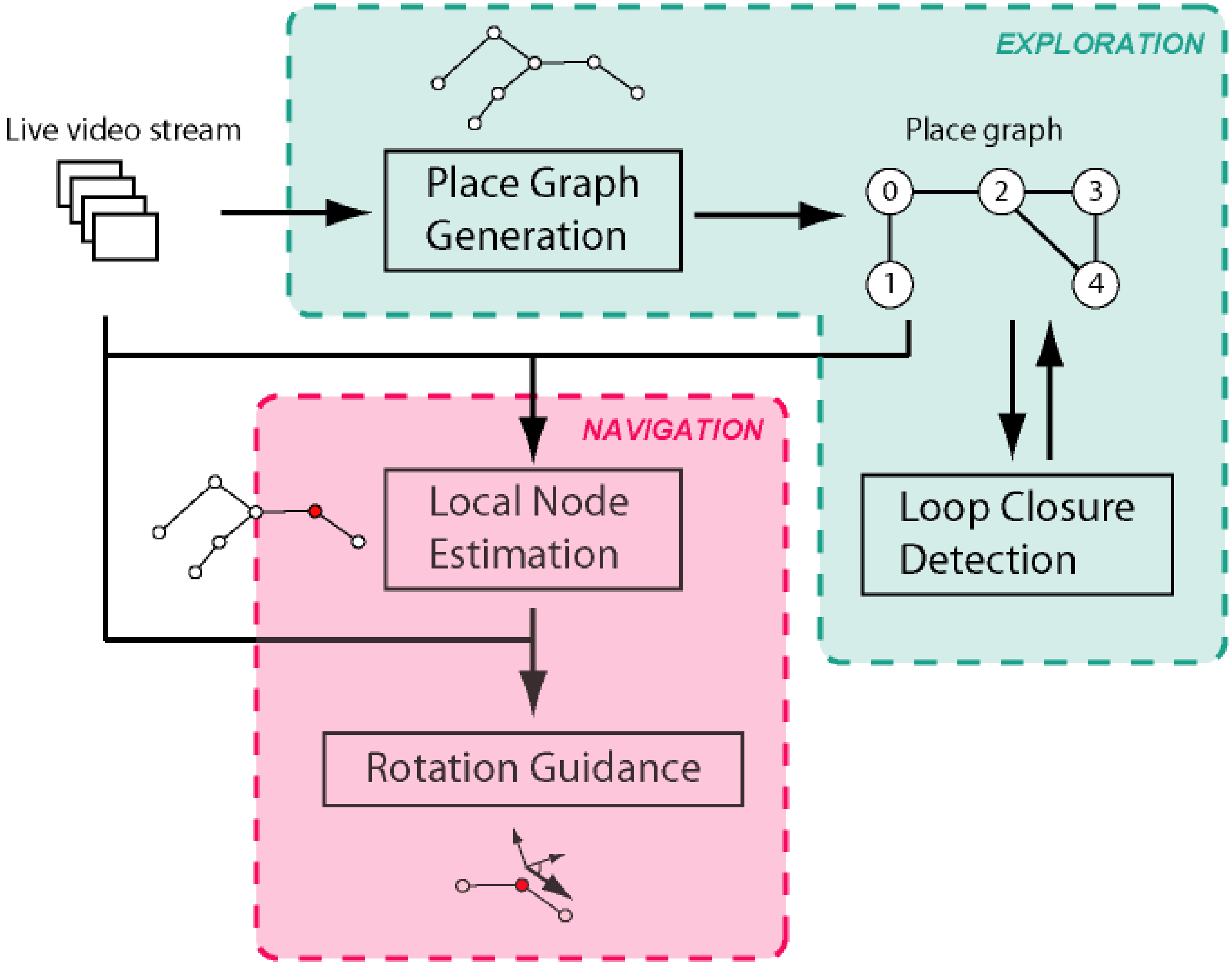


Rotation guidance



Body-relative navigation guidance using uncalibrated cameras*, O. Koch, S. Teller, ICCV 2009*

Method Overview

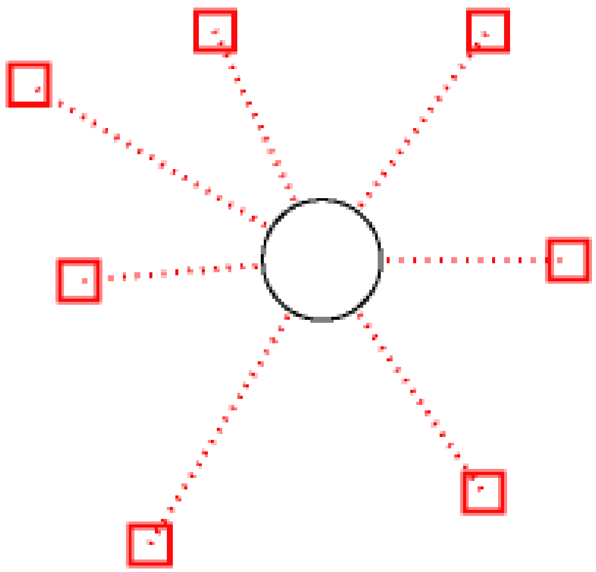


***Body-relative navigation guidance using uncalibrated cameras**, O. Koch, S. Teller, ICCV 2009*

Place Graph

Place Graph Generation

- Sparse representation of the exploration path



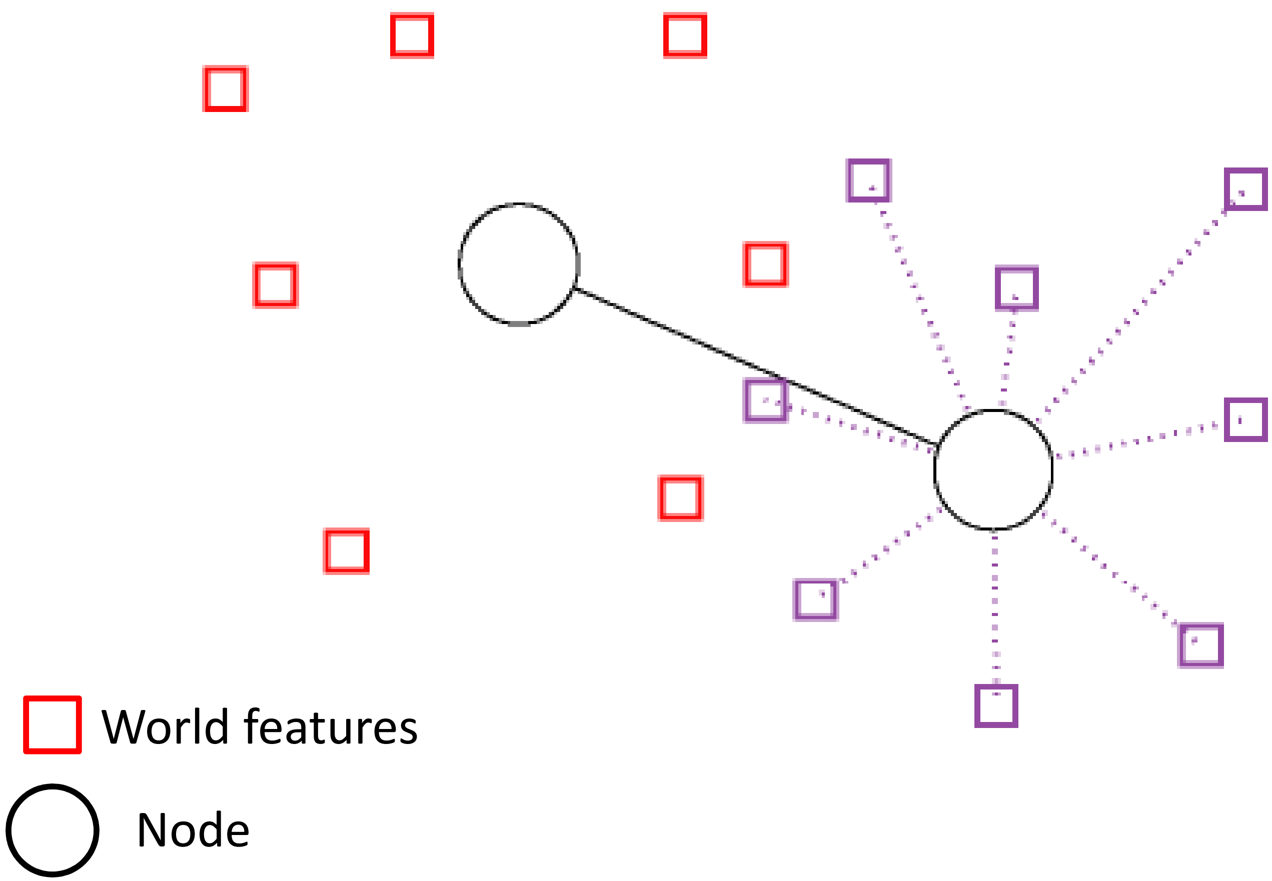
World features

Graph node

Place Graph

Place Graph Generation

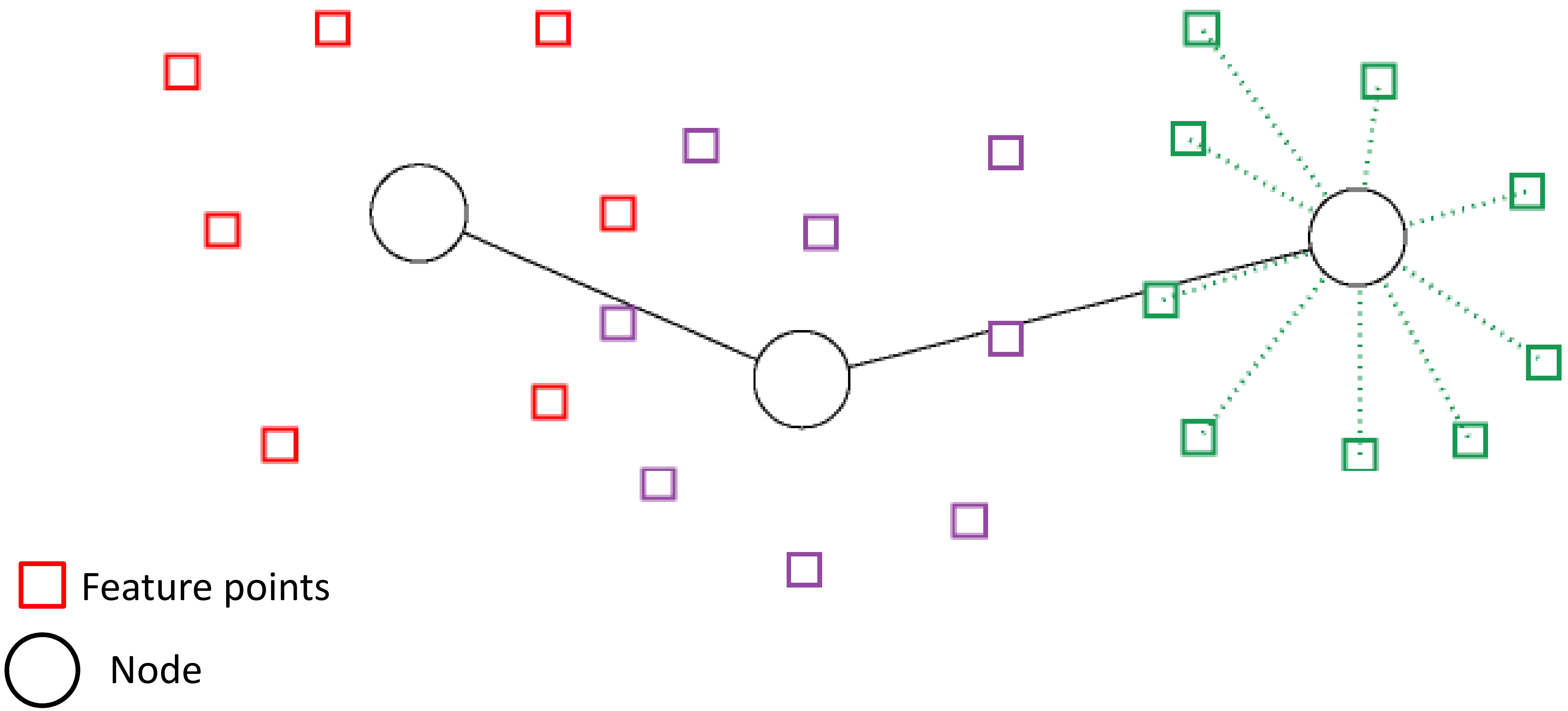
- Sparse representation of the exploration path



Place Graph

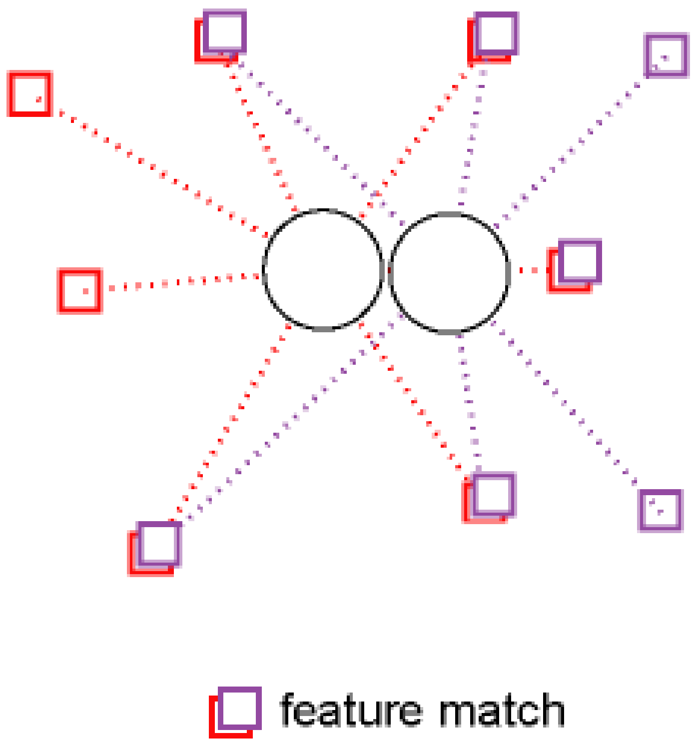
Place Graph Generation

- Sparse representation of the exploration path



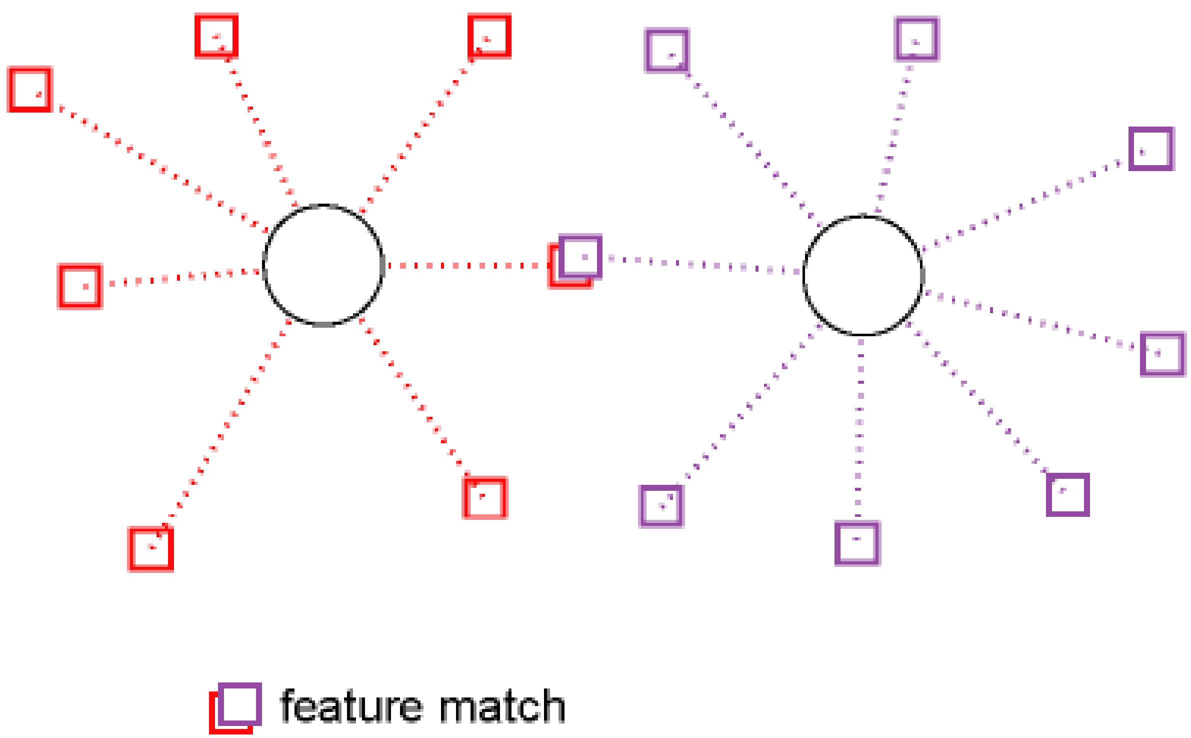
Place Graph Generation

- Variation in environment appearance is captured by feature matching



Place Graph Generation

- Variation in environment appearance is captured by feature matching

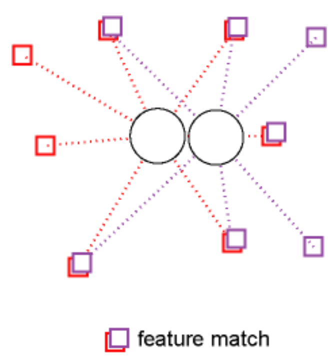


Place Graph Generation

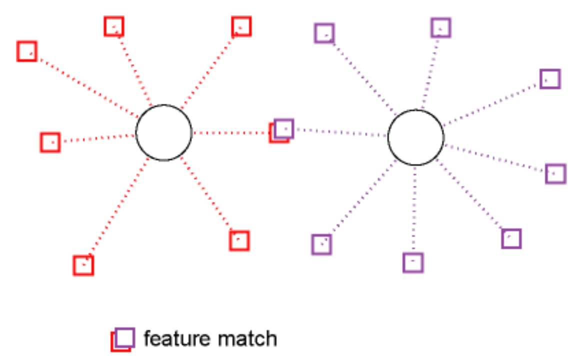
- Energy function Ψ measures visual similarity.

$$\Psi(F_1,F_2)=\frac{1}{|F_1|+|F_2|}\cdot\left(\underbrace{|F_1|+|F_2|-2\cdot|\Phi(F_1,F_2)|}_{\text{penalty for unmatched features}}+\underbrace{\sum_{(f_1^i,f_2^j)\in\Phi(F_1,F_2)}\left\|f_1^i-f_2^j\right\|_2}_{\text{average L2 distance in feature space}}\right)$$

F1, F2: feature sets
Φ: feature matching function



$0 < \Psi < \Psi_0$

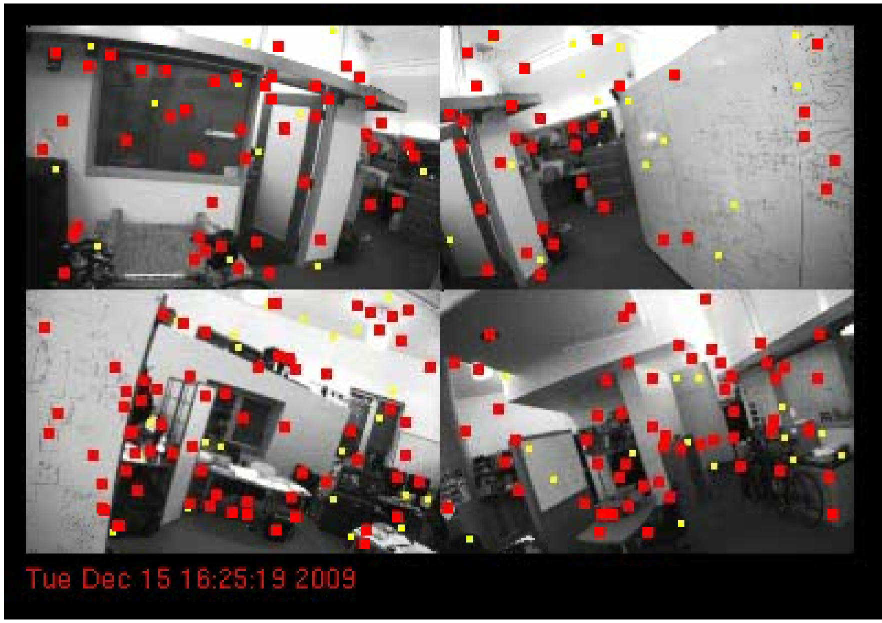


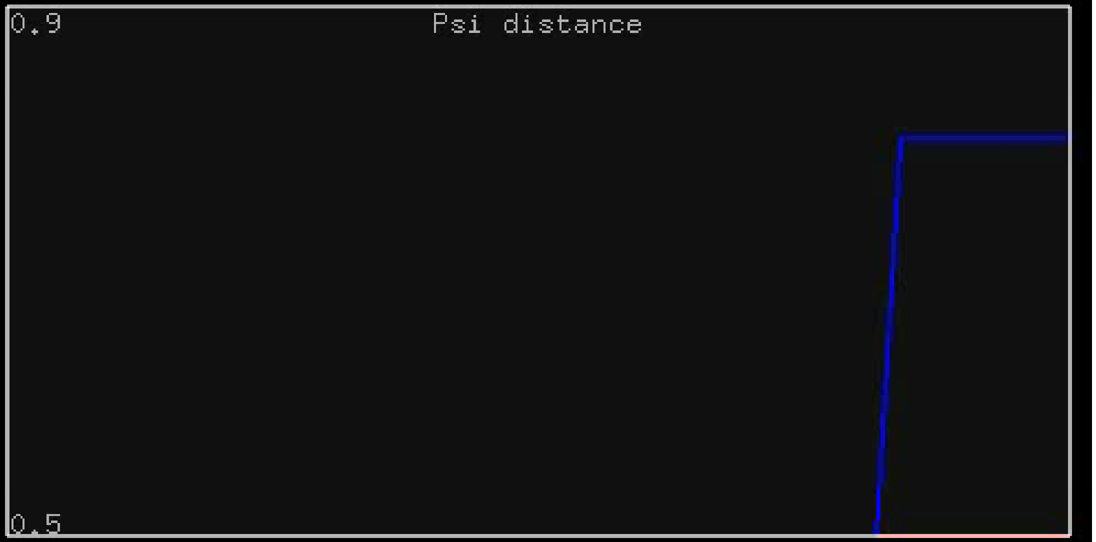
$\Psi_0 < \Psi < 1$

Place Graph Generation

Place Graph Generation

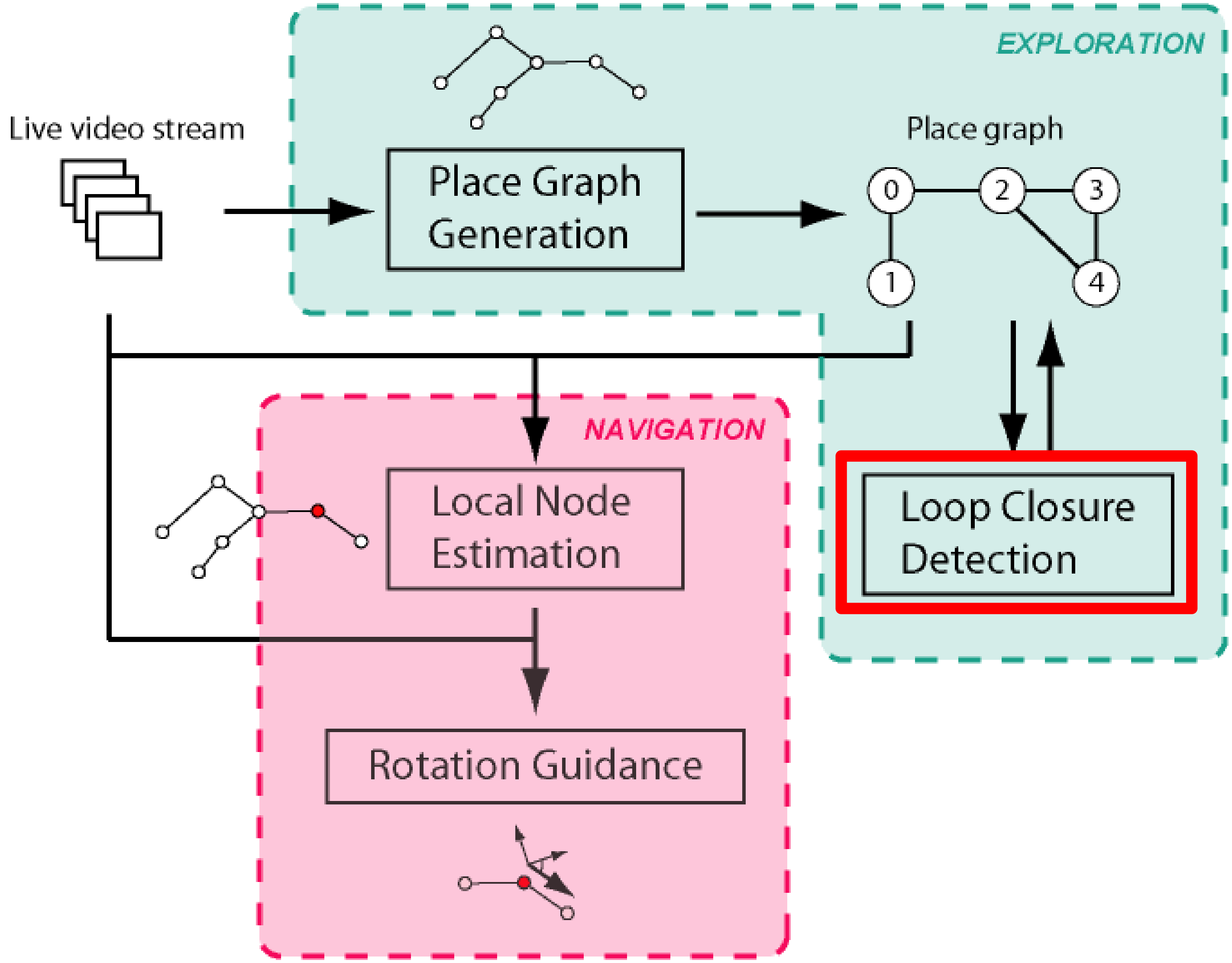
Live video stream **Energy Function**





- Features
- Matched features

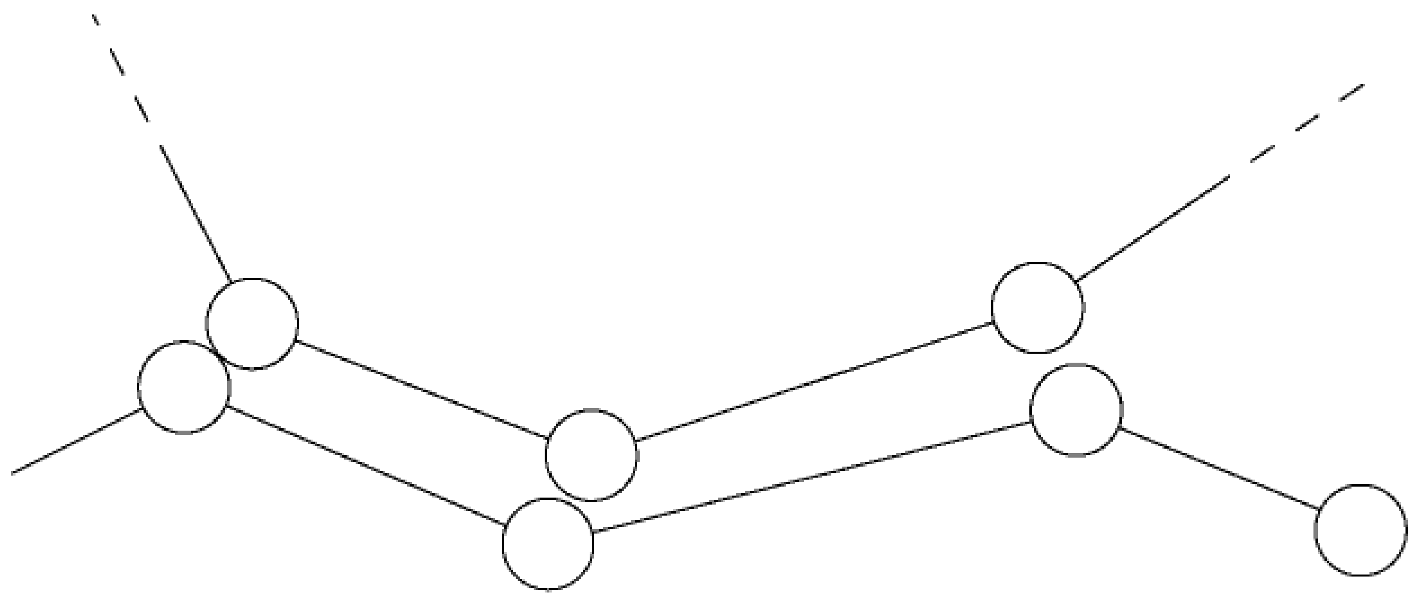
Method Overview



Loop Closure Detection

Approach

1. generate loop closure hypothesis using visual similarity between nodes
2. validate hypothesis using temporal consistency



Bag-of-words

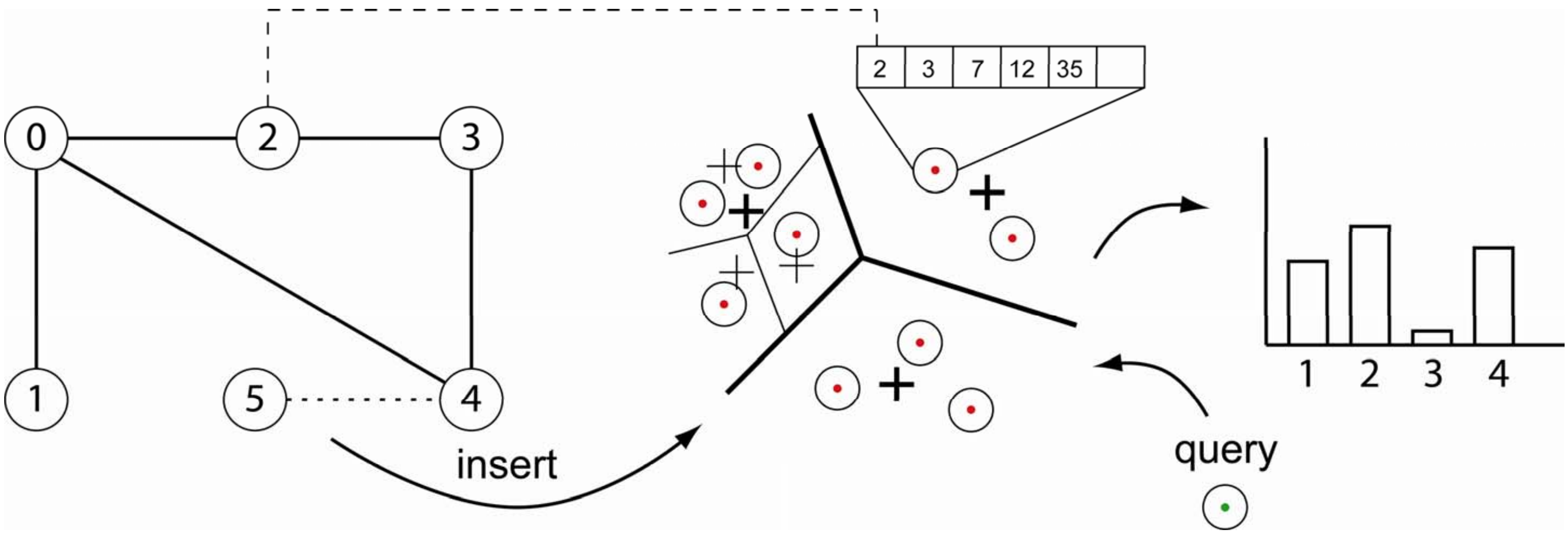
- Bag-of-words: model for data representation

▪ John will fly to San Francisco next week.
▪ Mary would like to fly to the moon with John.
Dictionary: {1: John, 2: will, 3: fly, 4: like, 5: to, 6: moon, 6: Francisco, 7: week, 8: Mary}
▪ [1, 1, 1, 0, 1, 0, 1, 1, 0]
▪ [1, 0, 1, 1, 1, 1, 0, 0, 1]

- Image="document", image feature="word"
- Applications
  - object categorization (generative/discriminative models)
  - object segmentation, localization

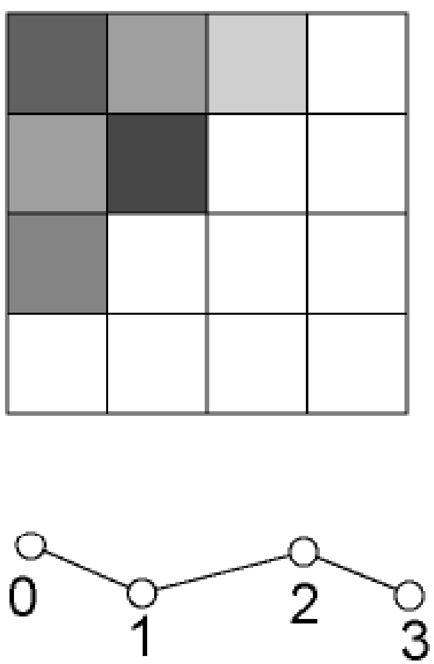
Loop Closure Detection

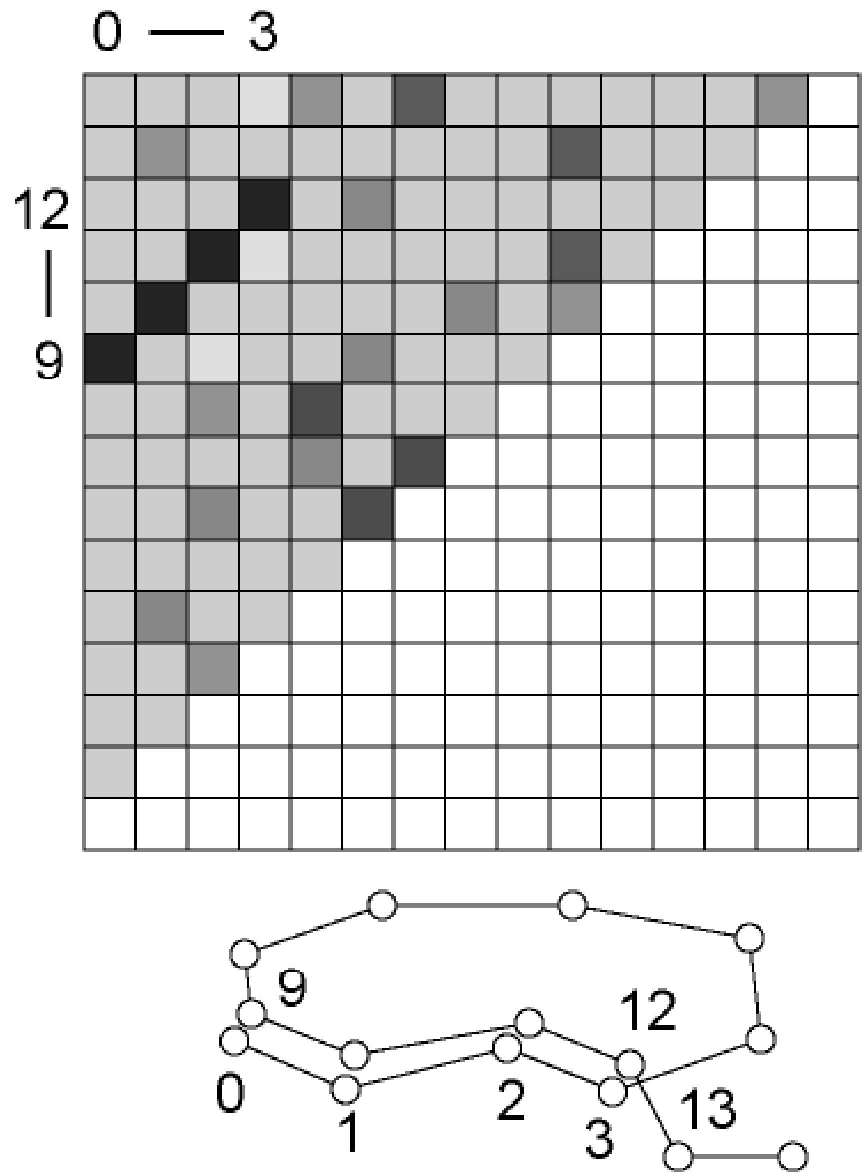
- Node="document", image feature="word"



Loop Closure Detection

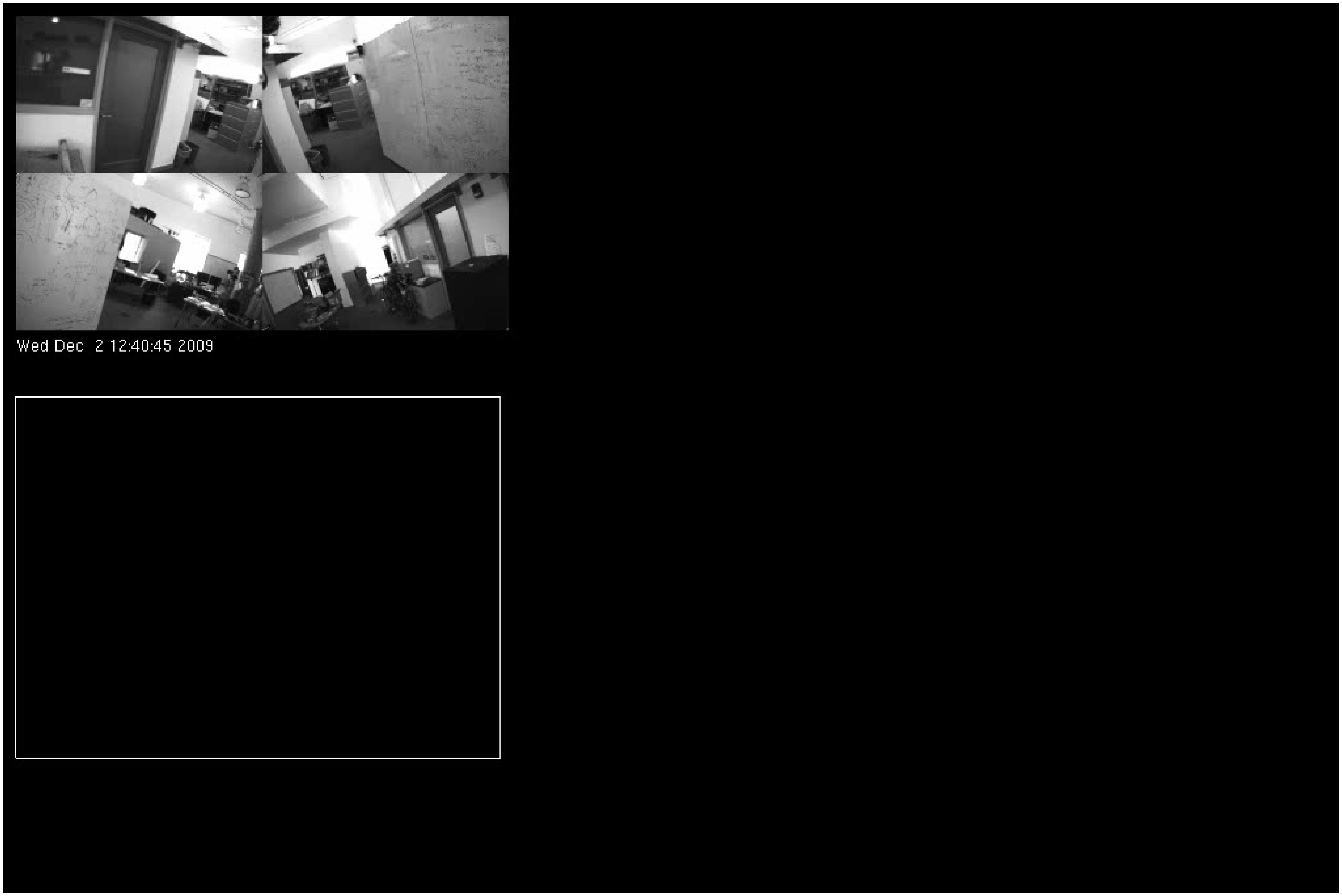
- Similarity matrix





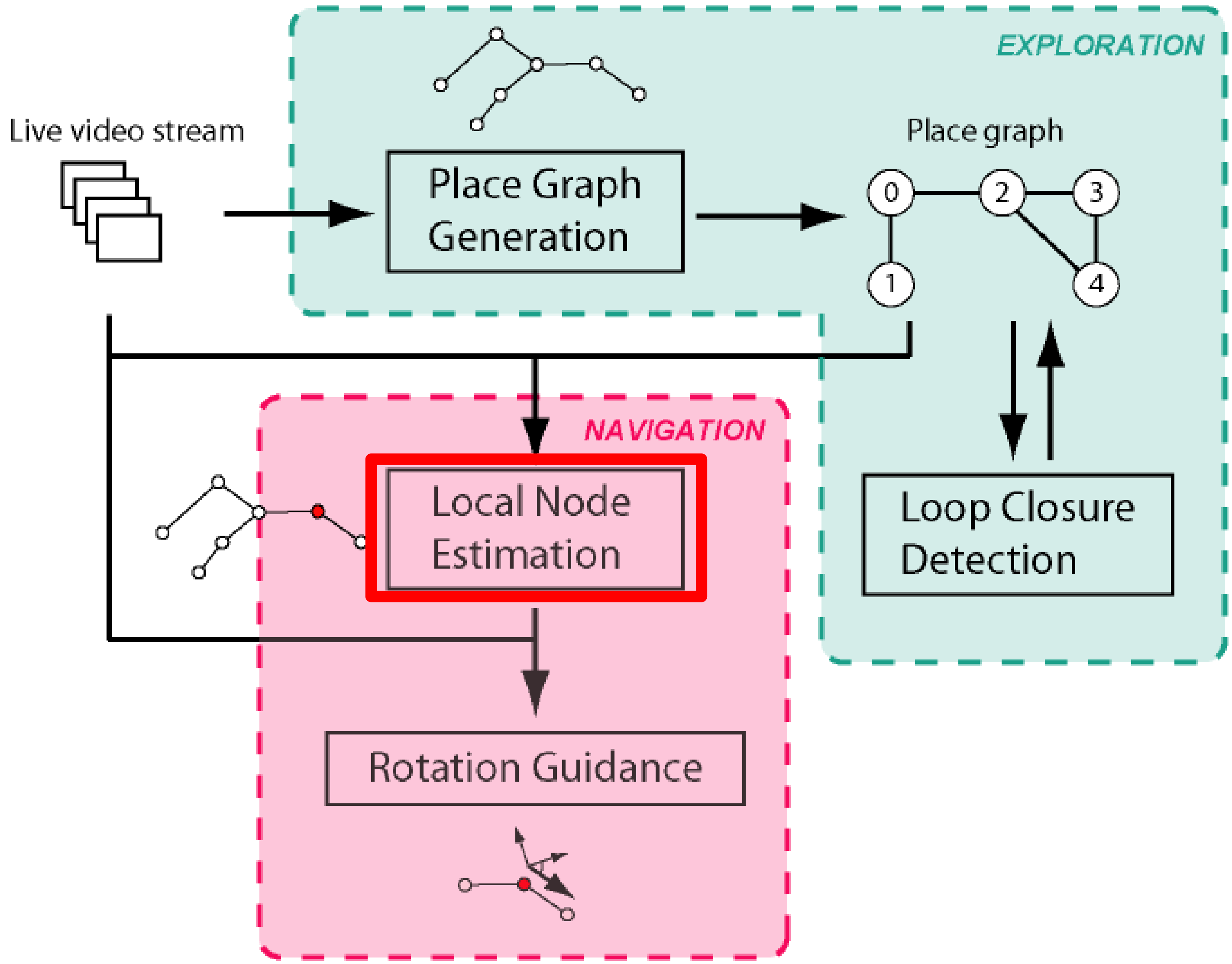
Loop Closure Detection

Live video stream



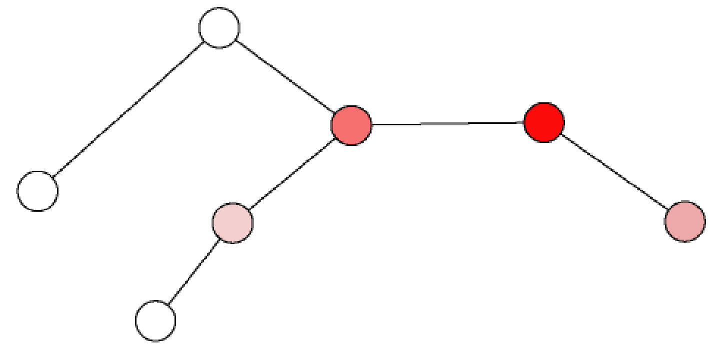
Place Graph ***Similarity Matrix***

Method Overview



Local Node Estimation

Discrete recursive Bayesian filtering



$p(x_k \mid z_k)$: probability of state x_k while observing z_k at time k

1. Prediction step

$$p(x_{k+1} \mid z_k) = \sum_k p(x_{k+1} \mid x_k) p(x_k \mid z_k)$$

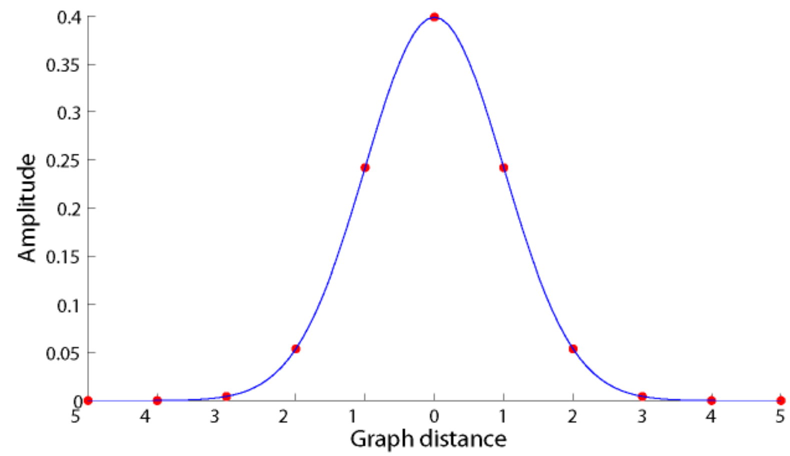
2. Observation step

$$p(x_{k+1} \mid z_{k+1}) = \lambda p(z_{k+1} \mid x_{k+1}) p(x_{k+1} \mid z_k)$$

Local Node Estimation

1. Prediction step

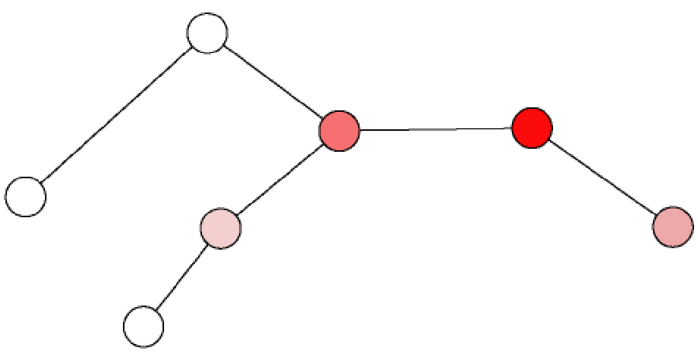
$p(x_{k+1} \mid x_k)$: motion model



Gauss Window Function

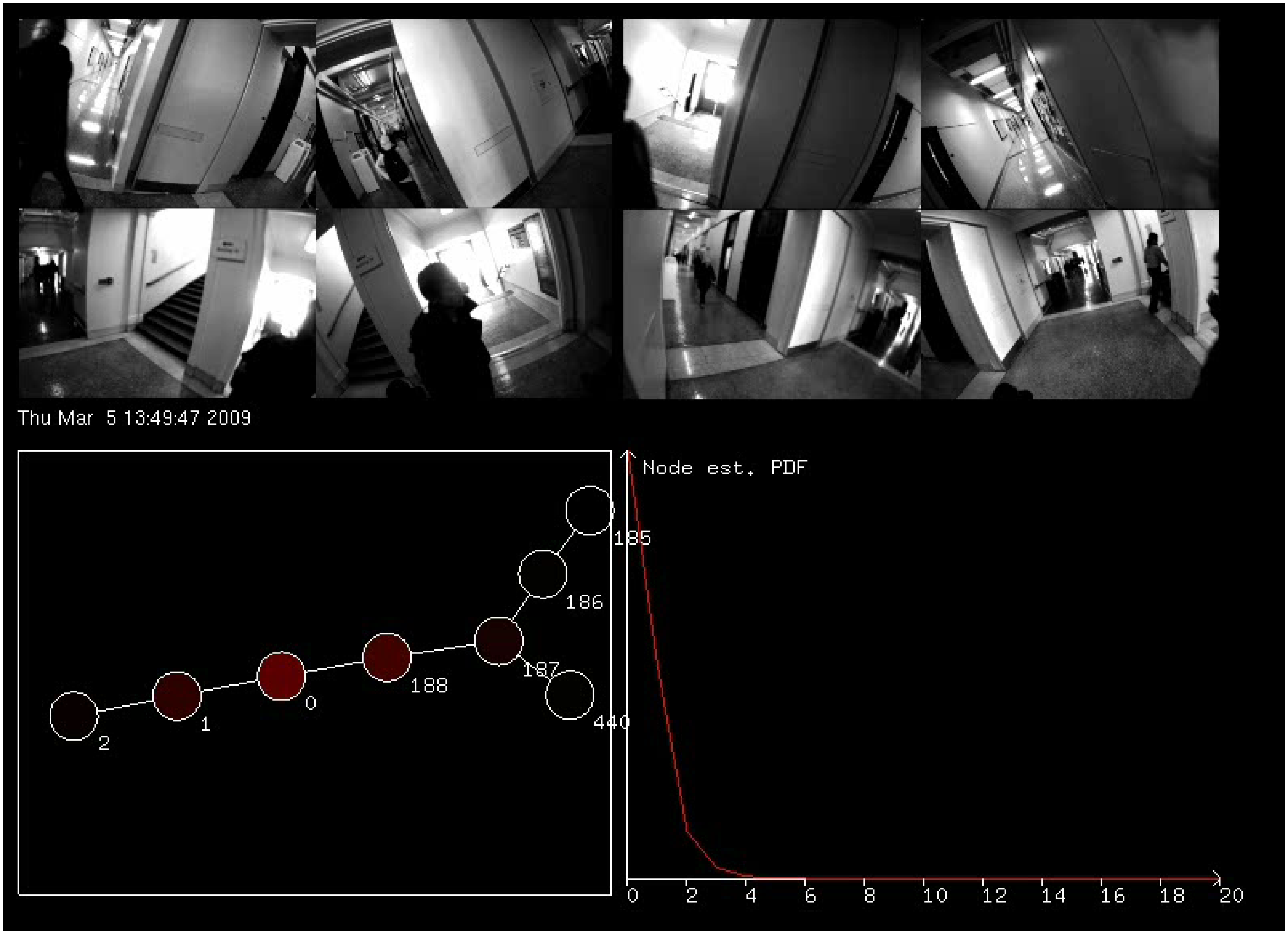
2. Observation step

$p(z_{k+1} \mid x_{k+1})$: observation model



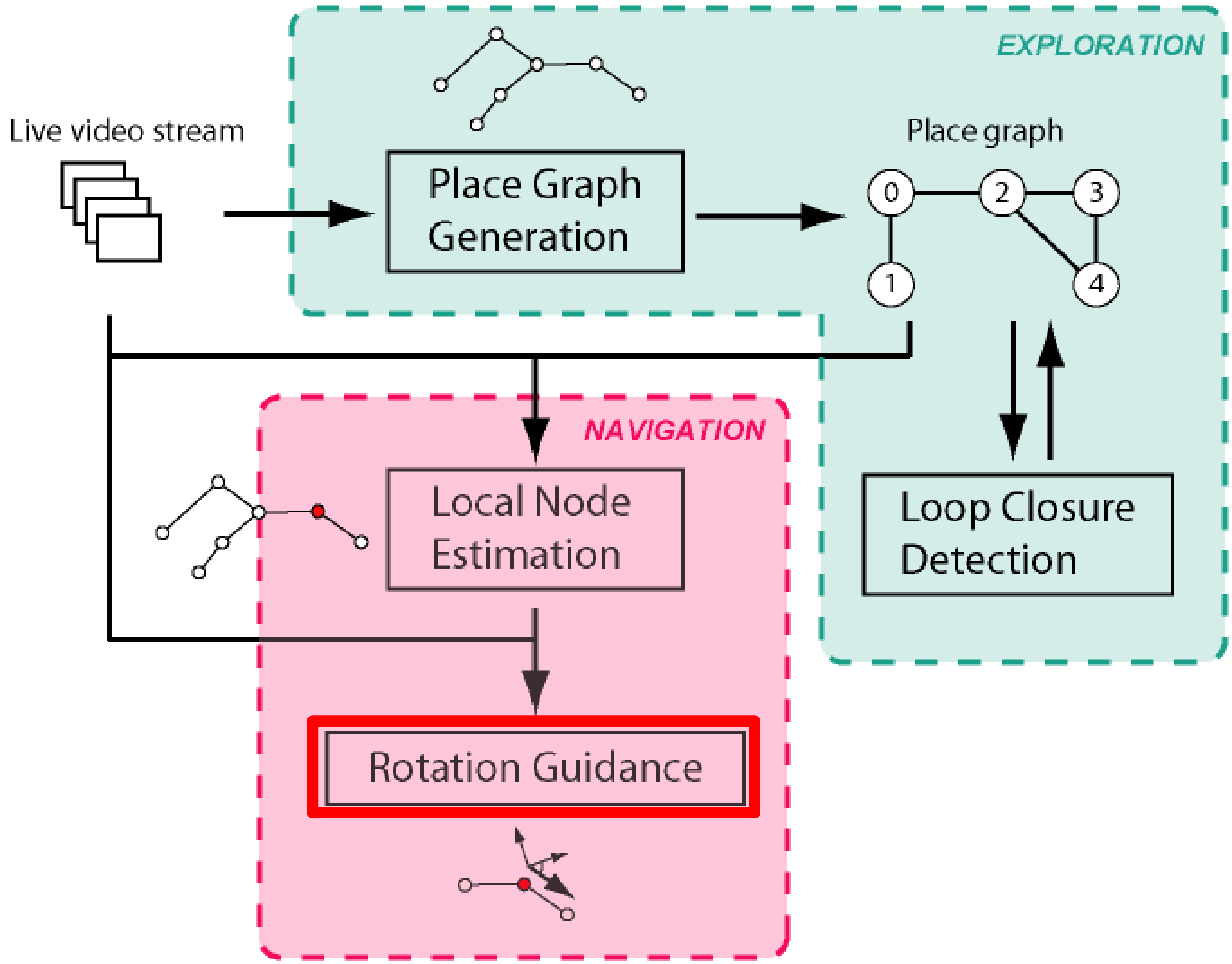
Local Node Estimation

Live video stream ***Localization***



Place Graph ***Node Estimation PDF***

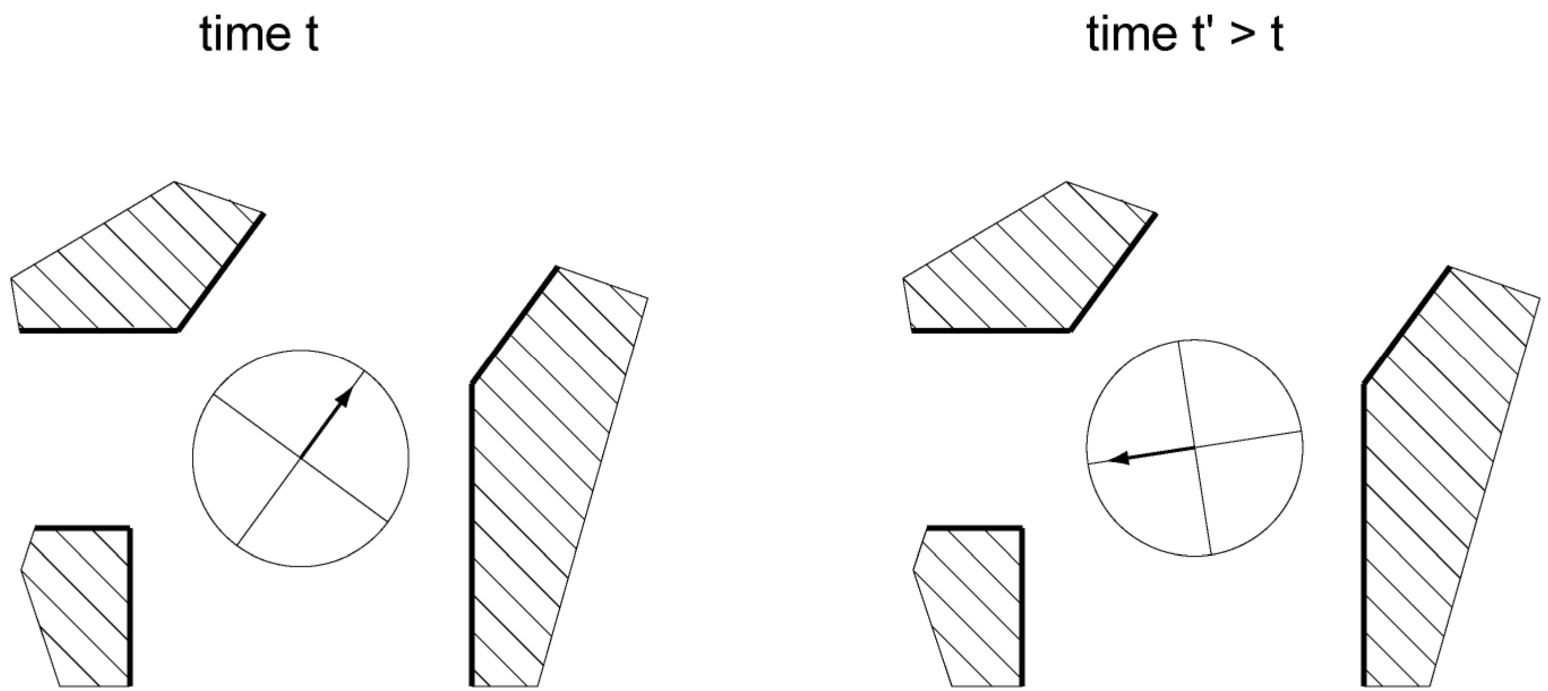
Method Overview



Rotation Guidance

Relative orientation problem

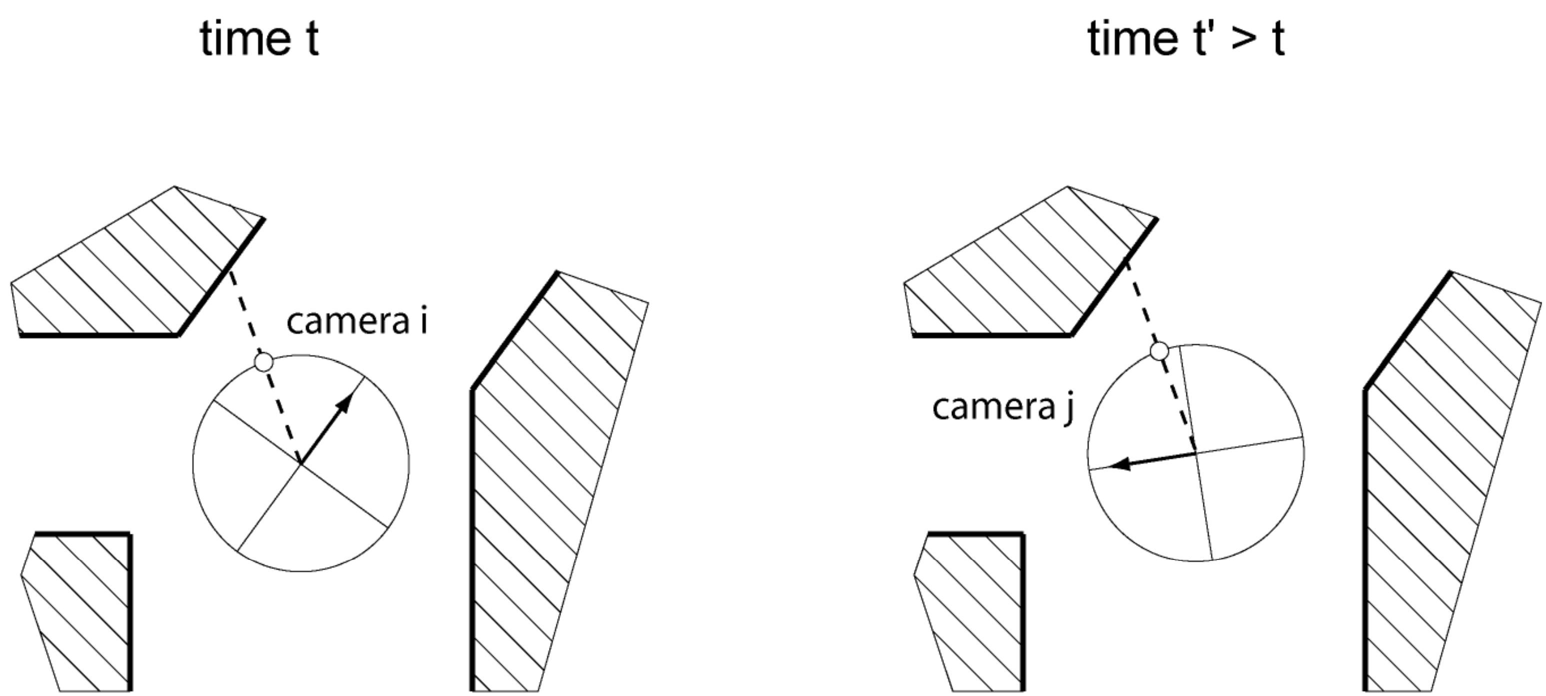
Determine the orientation of the user at time t' *relative to* the orientation at time t.



Rotation Guidance

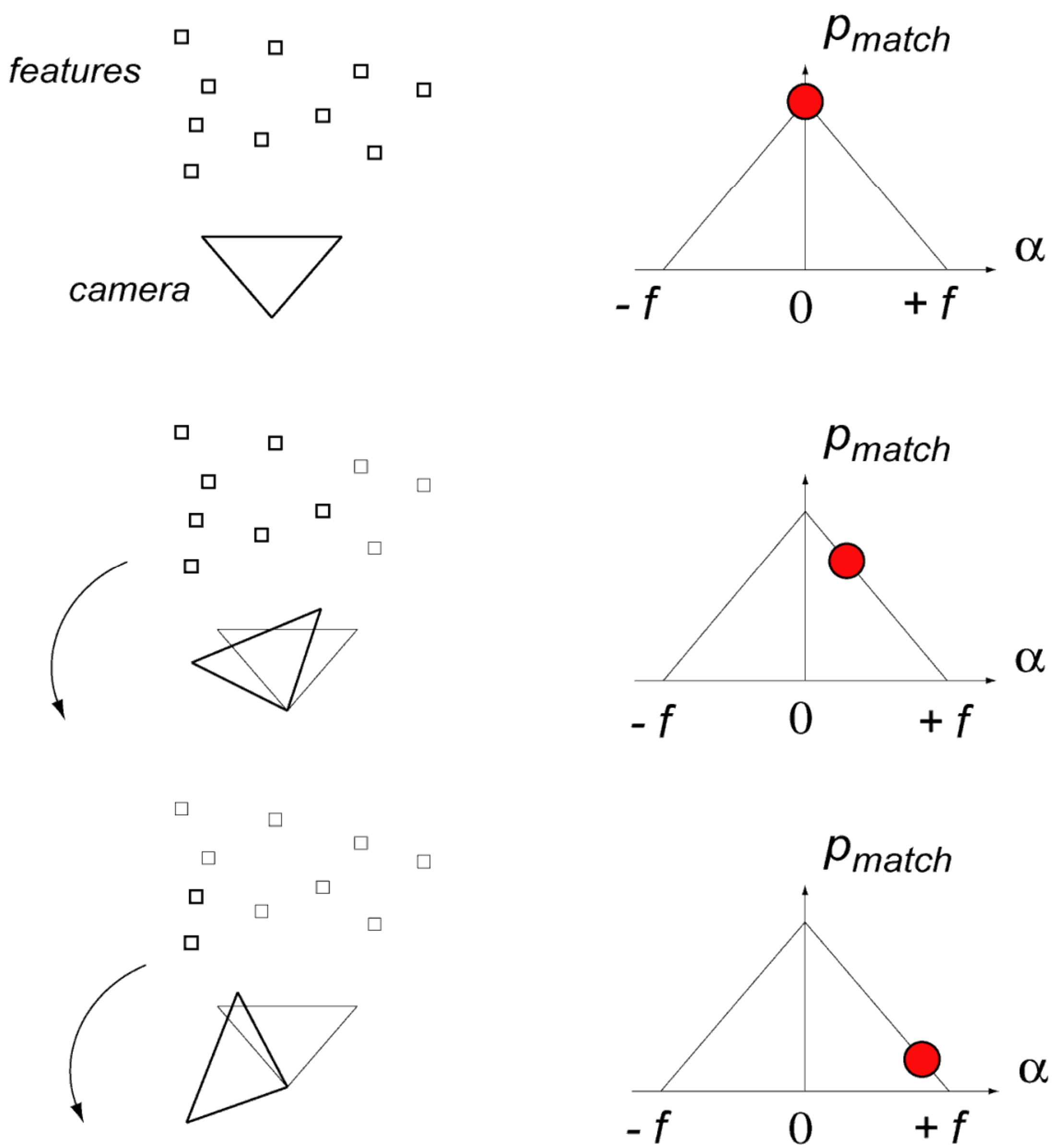
Idea

Learn the relationship between user rotation and feature matches across cameras.



Feature matching with one camera

Rotation Guidance



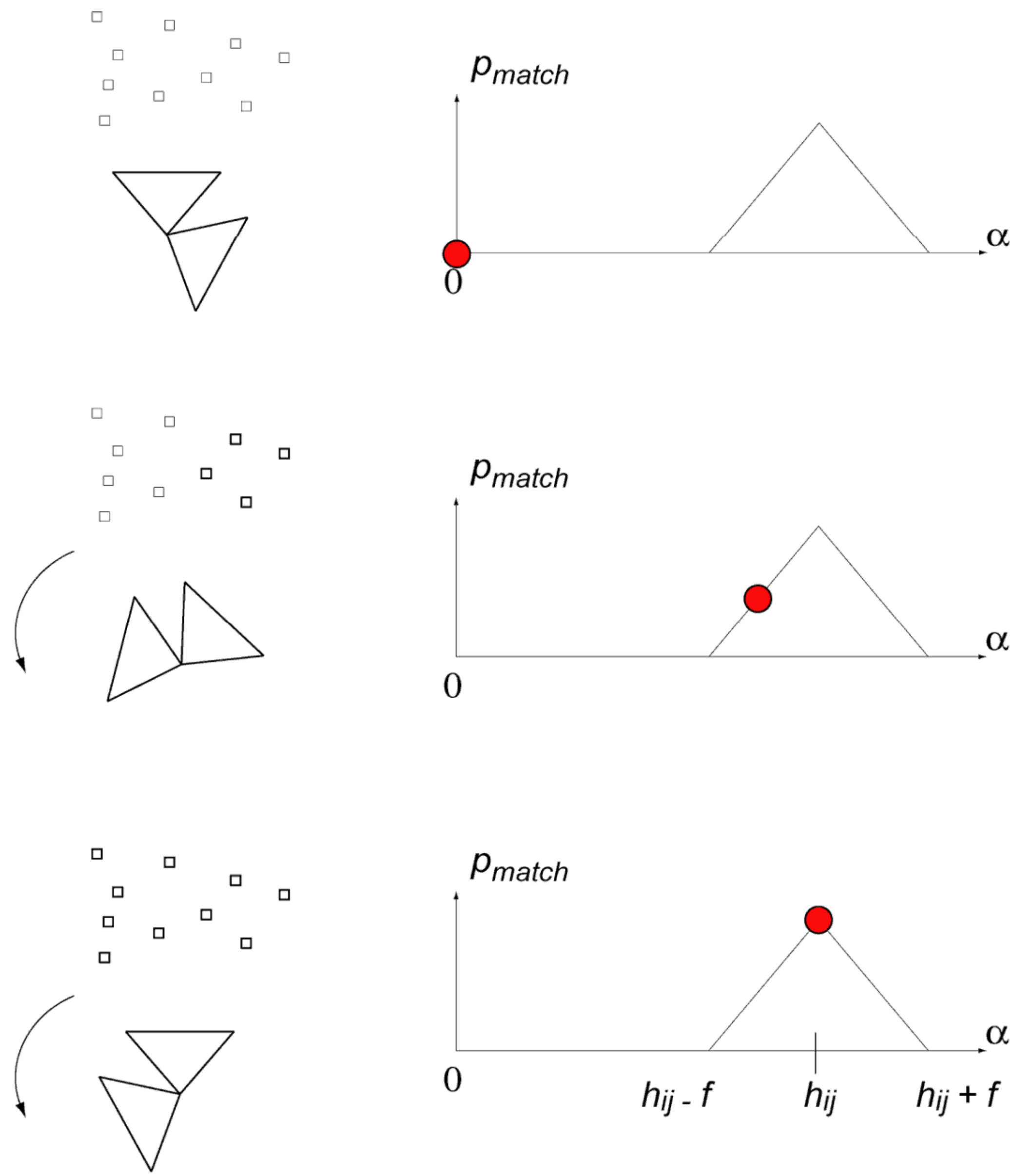
α: rotation angle

f: camera field of view

Assumption: features are uniformly distributed in image space.

Feature matching with two cameras

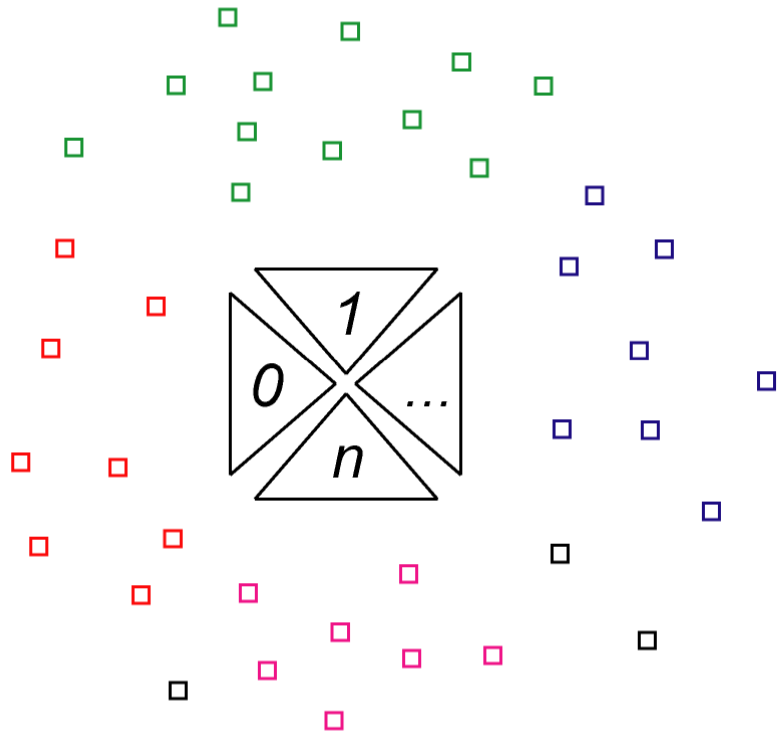
Rotation Guidance



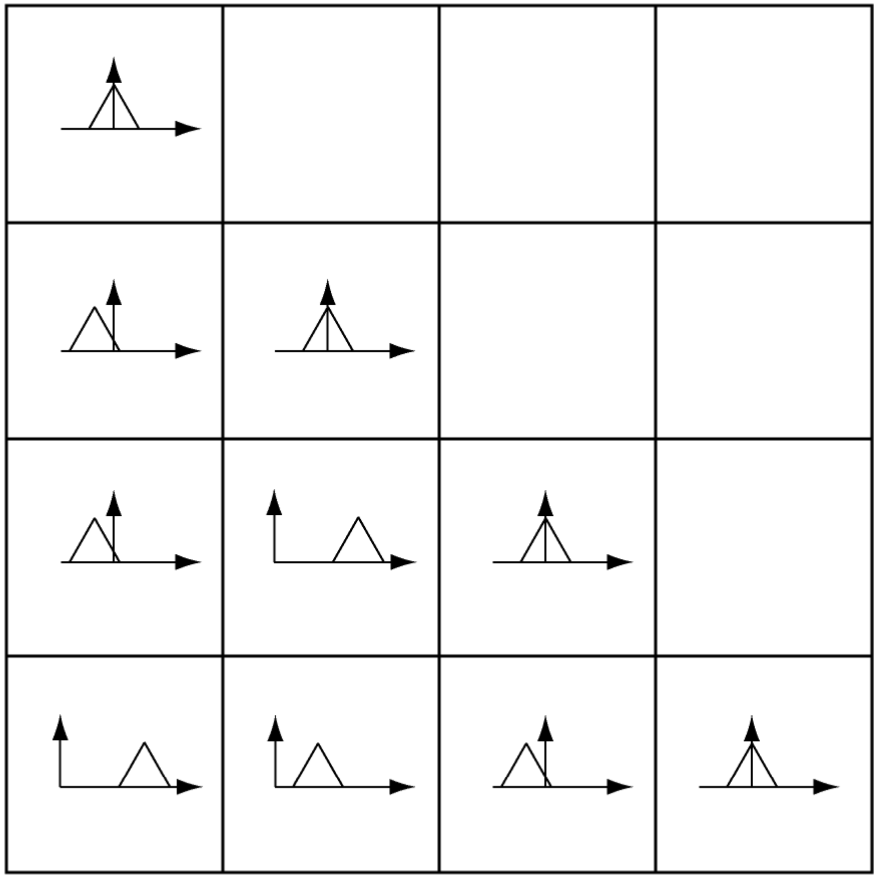
Assumption: cameras have the same field-of-view.

Feature matching with *n* cameras

Rotation Guidance

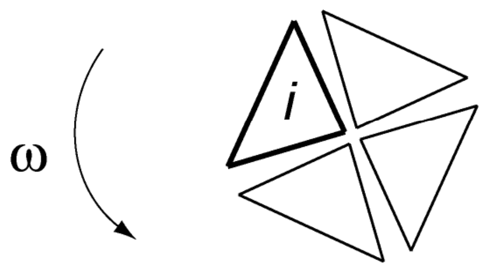


Match matrix (n x n)



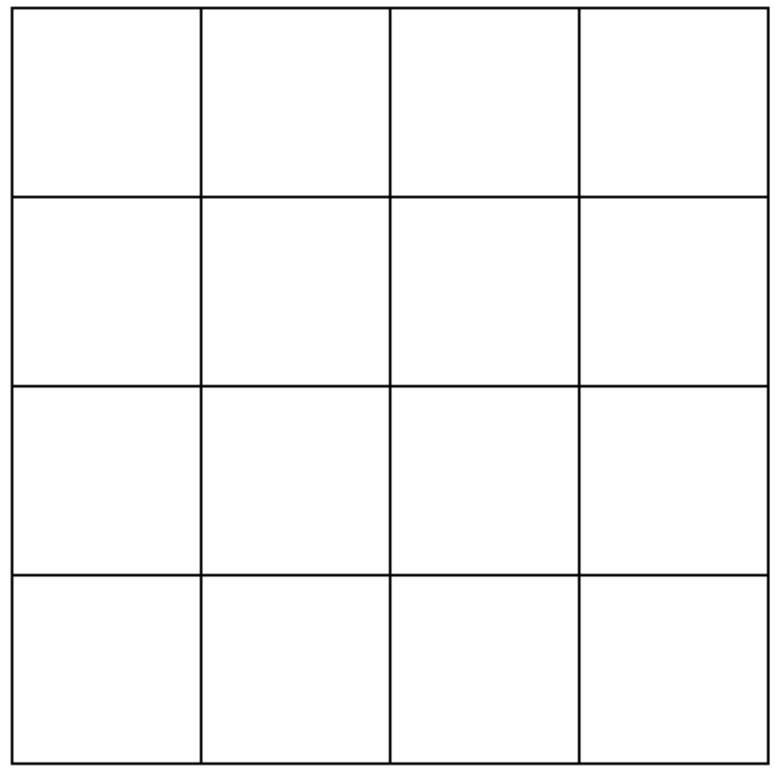
Learning the match matrix from training

System rotating in place at constant speed ω



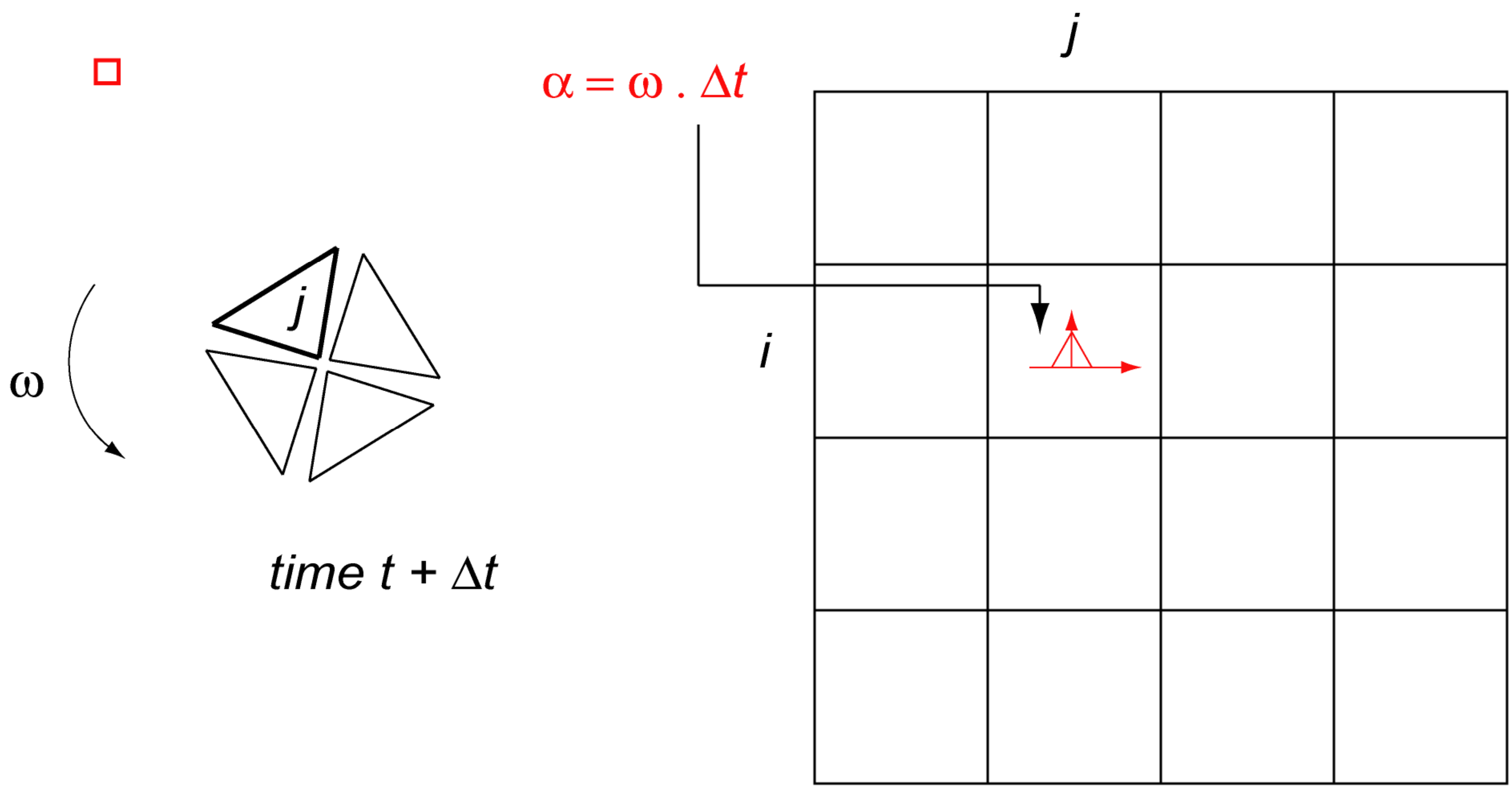
time t

Match matrix



□ *feature observed on camera i at time t*

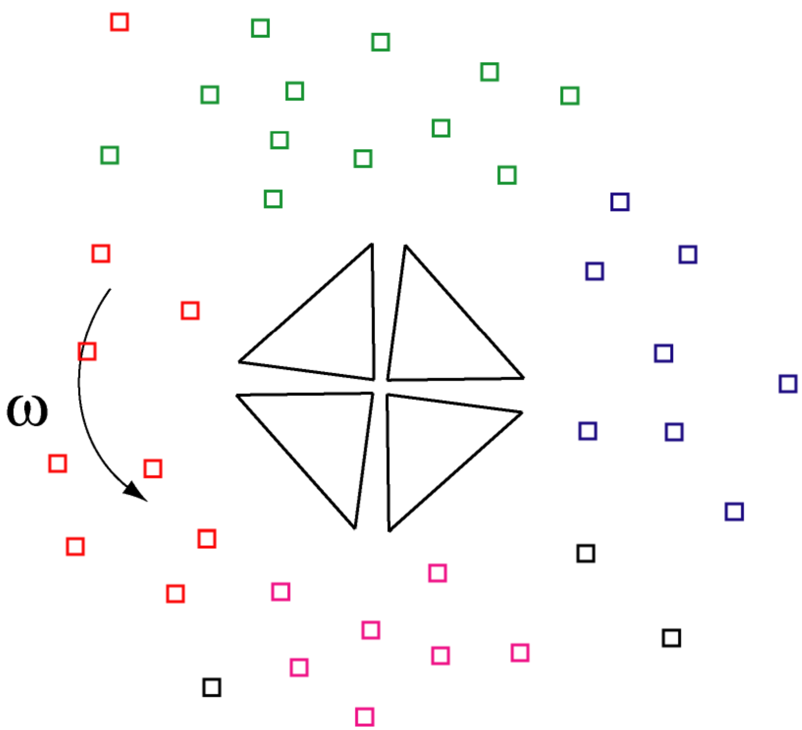
Learning the match matrix from training



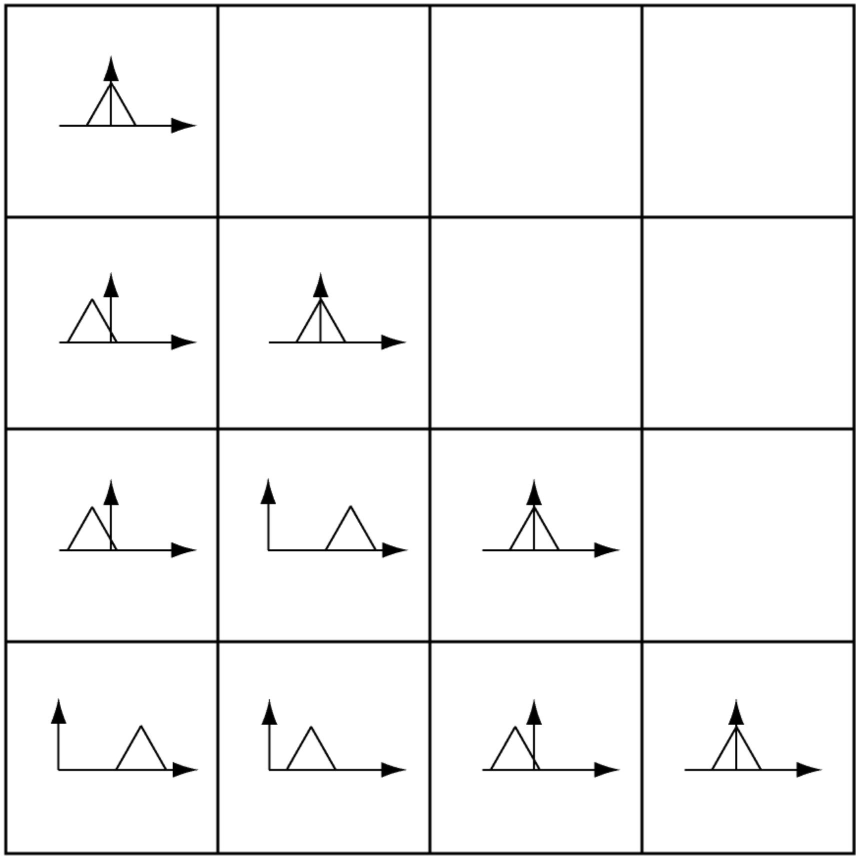
- (same) feature observed on camera *j* at time $t + \Delta t$

Learning the match matrix from training

Repeat for all features and all pairs of image frames (time sampling)



Match matrix (n x n)



Rotation Guidance

Match Matrix

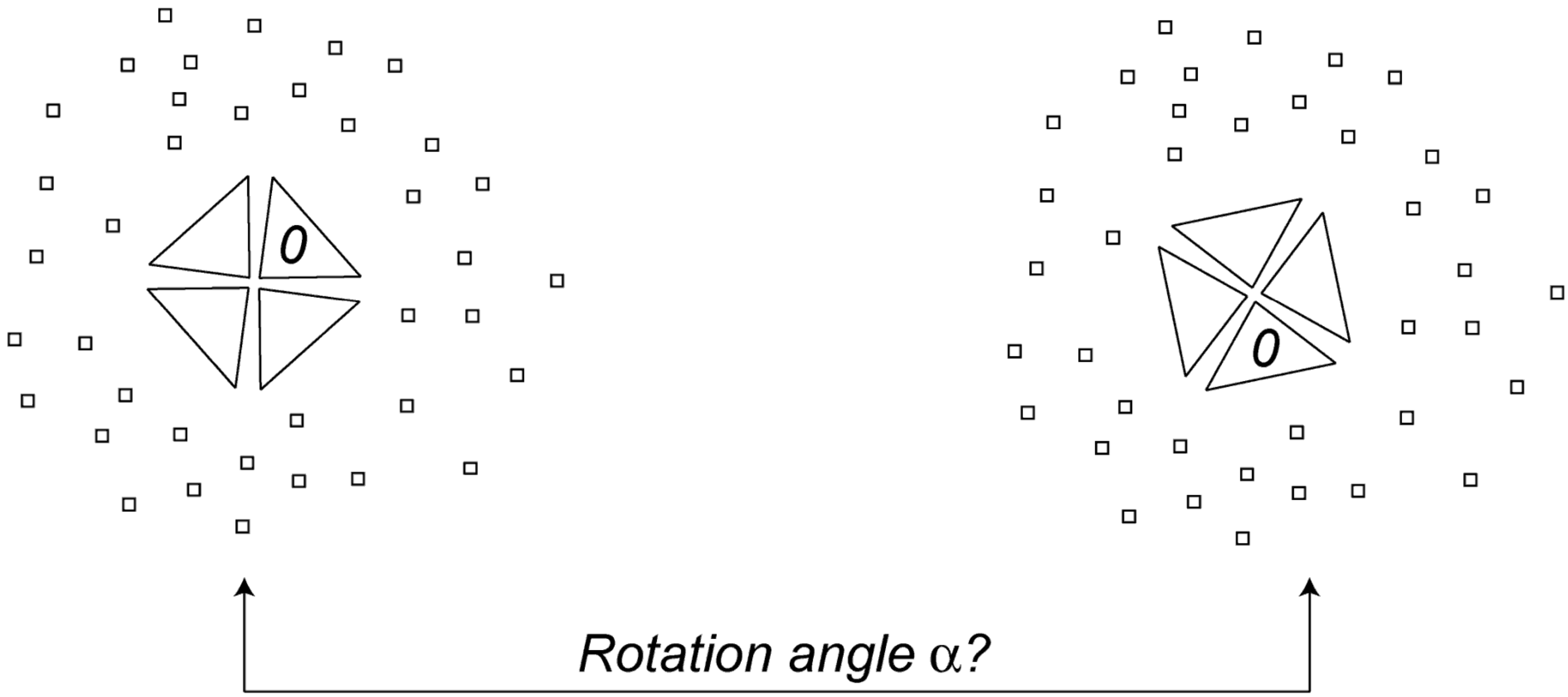
Live video stream

Close-up view of matrix cell

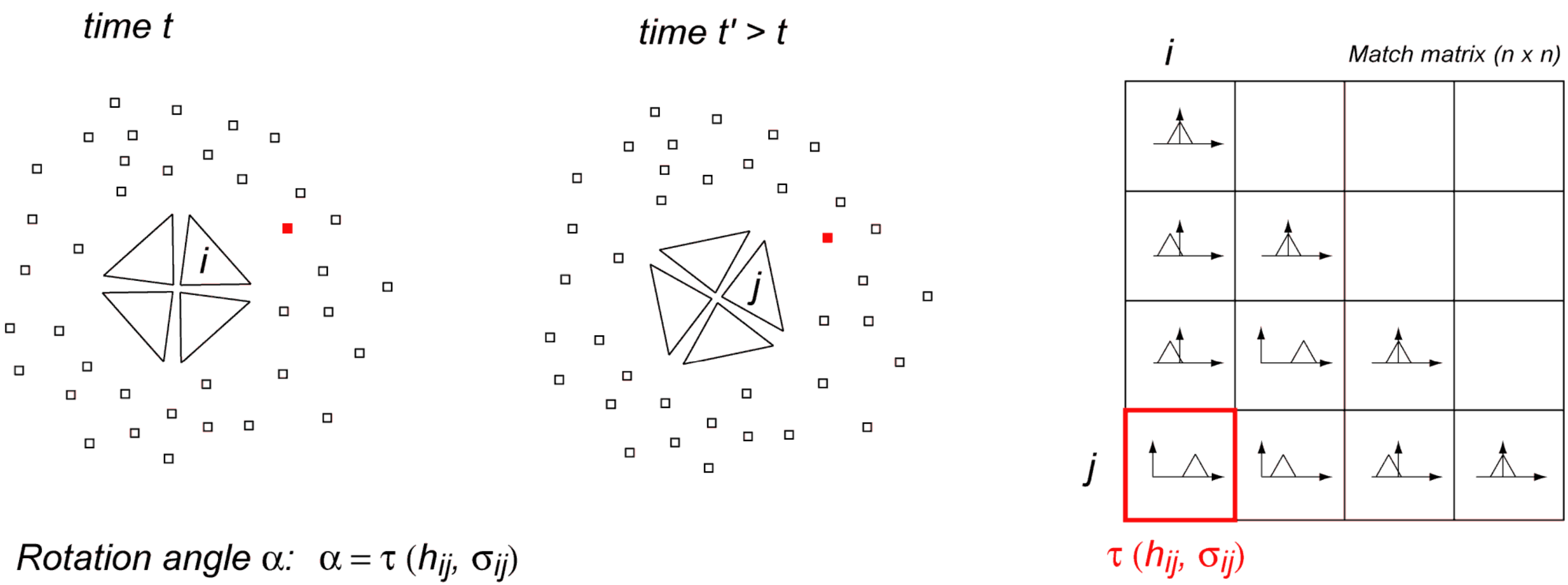
Relative orientation problem

Relative orientation problem

Determine the orientation of the user at time t' *relative to* the orientation at time t.



Orientation from a single correspondence

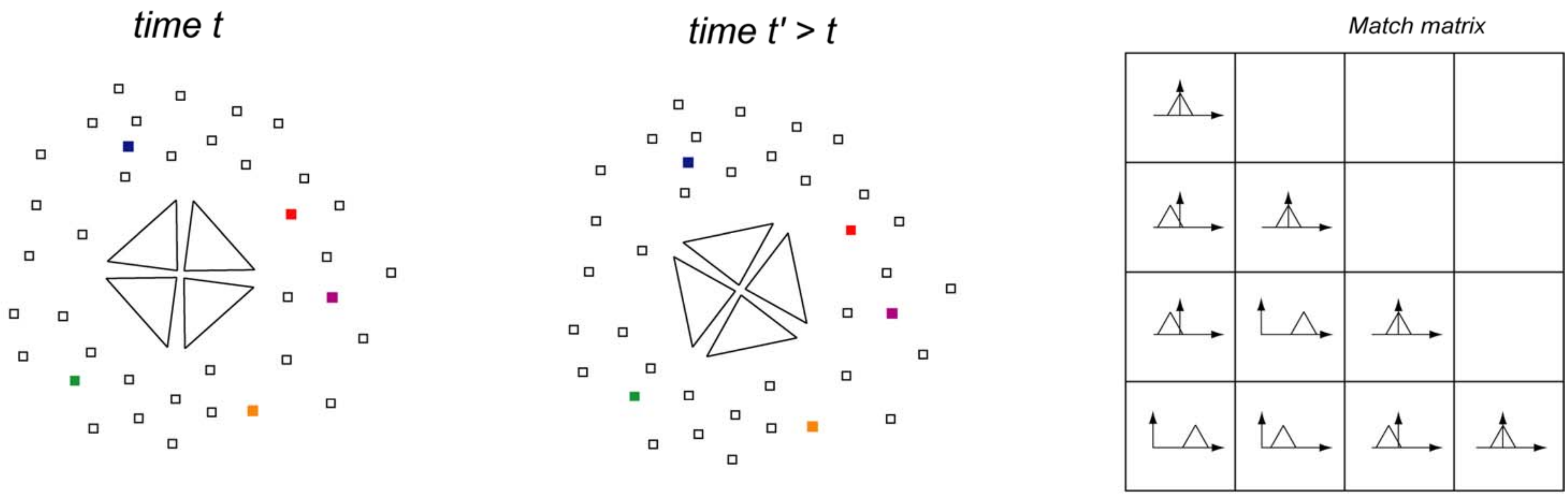


Rotation angle α: $\alpha = \tau\,(h_{ij}, \sigma_{ij})$

Estimate of α: $\bar{\alpha} = h_{ij}$

Error: $\varepsilon = \bar{\alpha} - \alpha = \tau(0, \sigma_{ij}) = \tau(0, \sigma)$

Orientation from N correspondences



Estimate of α: $\bar{\alpha} = \frac{1}{N} \sum^{N} h_{ij}$ (N observations)

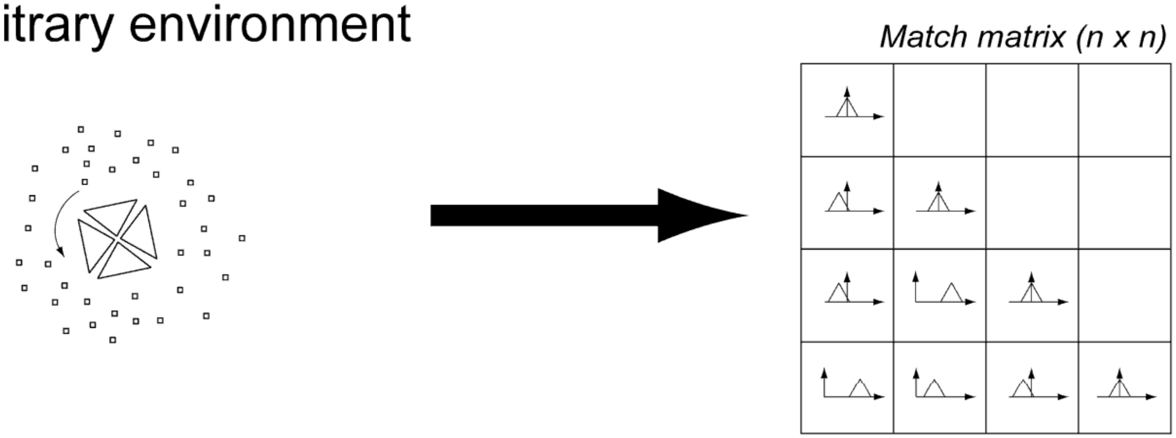
Error: $\varepsilon = \bar{\alpha} - \alpha = \frac{1}{N} \sum^{N} \tau(0,\sigma)$ $\xrightarrow[N \longrightarrow \infty]{\text{CLT}}$ $\varepsilon \sim N(0, \sigma_N)$ $\sigma_N = \frac{\sigma}{\sqrt{N}}$

$f = 90°$ $\sigma^2 = f^2/6$ $\sigma = 36.7°$ $N = 200$ $\sigma_N = 2.6°$

Rotation Guidance

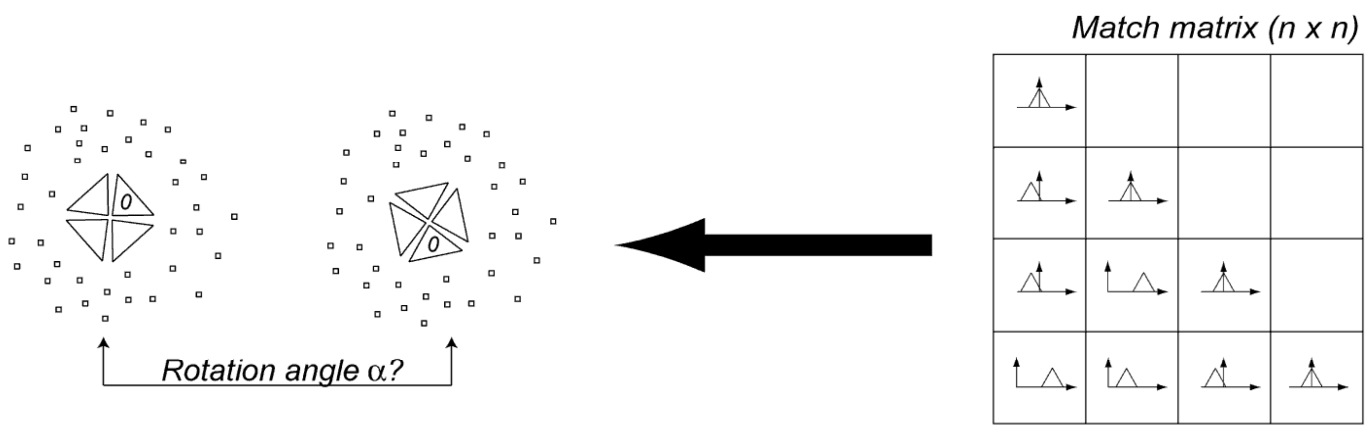
Training

- Learn match matrix from training data
- Once and for all for a given camera configuration
- Arbitrary environment



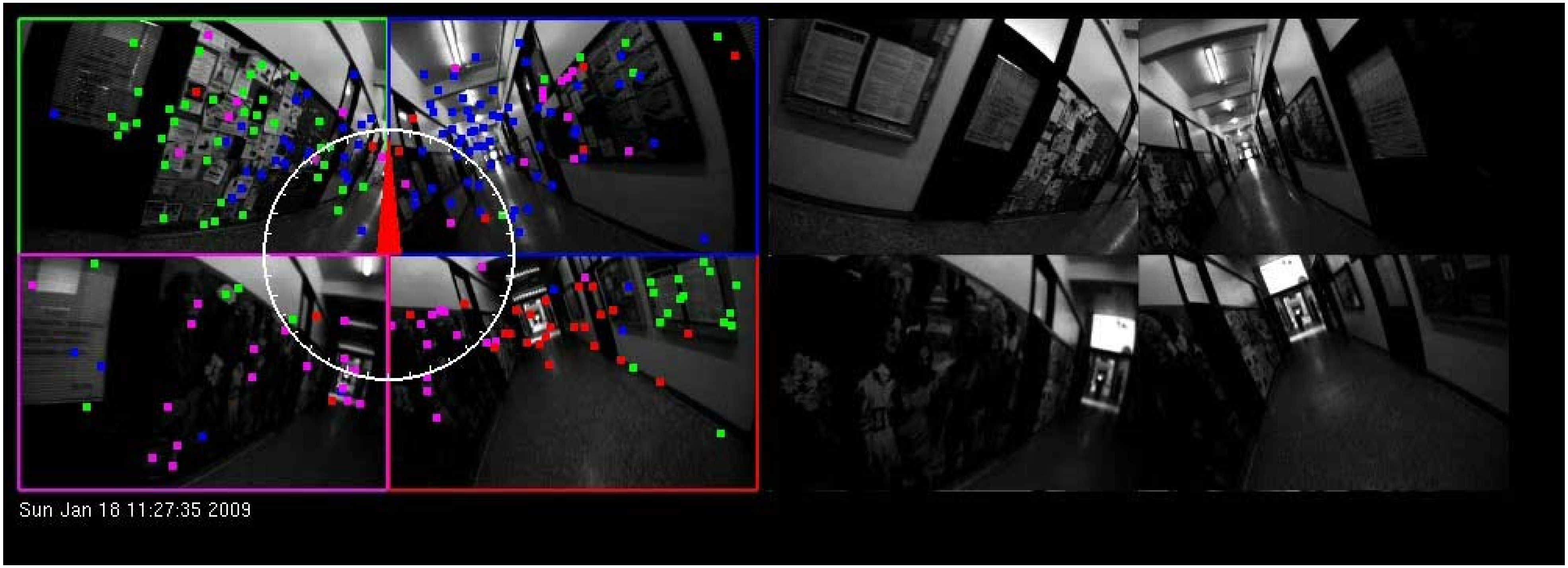
Navigation

- Use match matrix to estimate *relative* user orientation



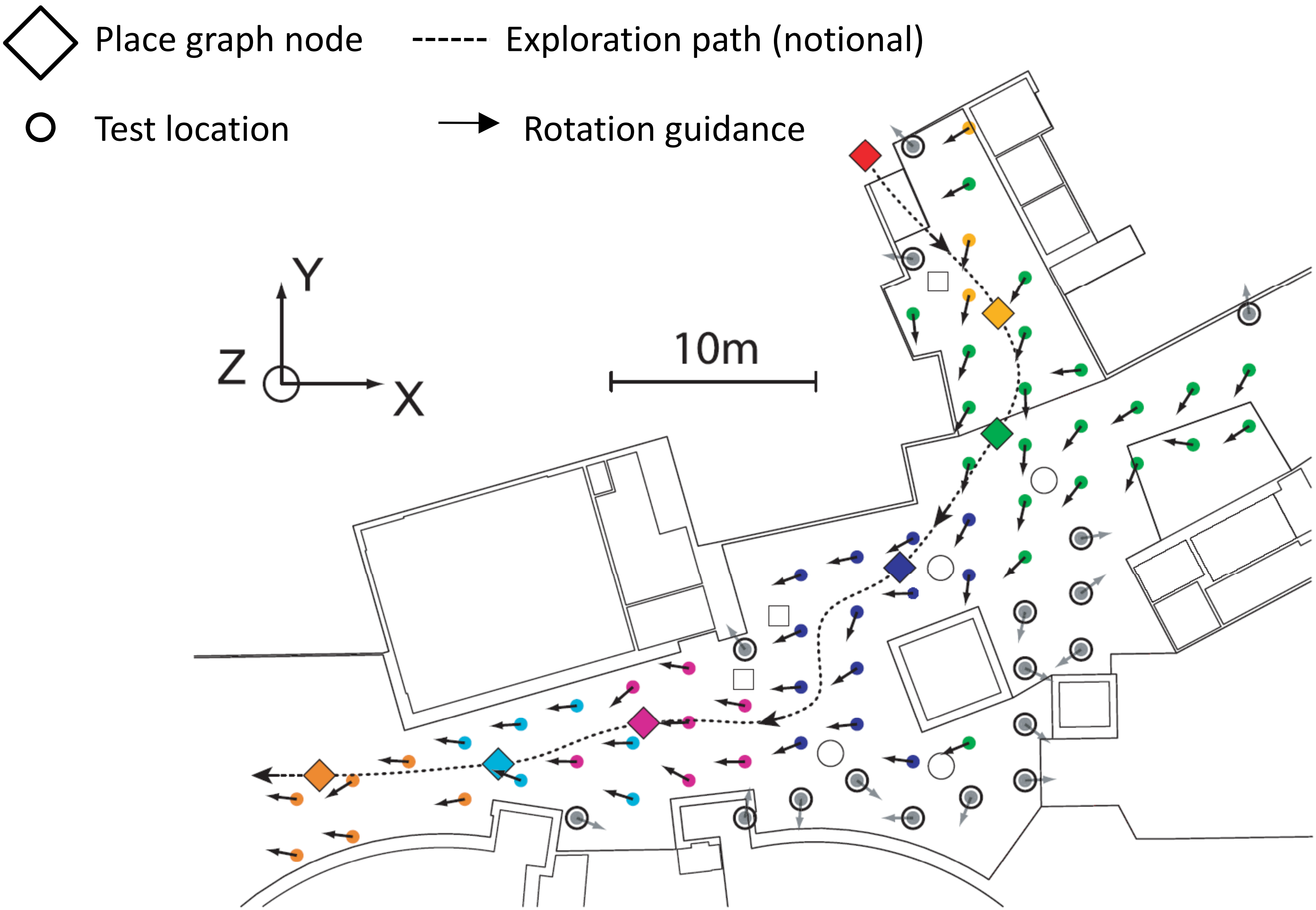
Rotation Guidance: Video

Live video stream ***Localization***

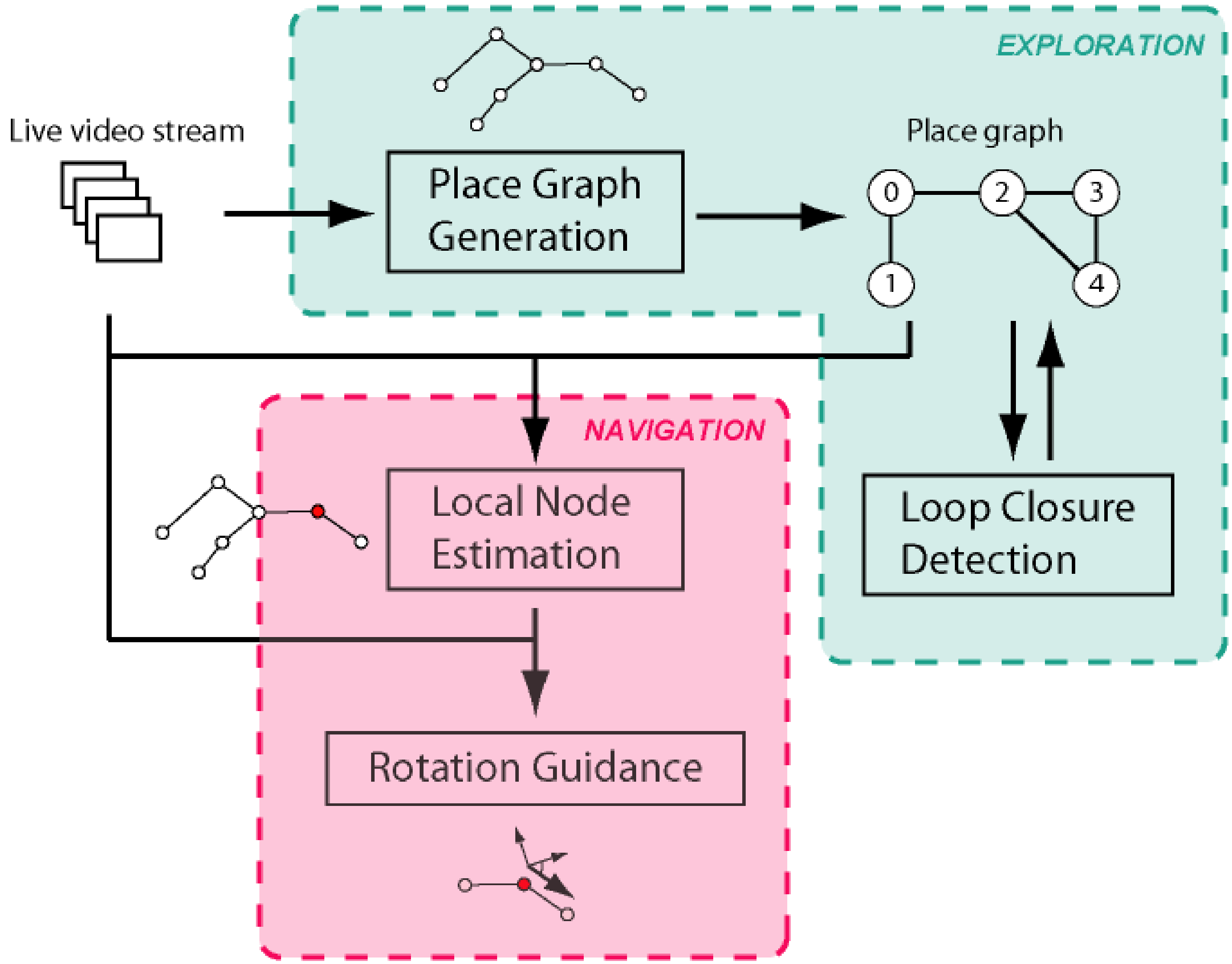


Features

Rotation Guidance

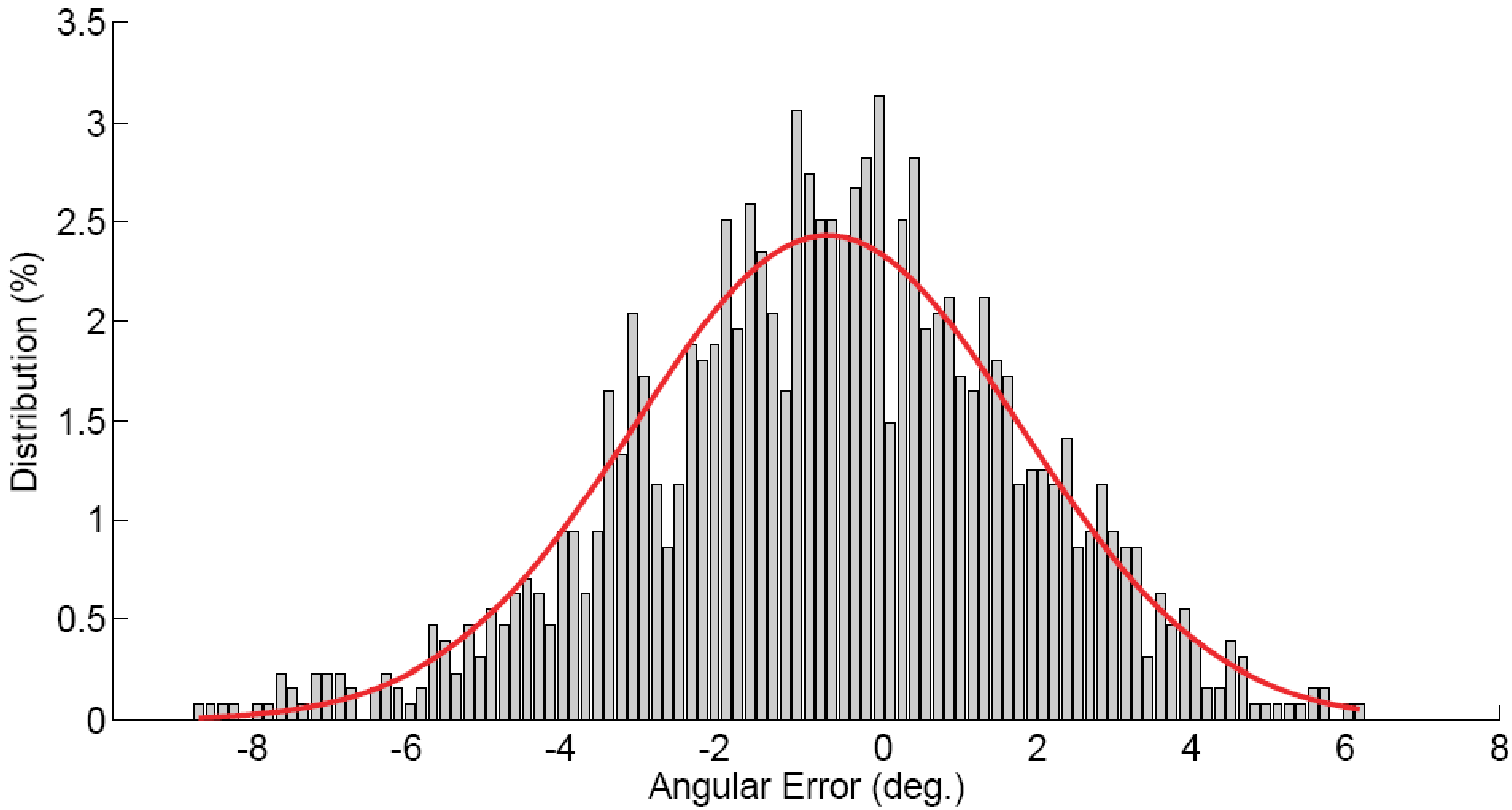


Method Overview



Method Evaluation

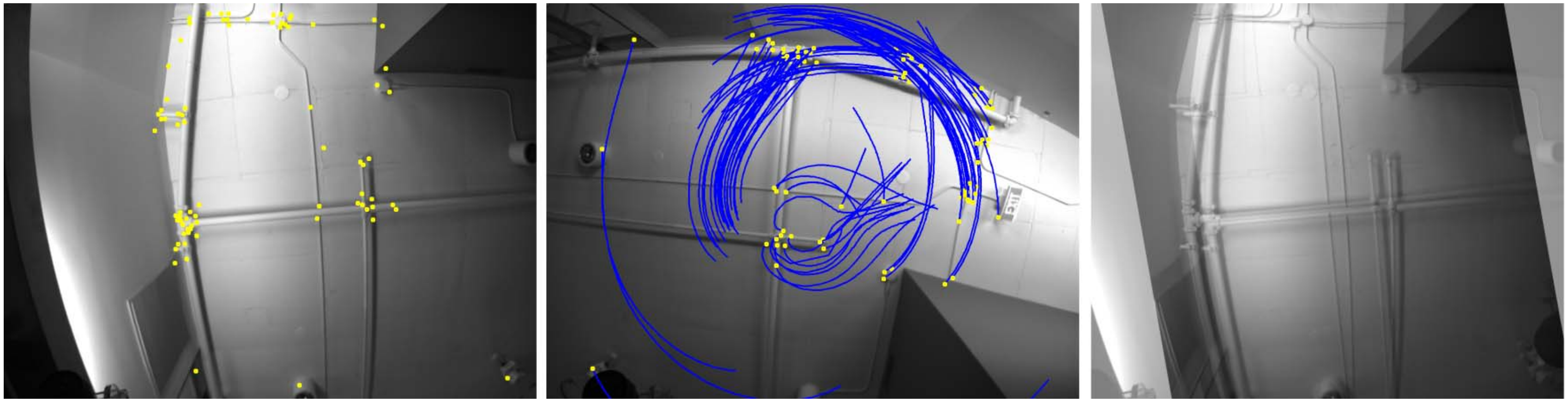
Rotation Guidance evaluation using IMU-derived ground truth (30 sec. sequence)

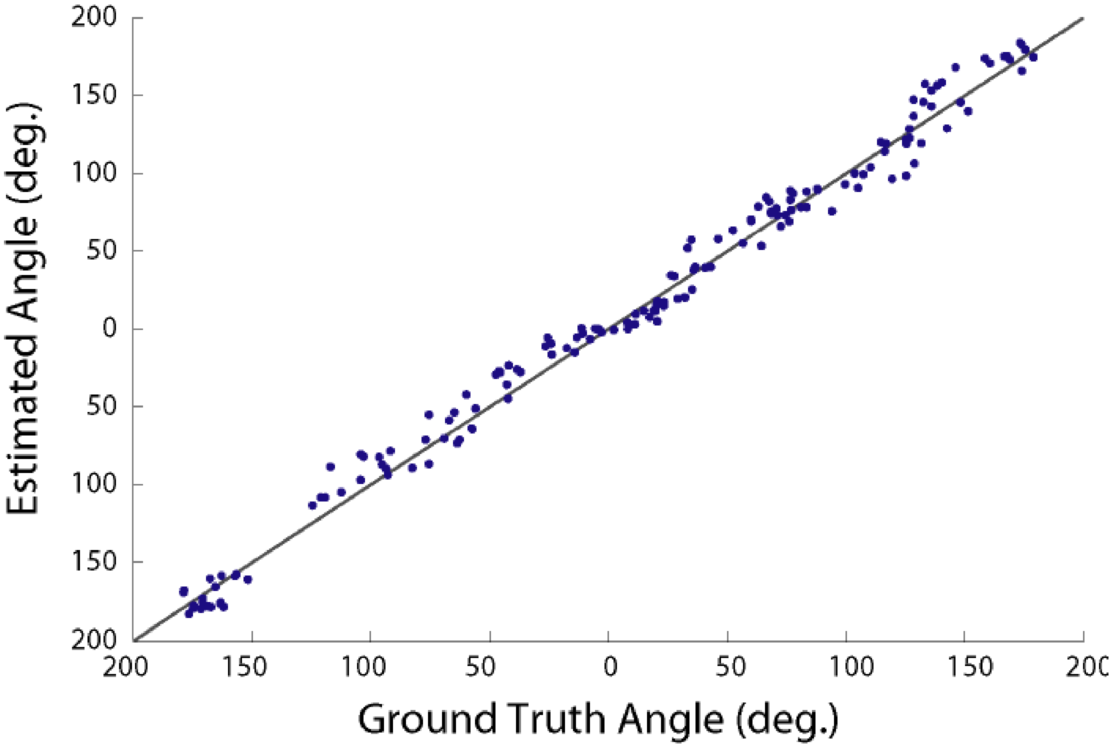


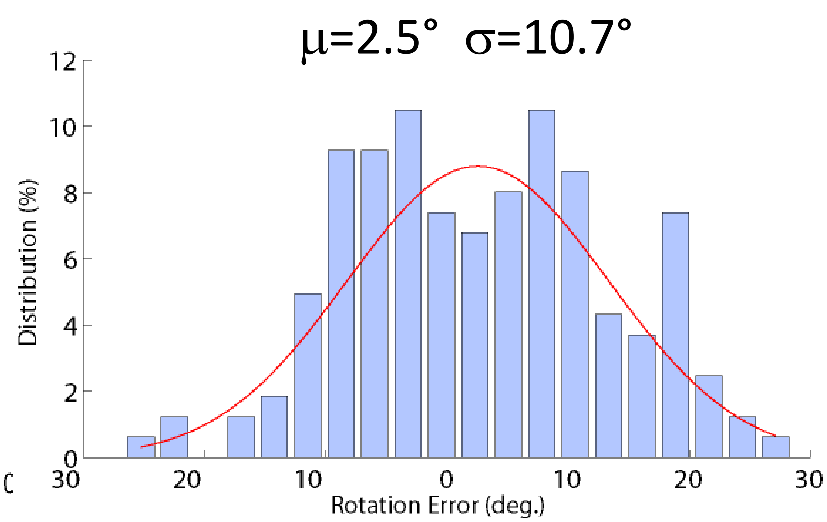
μ = 0.7° σ = 2.5° max = 8°

Method Evaluation

Large-scale rotation ground-truth using upward images

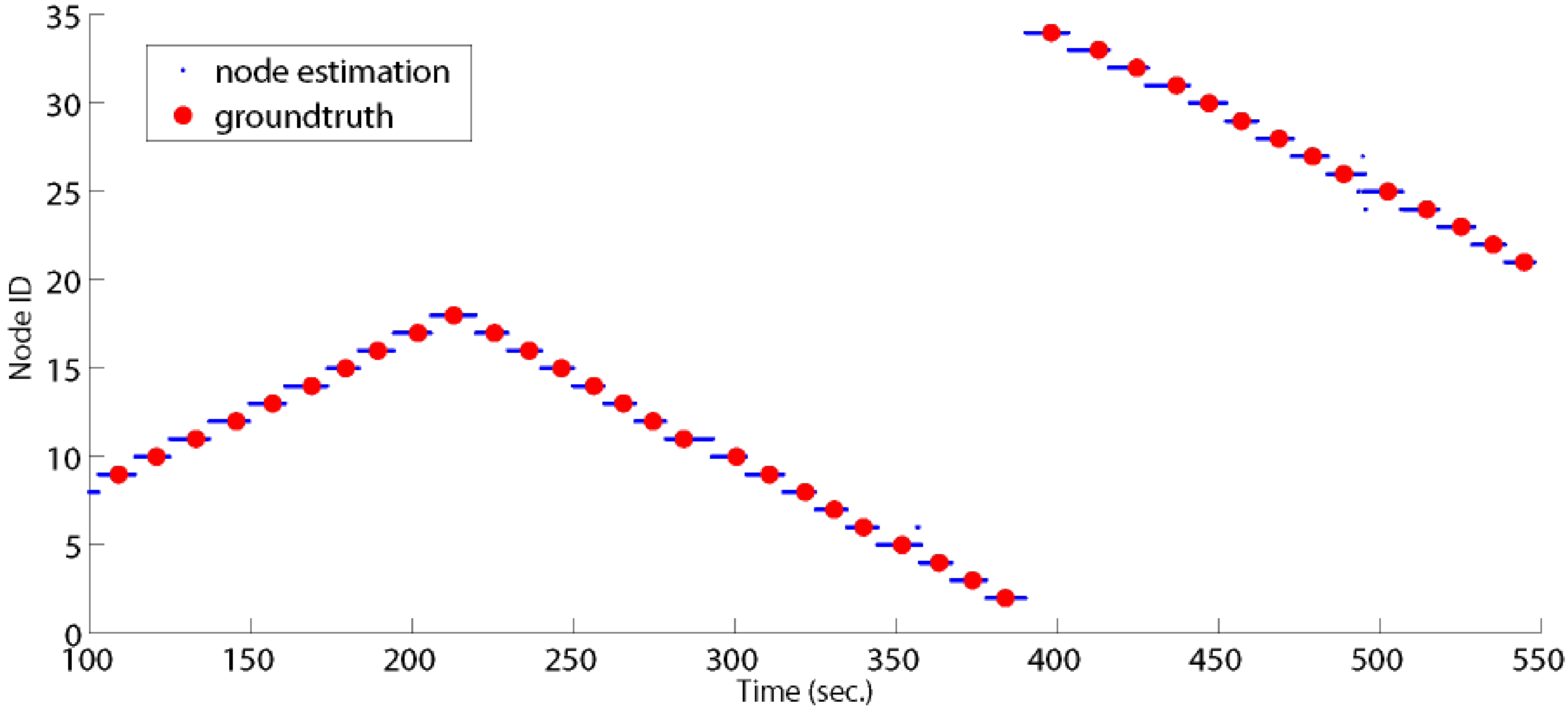




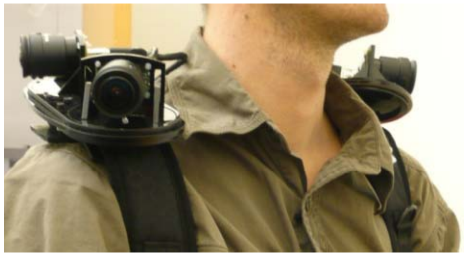


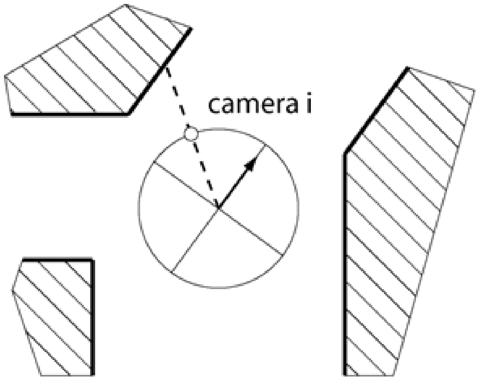
Method Evaluation

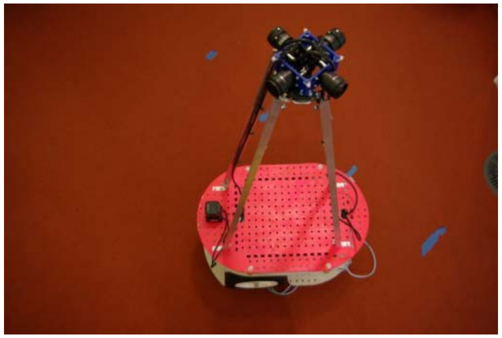
Large-scale localization ground-truth



Outline





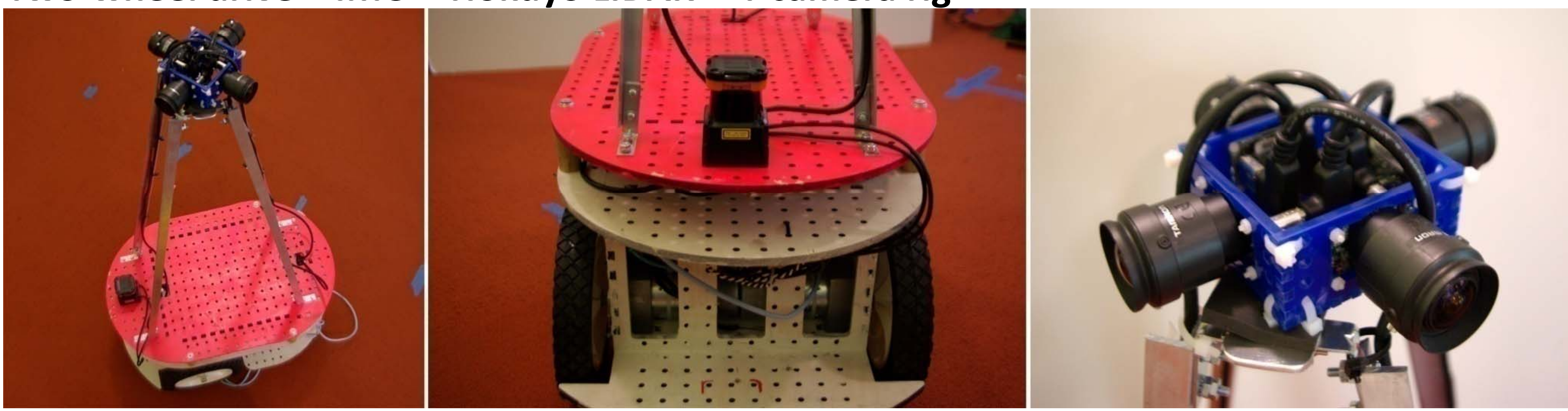


- Sensor suite for vision-based navigation
- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance
- **Application to ground robot navigation**
- User Study

Ground Robot Navigation

- Demonstrate approach on robot equipped with low-level obstacle avoidance capability
- Collect extended ground-truth data

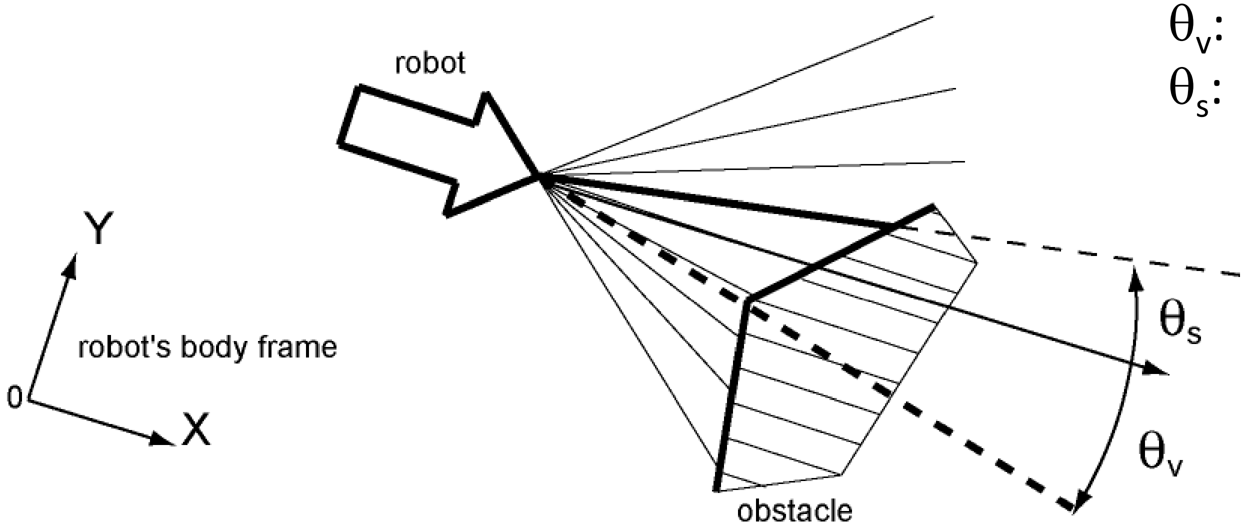
Two-wheel drive – IMU – Hokuyo LIDAR – 4-camera rig



Ground robot navigation using uncalibrated cameras*, Koch, Walter, Huang, Teller, ICRA 2010*

Ground Robot Navigation

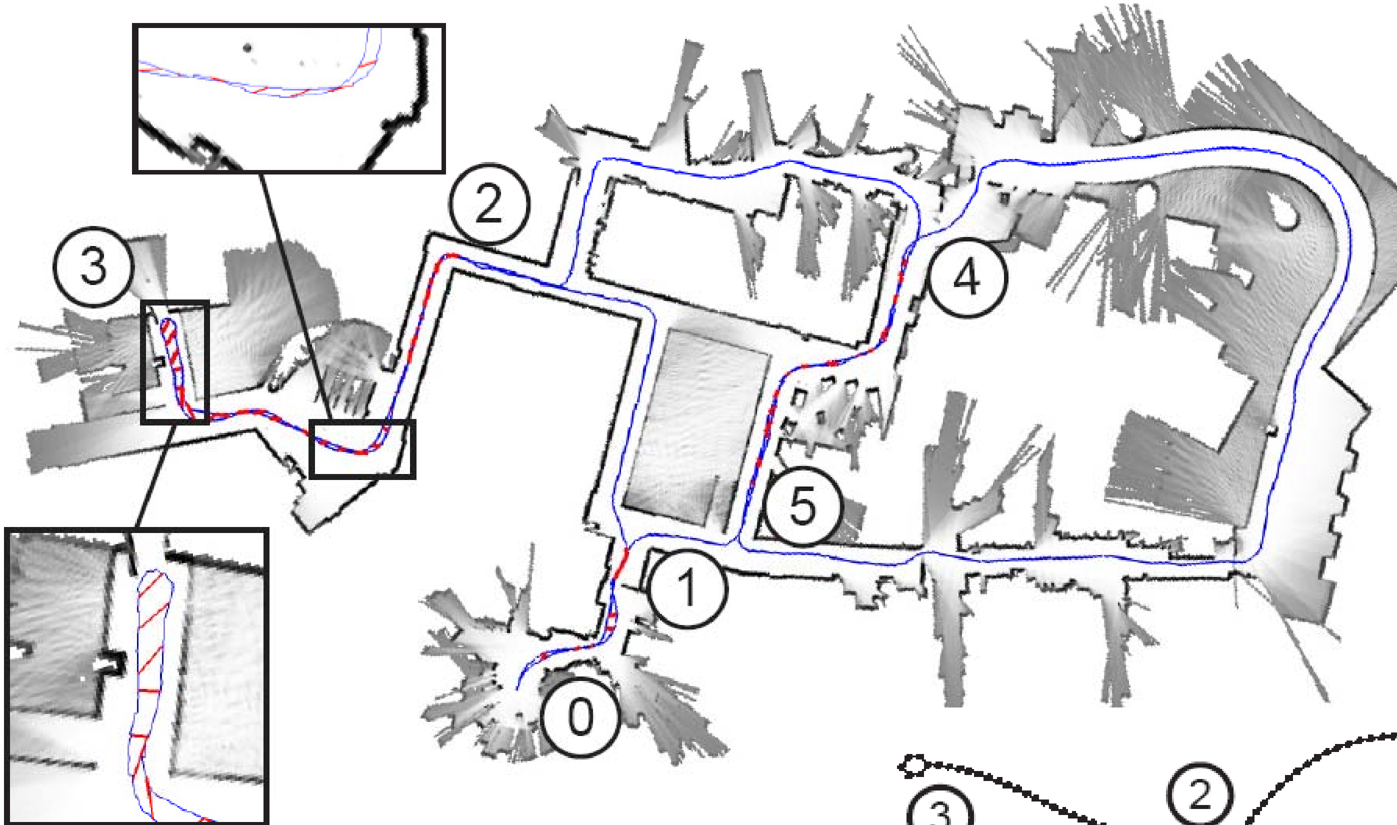
1. Localization & high-level guidance using vision
2. "Greedy" local obstacle avoidance using laser

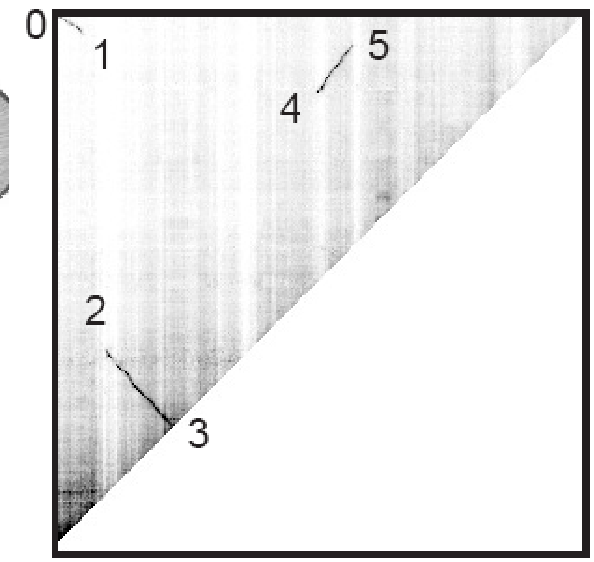


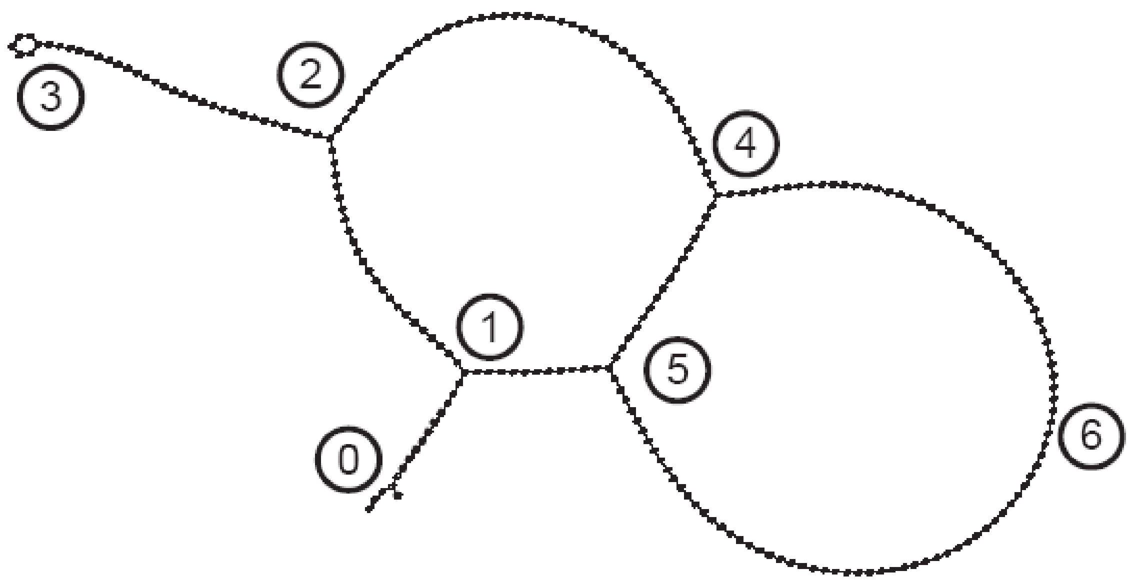
θ_v: vision-based guidance
θ_s: selected direction

***Ground robot navigation using uncalibrated cameras**, Koch, Walter, Huang, Teller, ICRA 2010*

Ground Robot Navigation







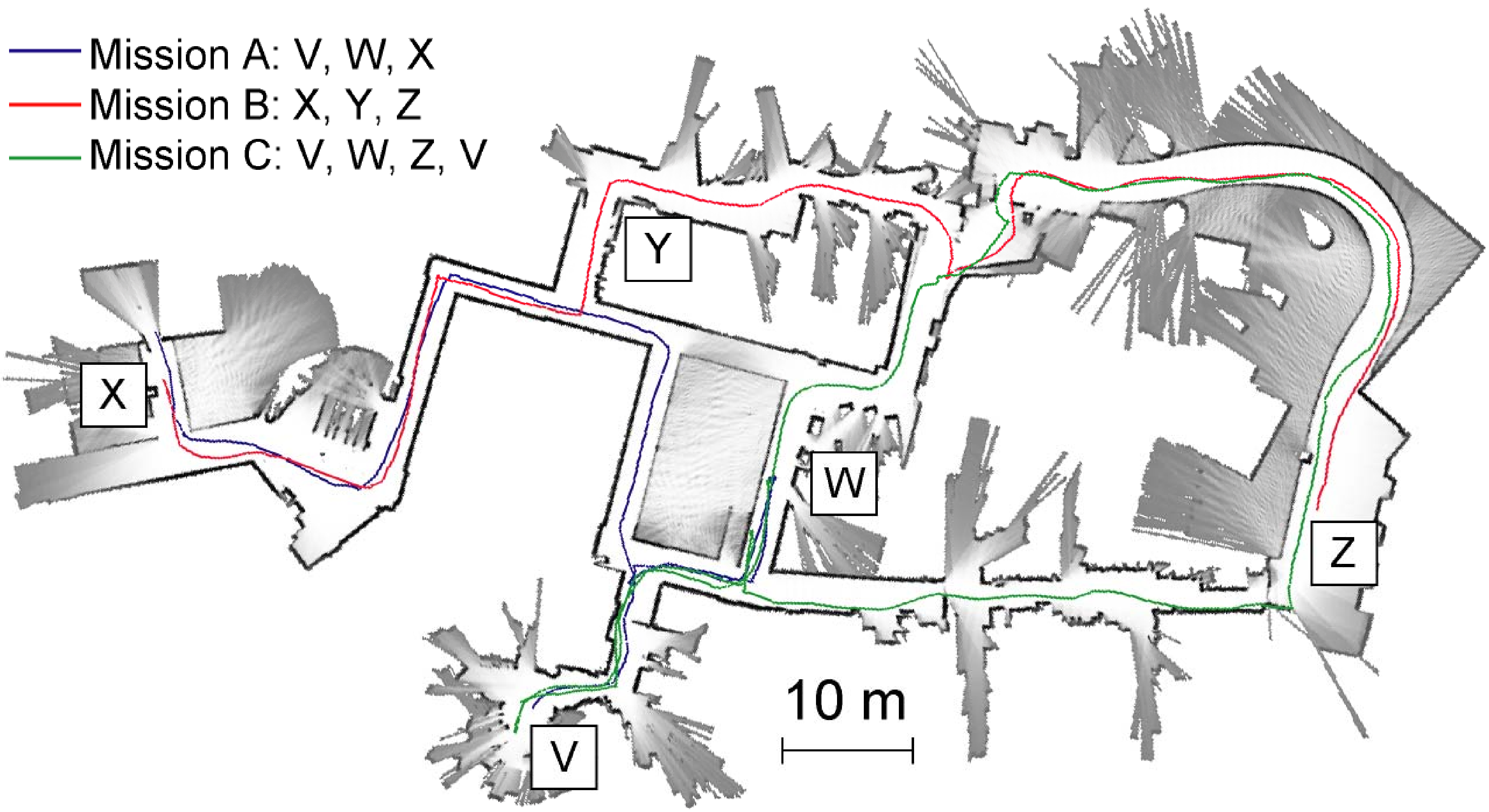
Loop closure – 24 minutes – 290 m
MIT Stata Center 3rd floor

Exploration path (ground-truth)

Loop closure

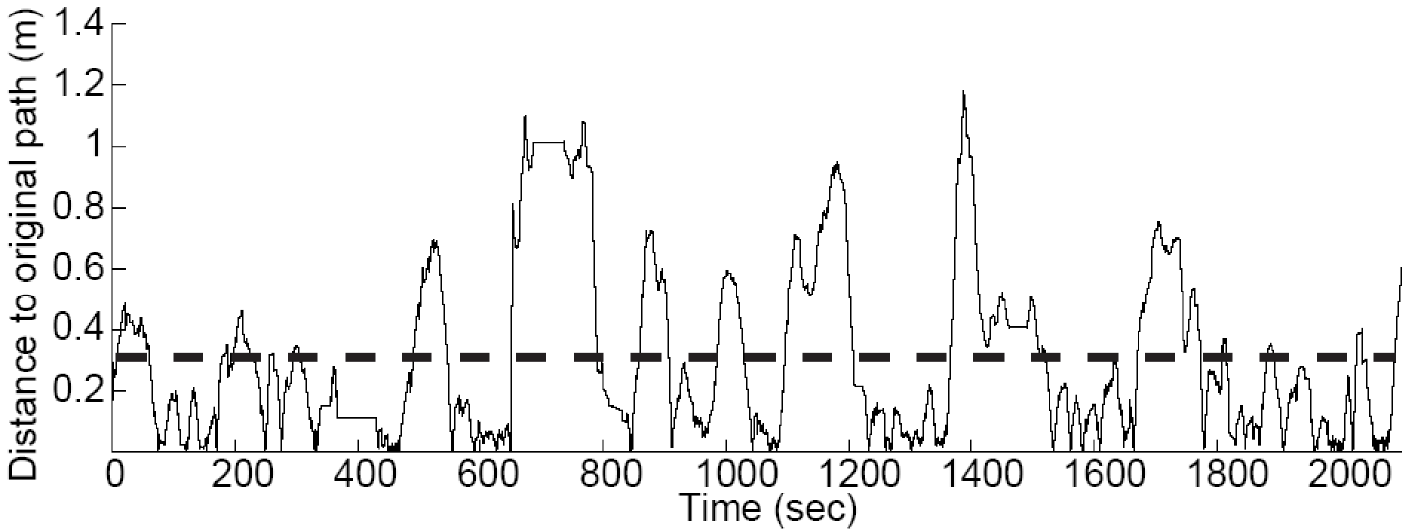
Ground Robot Navigation

3 missions   74 min – 380 m
MIT Stata Center 3rd floor

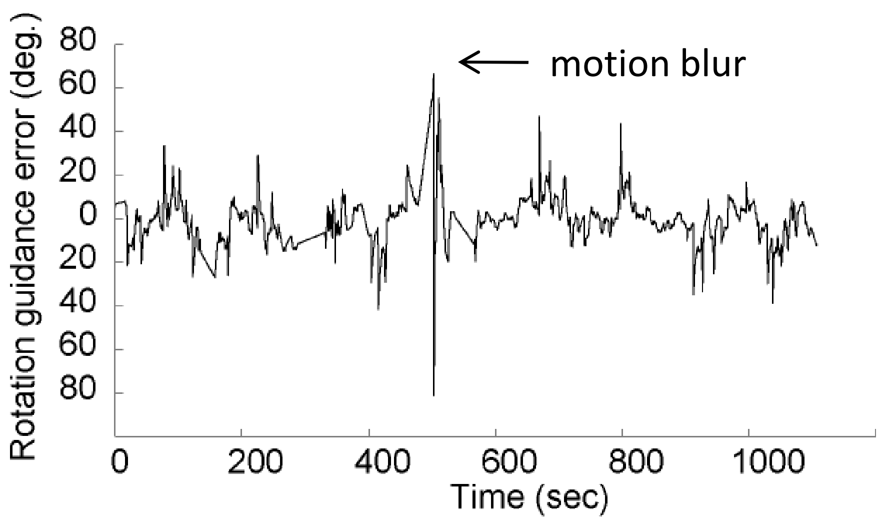


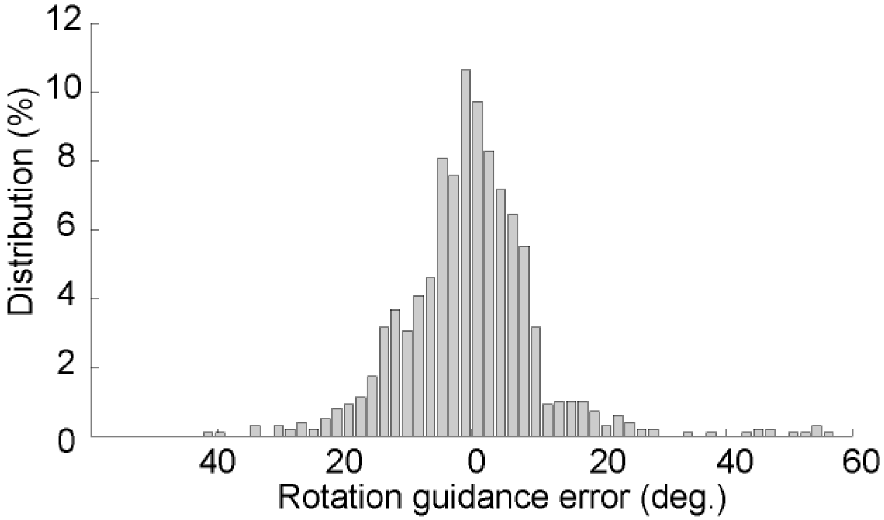
Ground Robot Navigation

Distance to original path (m)



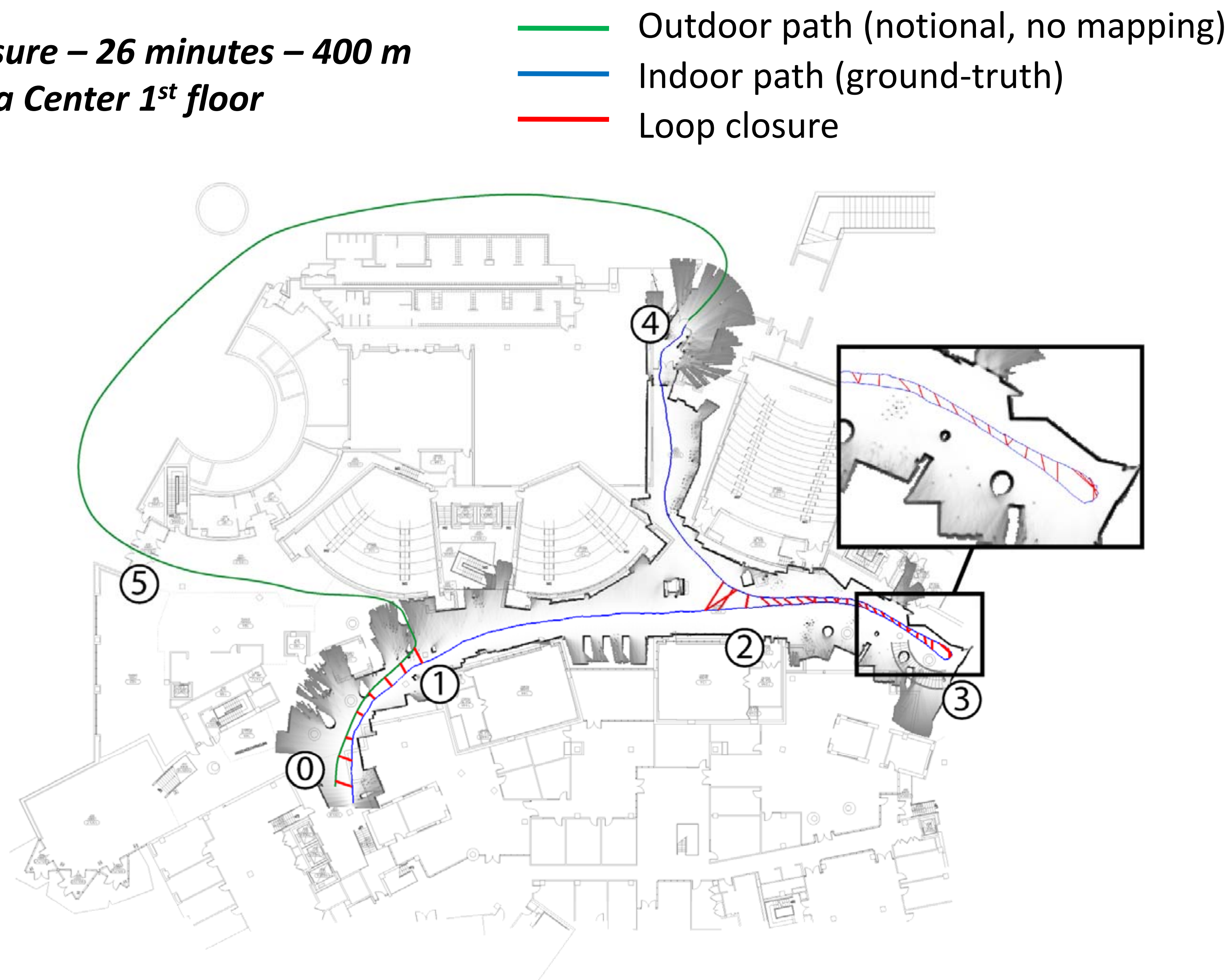
Rotation guidance error (deg.)





Ground Robot Navigation

Loop closure – 26 minutes – 400 m
MIT Stata Center 1st floor

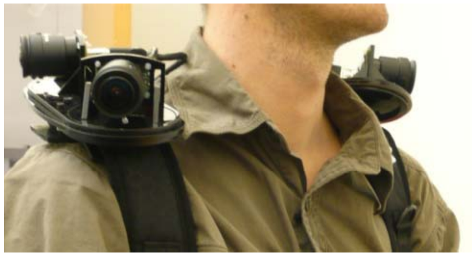


Ground-Robot Navigation

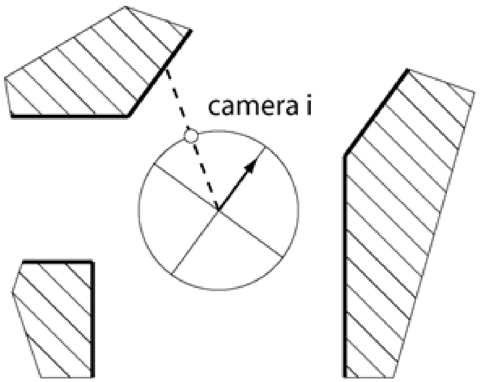
Ground Robot Navigation
Using Uncalibrated Cameras

O. Koch, M. Walter, A. Huang, S. Teller
ICRA 2010

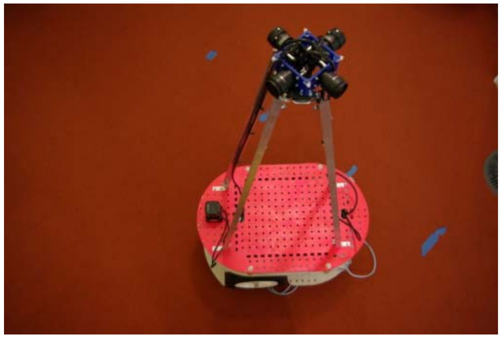
Outline



- Sensor suite for vision-based navigation



- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance



- Application to ground robot navigation
- User Study

User Study

Experiment Setup

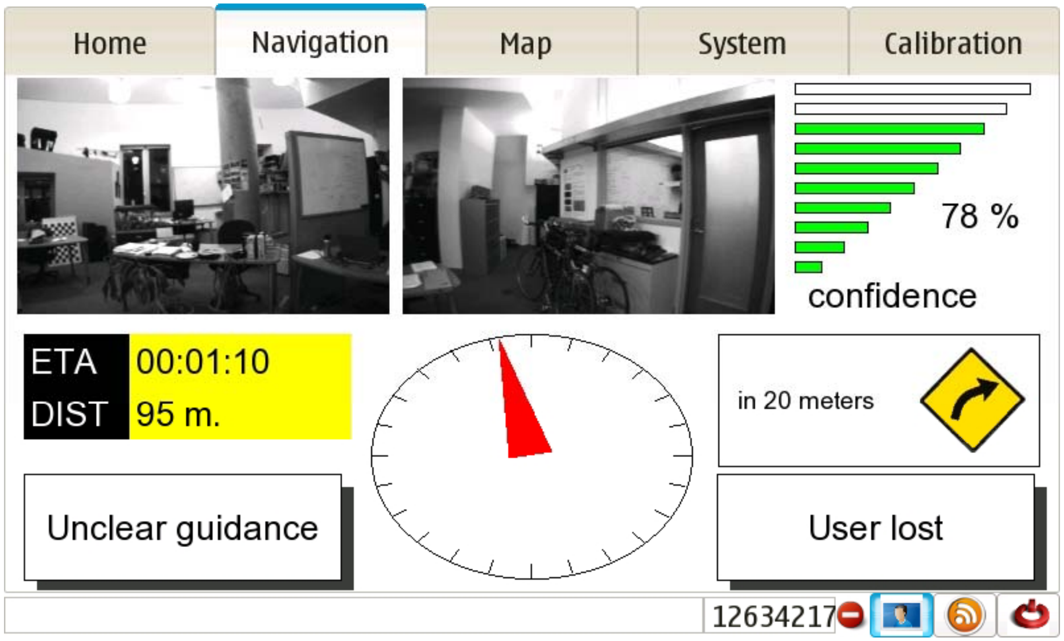
- Follow directions to a series of unknown target destinations in the building.

Effectiveness (qualitative metrics)

- User survey

Efficiency (quantitative metrics)

- # successful missions
- # unclear guidance events
- Speed ratio



Tablet PC User Interface






User Study

9 users, 2.5 hours, 6 km of exploration, 59 missions

Quantitative Analysis (log data)

	Worst	Mean	Best
Success rate	50%	70%	100%
Speed ratio	0.37	0.51	0.61
Unclear guidance	1.5 min.	3 min.	Never
User lost	3 min.	5 min.	Never

Qualitative Analysis (survey)

Positive Feedback
High-level guidance and non-metrical information are crucial.

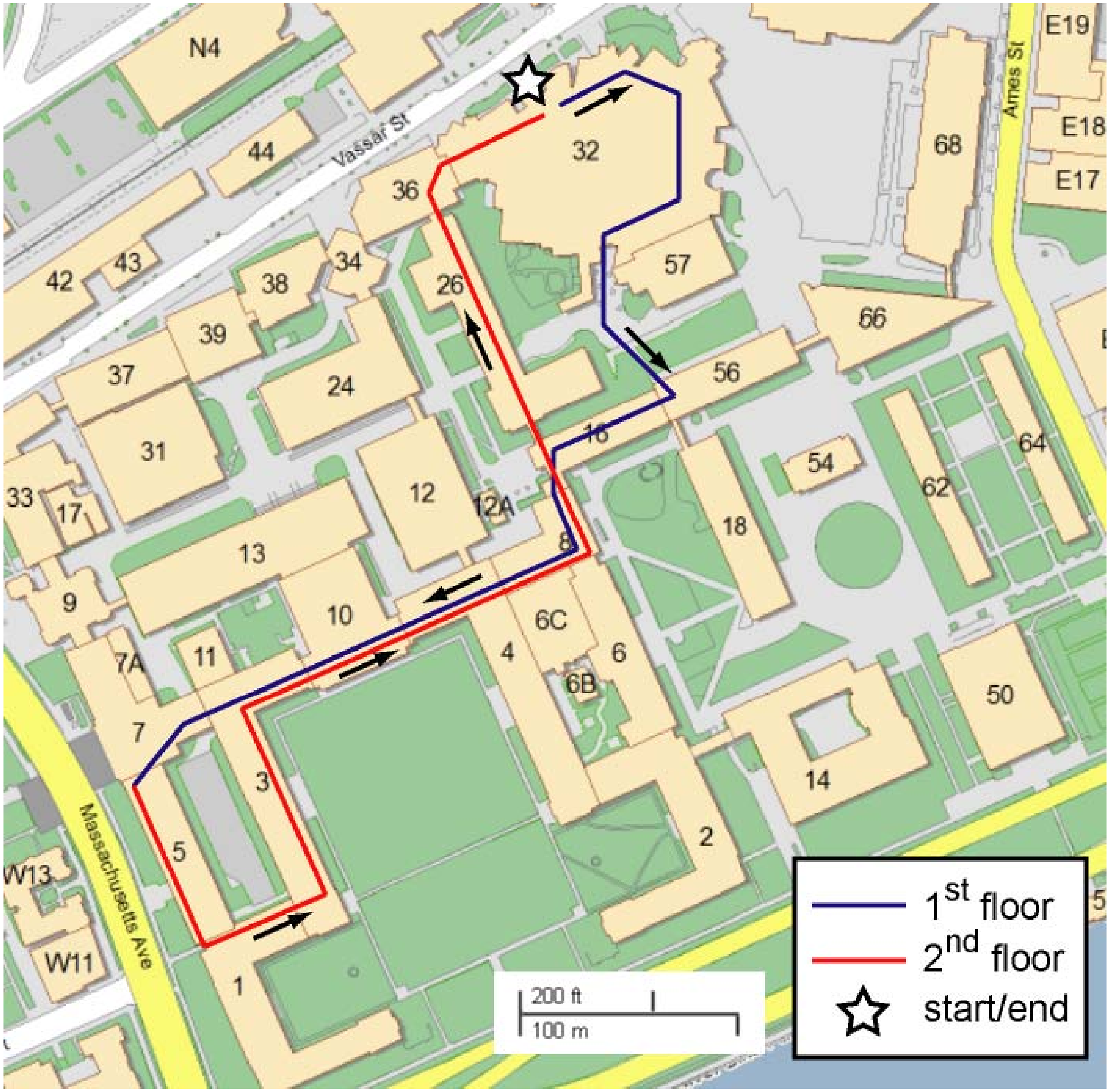
Suggested improvements
More advanced turn-by-turn directions (Google Maps) Compass/turn-by-turn inconsistency. Drawing directions on images.
Simplify GUI.

User Study

Exploration Mission

23 minutes - 1. 1 km

Indoor/outdoor
Dynamic scenes

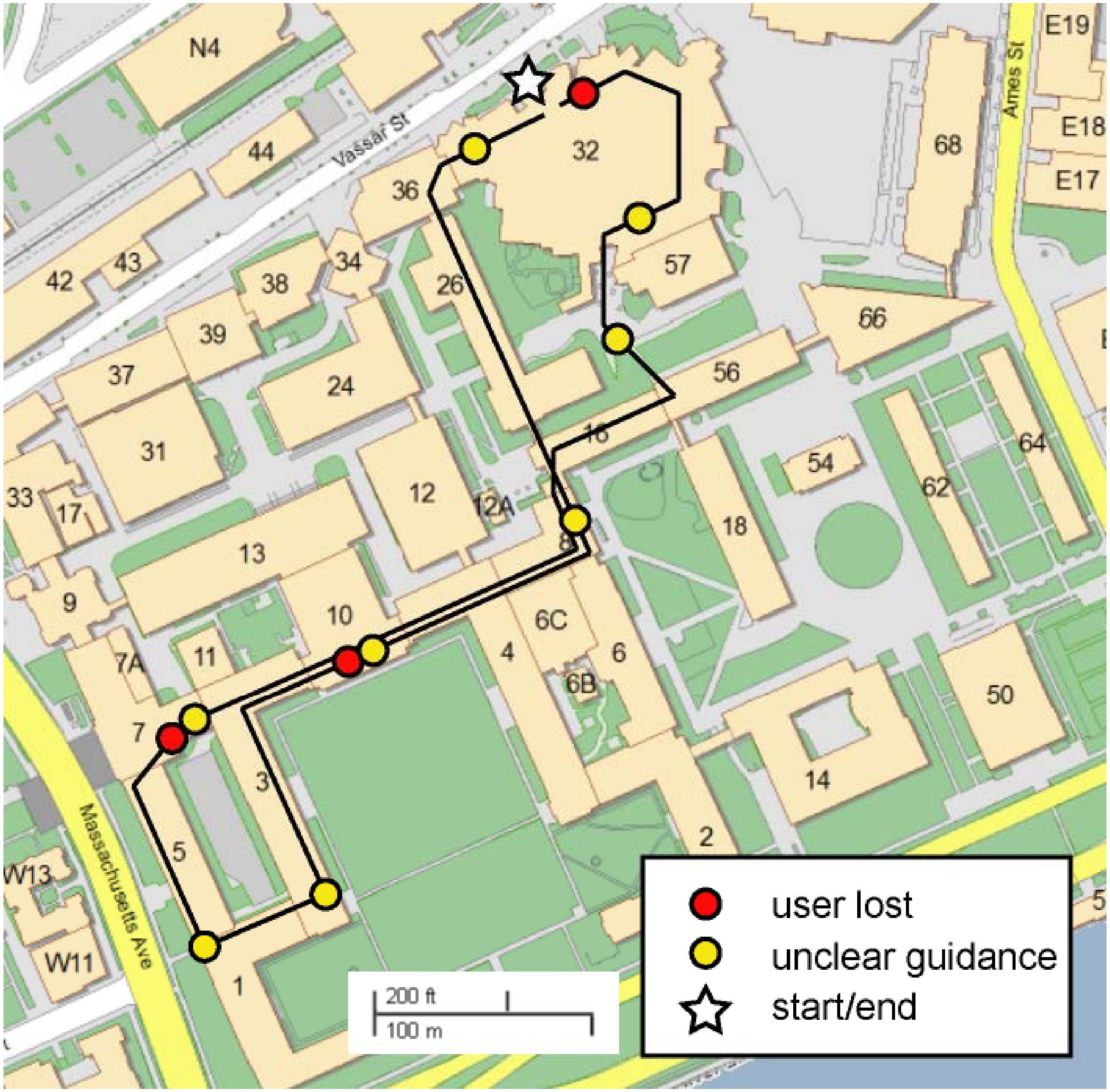


User Study

Navigation Mission

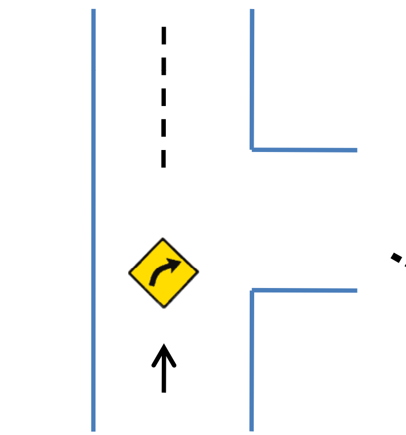
32 minutes

user lost: 3
unclear guidance: 8
Speed ratio: 0.70

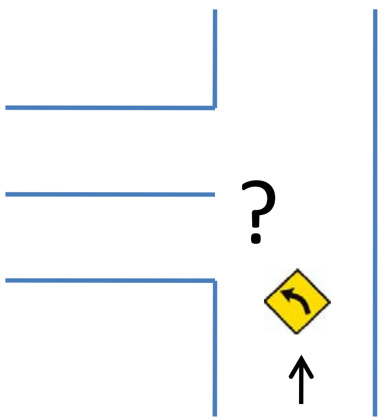


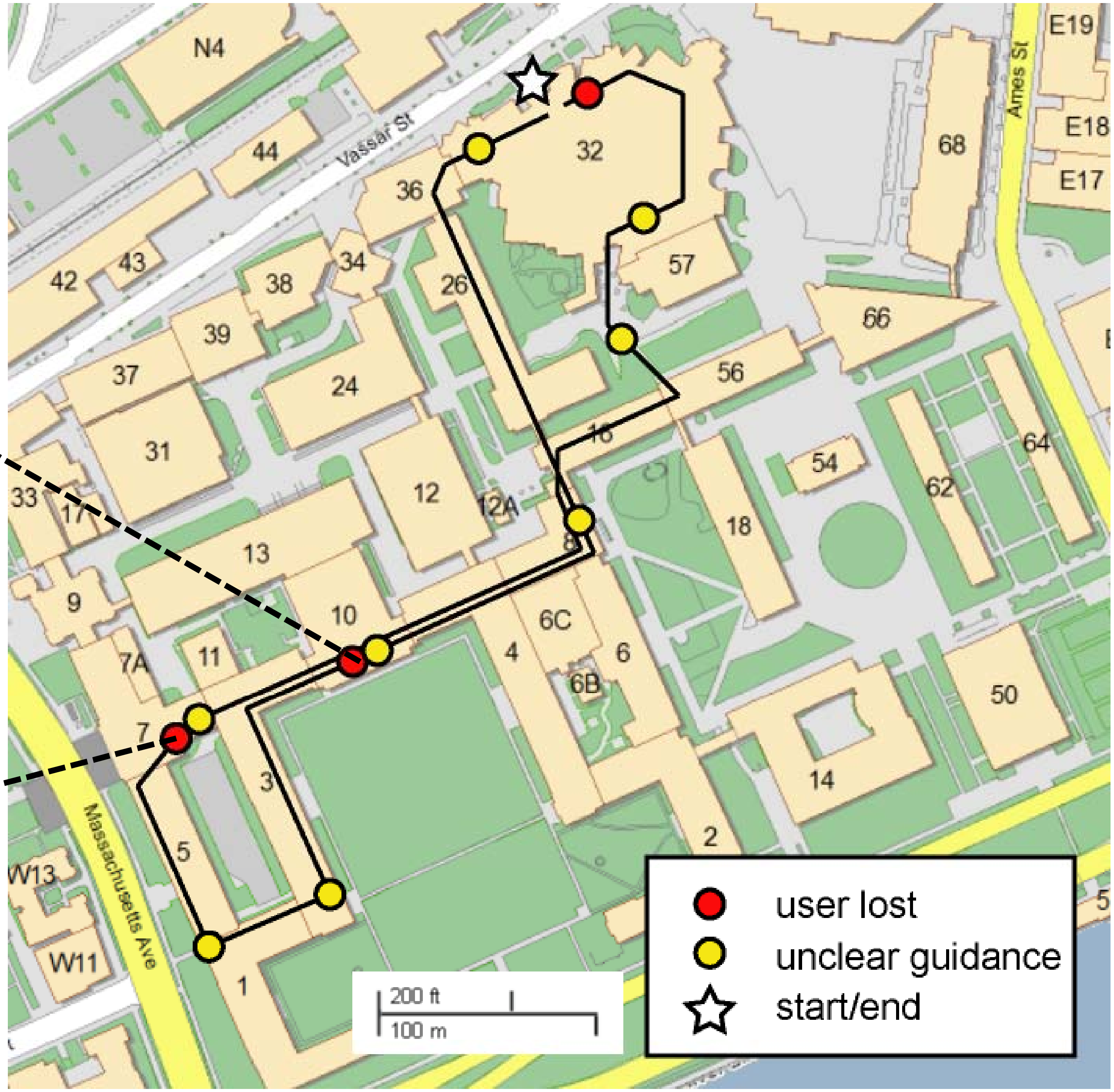
User Study

False positive



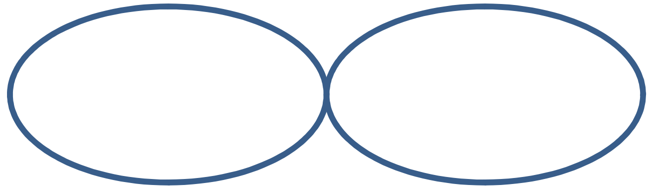
Ambiguous world configuration

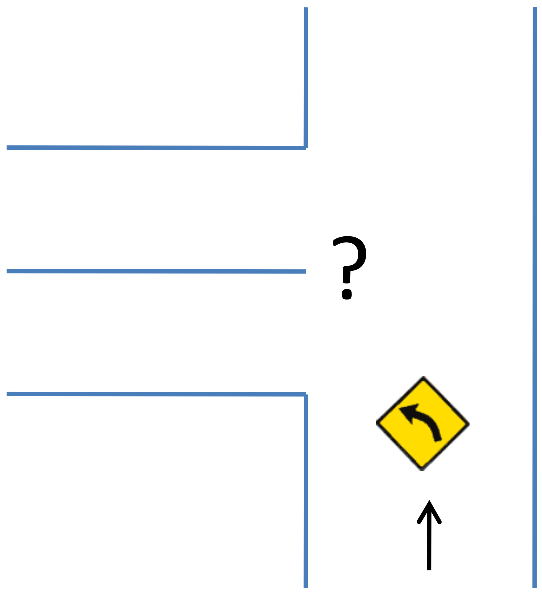




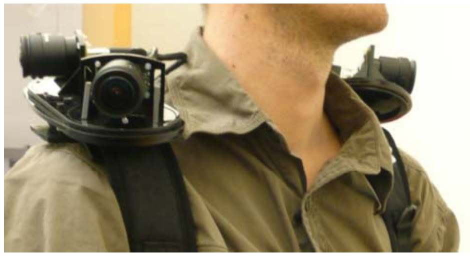
Failure Modes & Limitations

- Featureless environments
- Visually repetitive environments
- Non-isotropic feature distributions
- Ambiguous world configurations
- “One-shot” loop closure events

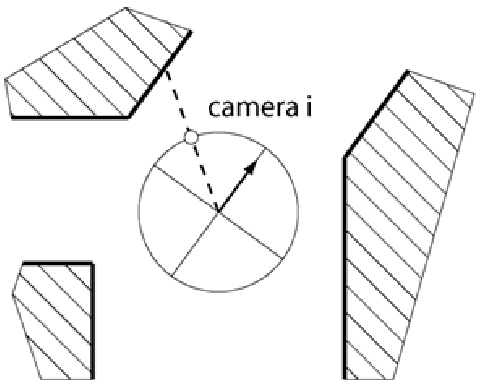




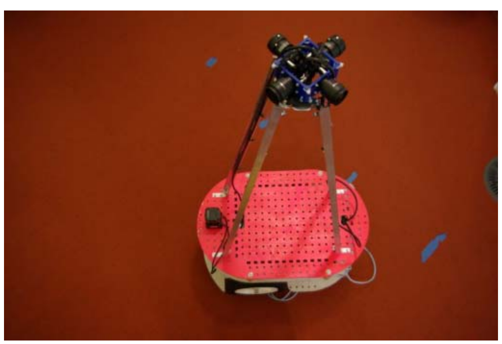
Contributions



- Sensor suite for vision-based navigation



- Body-relative navigation guidance using uncalibrated cameras
  - Topological mapping & loop closure
  - Localization & Rotation guidance
- Application to ground robot navigation



- User Study

Future Work

- Incorporate metrical constraints in the place graph
- Account for spatial relationship of features
- Bound expansion of visual vocabulary
- User interface
- Application to the visually impaired and in natural environments

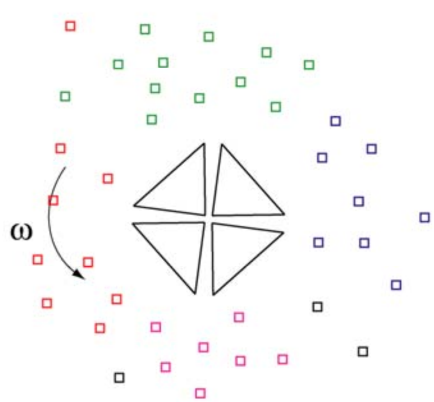
Thank you

- Seth Teller
- Bill Freeman, Rob Miller
- The Draper Laboratory
- TIG, Bryt (Bradley), Ron (Wiken)

The 33x neighborhood

Questions?

Repeat for all features and all pairs of image frames (time sampling)



Match matrix (n x n)