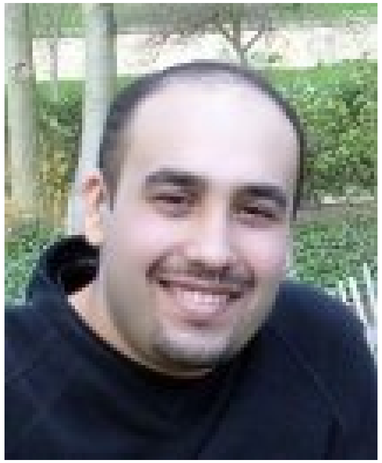
Toward an Object-Based Semantic Memory for Long-Term Operation of Mobile Service Robots

Feras Dayoub, Tom Duckett and Grzegorz Cielniak

School of Computer Science
Lincoln University, UK



UNIVERSITY OF LINCOLN

LINCOLN ROBOTICS

Outline

- Why Semantic Memory?
- Hybrid Map Representation
- Methods
- Results
- Conclusions & Future Work

Why Memory?

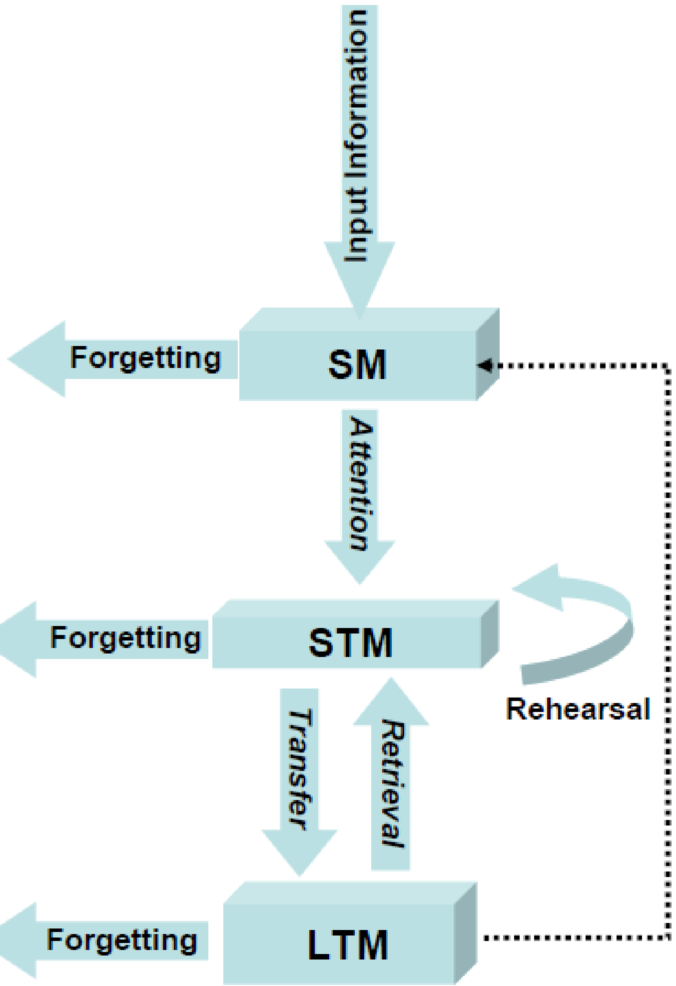
- Robotic helpers and companions need to deal with a dynamic and ever-changing world, including:
  - Changes in the arrangement of objects.
  - Changes in the appearance of the environment.

- Efficient methods are required for filtering, acquiring, storing and updating a robot's spatial semantic knowledge of its working environment.
  - The amount of sensory information to be processed in a lifetime is vast.

- Providing cognitive assistance to users
  - "Where was object X the last time you have seen it?"
  - "What are the most likely locations to find object X in the map?"

Types of Human Memory

- Modal Model
  - Sensory Memory
  - Short-Term Memory
  - Long-Term Memory



R. Atkinson and R. Shiffrin, “Human memory: A proposed system and its control processes,” In K.W. Spence & J.T. Spence (Eds.), The Psychology of Learning and Motivation, vol. 2, pp. 89–195, 1968.

Types of Human Memory

Sensory Memory

Short-Term Memory

Long-Term Memory

Explicit (declarative)

Implicit (non-declarative)

Episodic
(biographical events)

Semantic (concepts)

.
.
.

E. Tulving, “Episodic and semantic memory,” in Organization of Memory. New York: Academic Press, 1972, p. 89101.

Episodic and Semantic Memory

- Episodic memory provides the capacity to remember specific events, e.g.
  - specific experiences of objects and places
  - (and, of course, people – not covered in this paper)
- Semantic memory stores accumulative knowledge of the world
  - generalised representation of the episodes experienced
- Forgetting plays an important role in maintaining a compact representation of the world for subsequent reasoning.
- Generalisation is believed to be one of the important processes involved for improving the efficiency, scalability and adaptability of cognitive systems operating in dynamic environments

Map Representation

- Hybrid map that represents the global topology and local geometry of the environment, as well as the relative 3D location of objects.
  - Appearance-based / topological level
  - Spherical view representation / metric level (submaps)
  - Objects / semantic level

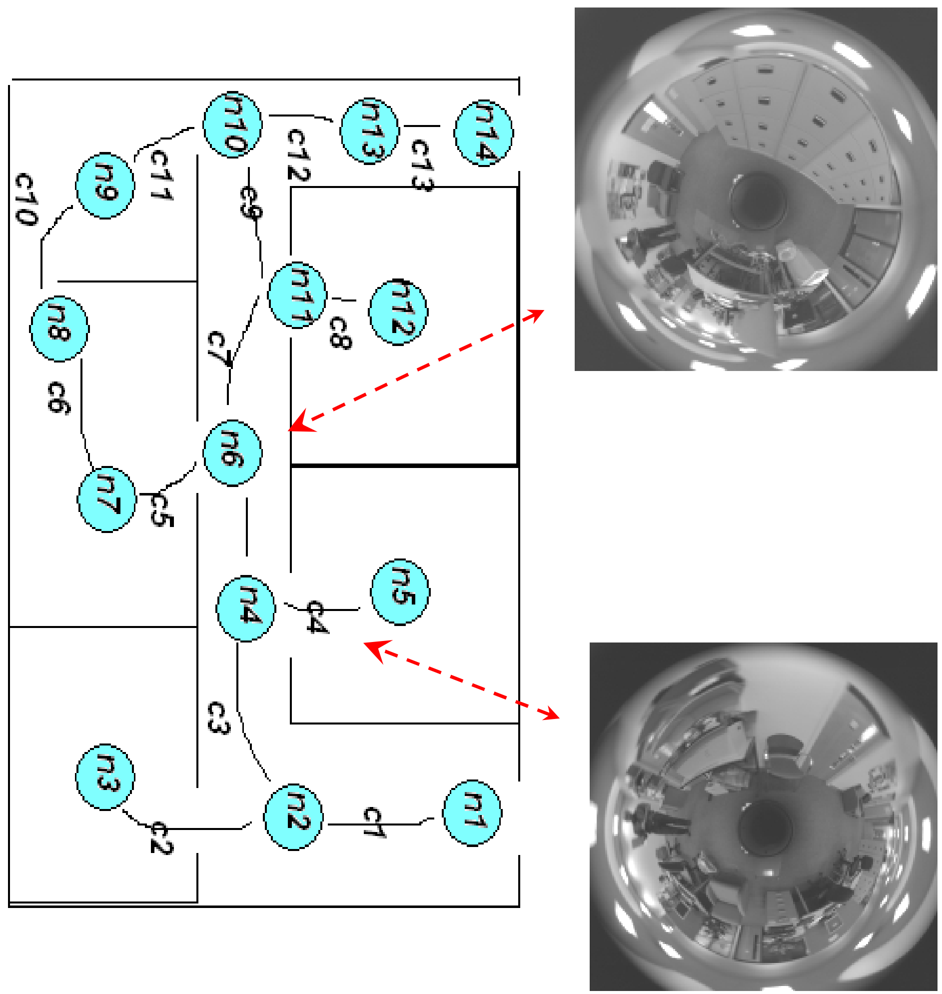
Appearance-based / topological level

- o Represent the environment as a adjacency graph.

- o Each node corresponds to a certain place and each link represents a traversable path.

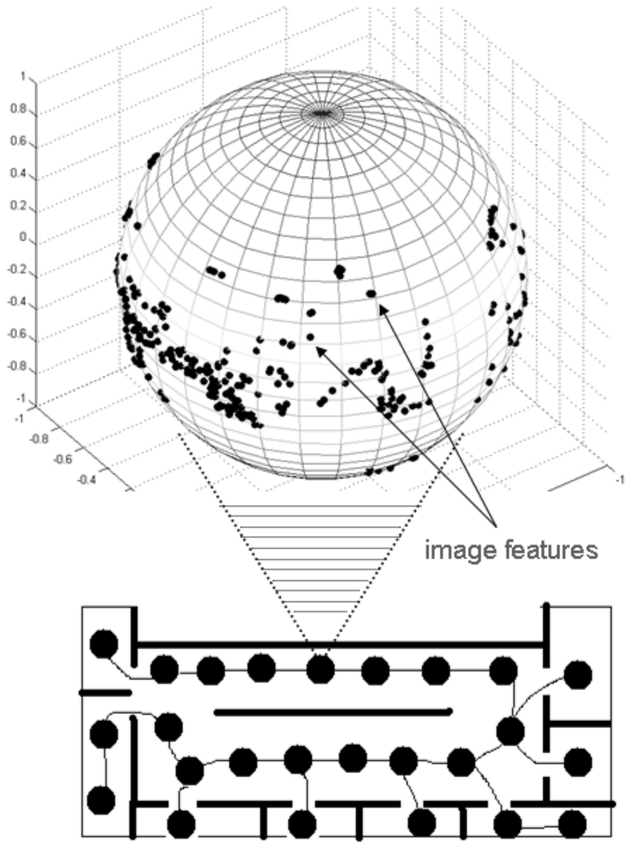
- o A group of image features with their descriptors is used as a signature for the node.

- o A similarity score based on the number of matched points is used for localization.



Spherical view representation / metric level

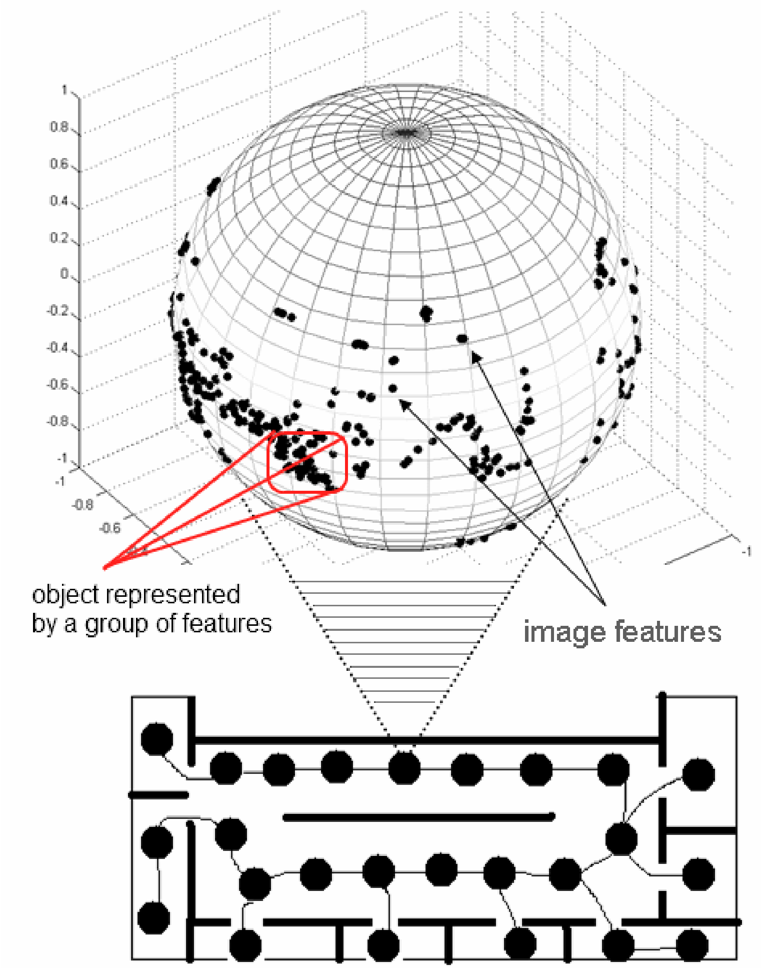
- Hybrid metric-topological map.
- Using the spherical camera model, re-project the image features onto a sphere.
- The group of features on the sphere are used both for global localization and for visual navigation (heading estimation).



F. Dayoub, G. Cielniak, and T. Duckett, “Long-term experiments with an adaptive spherical view representation for navigation in changing environments,” Robotics and Autonomous Systems (to appear), 2010.

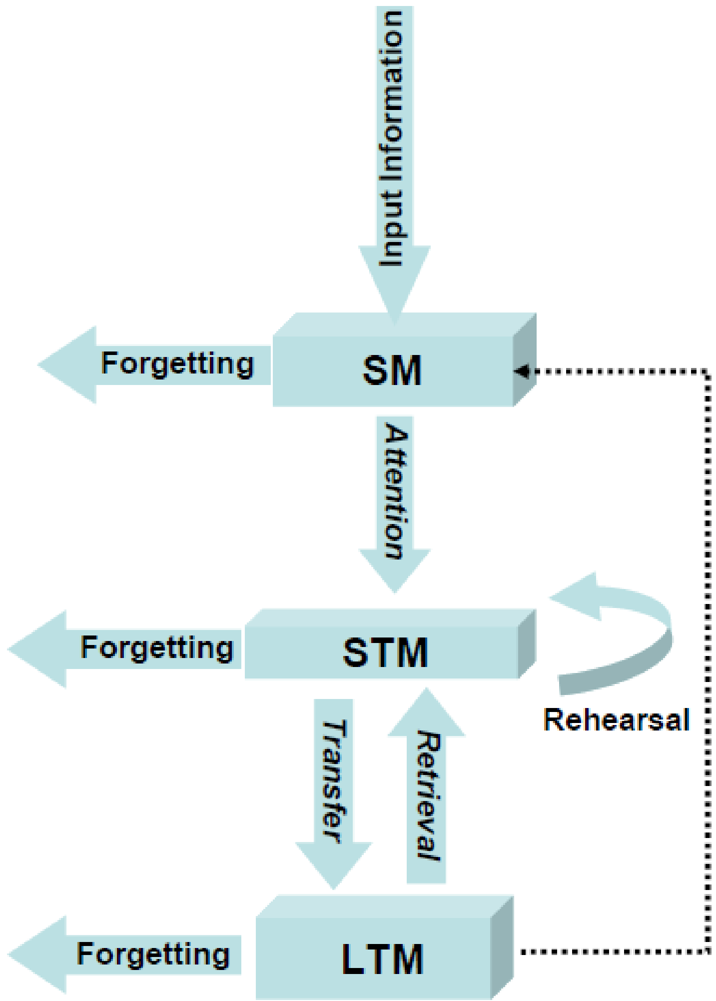
Objects / semantic level

- Our robot is equipped with an omni-directional vision sensor, and uses collections of local image features to represent objects as well as the background of places in the map.
- Objects are represented by using a bag-of-features approach.

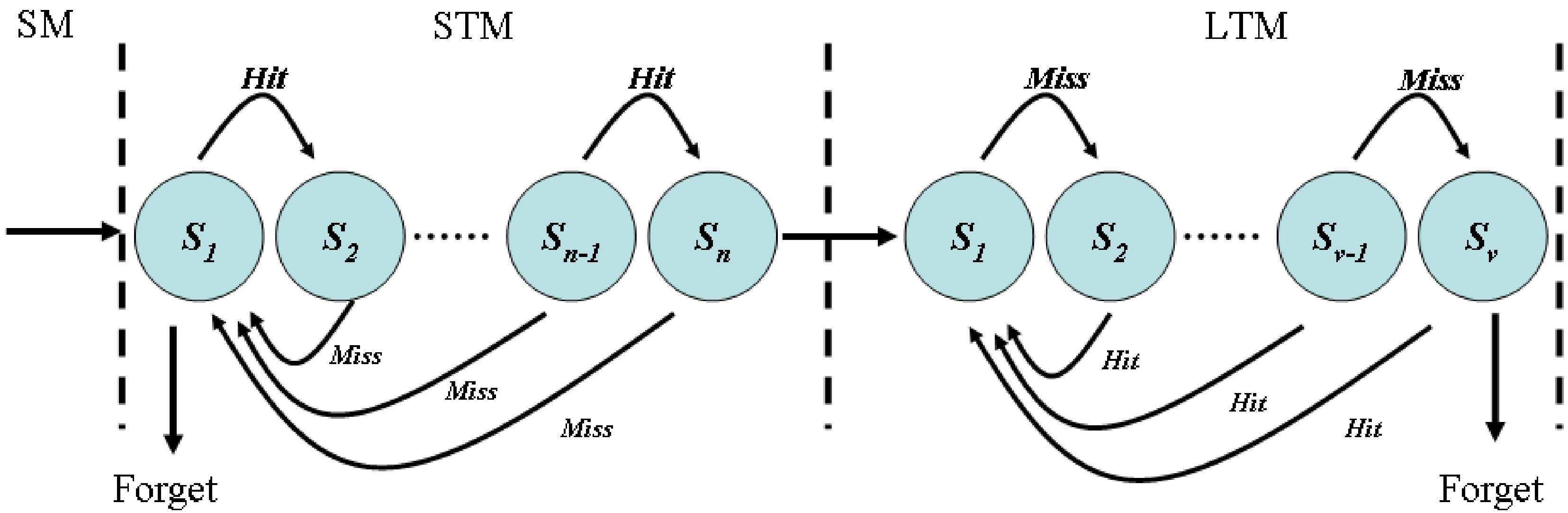


An overview of the memory model

- SM contains the features extracted from the current image.
- STM is used as an intermediate store where new observations are kept for a short time.
- Over time the system uses a rehearsal mechanism to select information that are more stable for transfer from STM to LTM.
- LTM is used in turn by the attentional mechanism for selecting the new sensory information to update the map.



Recall, Rehearsal & Transfer

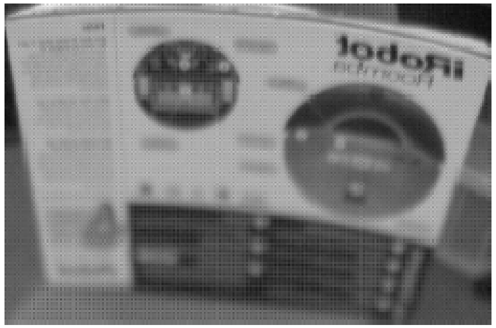
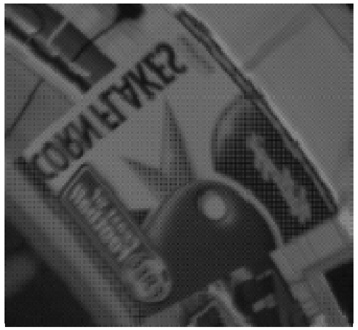
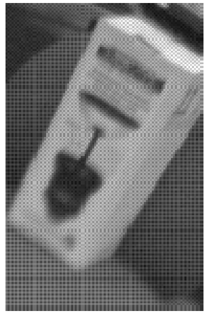


Map Updating

1. Global localisation
2. Object detection
3. Background image registration
4. Projection of observed features into map coordinates
5. Memory update
   - add new object instances and delete instances which have not been observed recently. (an object instance is defined as an object type plus location).
   - add new background features and delete features which have not been observed recently

Object Recognition

The objects used in our experiments.

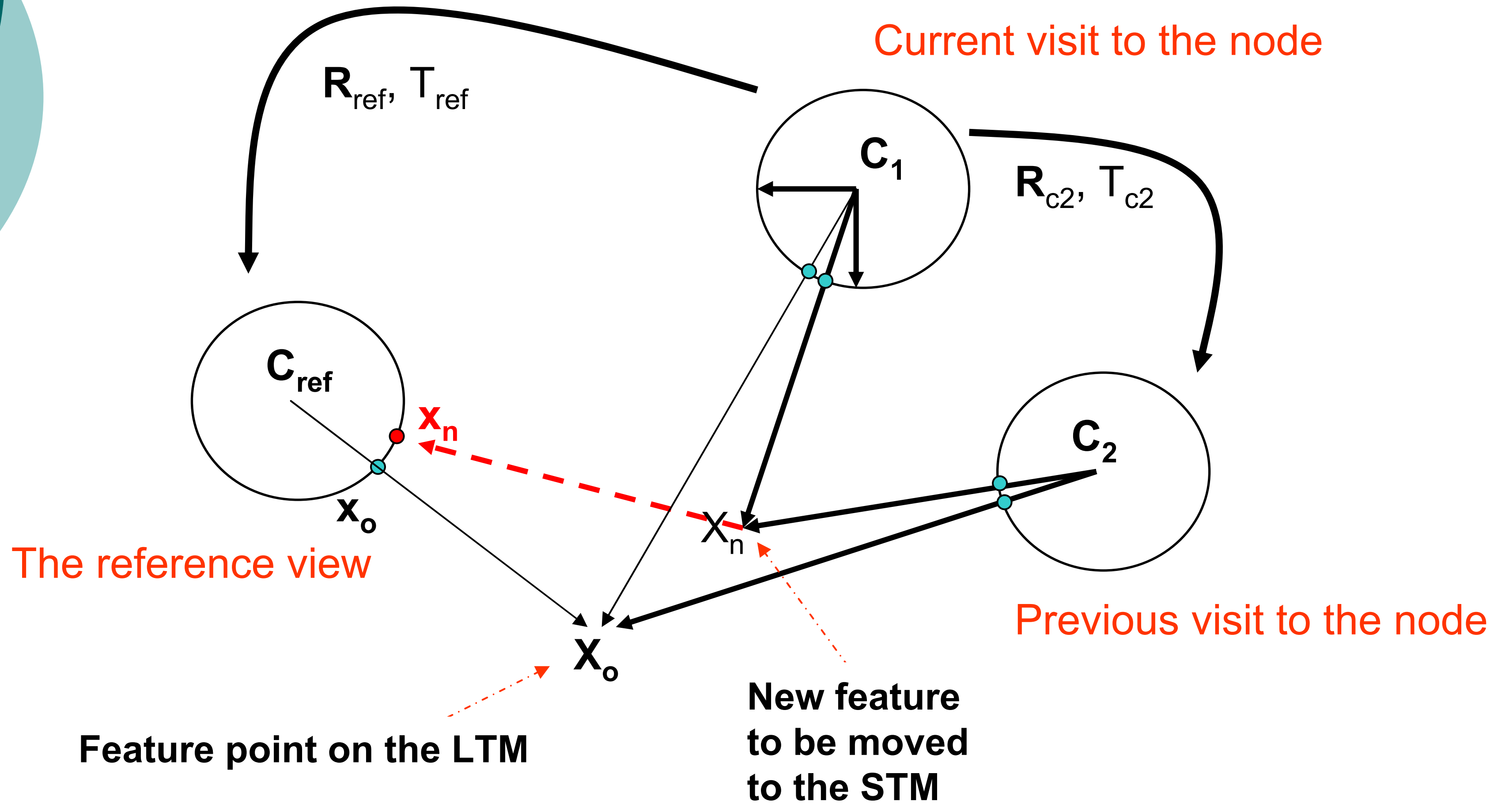
The number of stored SURF features for each object was as follows:

- Roomba box: 305.
- Cornflakes box: 259.
- Panoramic Mirror box : 147.

These features were generated from 3 views for each object.

Object recognition is realised by feature matching.

Updating the background (reference views)

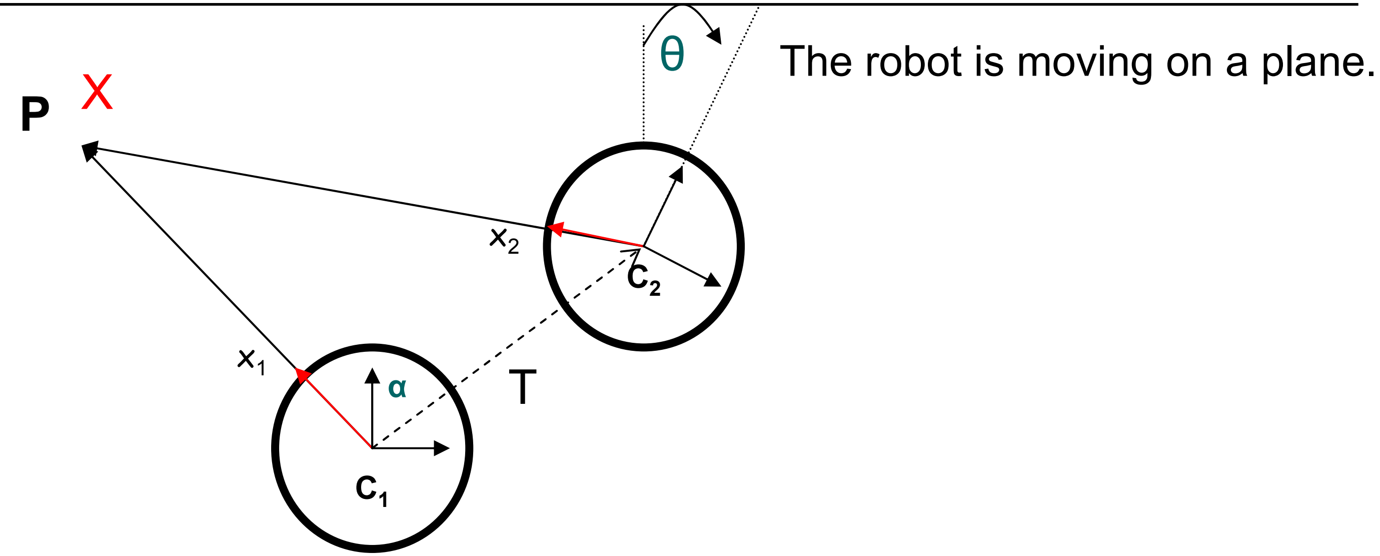


UKF is used to track the position x_n of each feature as it moves through STM and LTM

Heading estimation for visual navigation

- Common method used in the literature.
- After the localization step, the robot uses the reference view of the node to navigate.
- Using the epipolar geometry for spherical cameras, the **essential matrix** can be estimated using the correspondences points between the spherical view of the node and the current view.
- The essential matrix then can be decomposed to give a rotation matrix and a unit vector as translation direction.

Epipolar Geometry



- $X_1 = \lambda_1 \mathbf{x}_1$; $\lambda_1 \in \mathbf{R}_+$; in the reference frame of C_1
- $X_2 = \lambda_2 \mathbf{x}_2$; $\lambda_2 \in \mathbf{R}_+$; in the reference frame of C_2
- $X_2 = \mathbf{R}_\theta X_1 + T$
- $\lambda_2 \mathbf{x}_2 = \mathbf{R}_\theta \lambda_1 \mathbf{x}_1 + T$
- $(\mathbf{x}_2)^T \mathbf{E} \mathbf{x}_1 = 0$
- $\mathbf{E} = [T]_x \mathbf{R}_\theta$

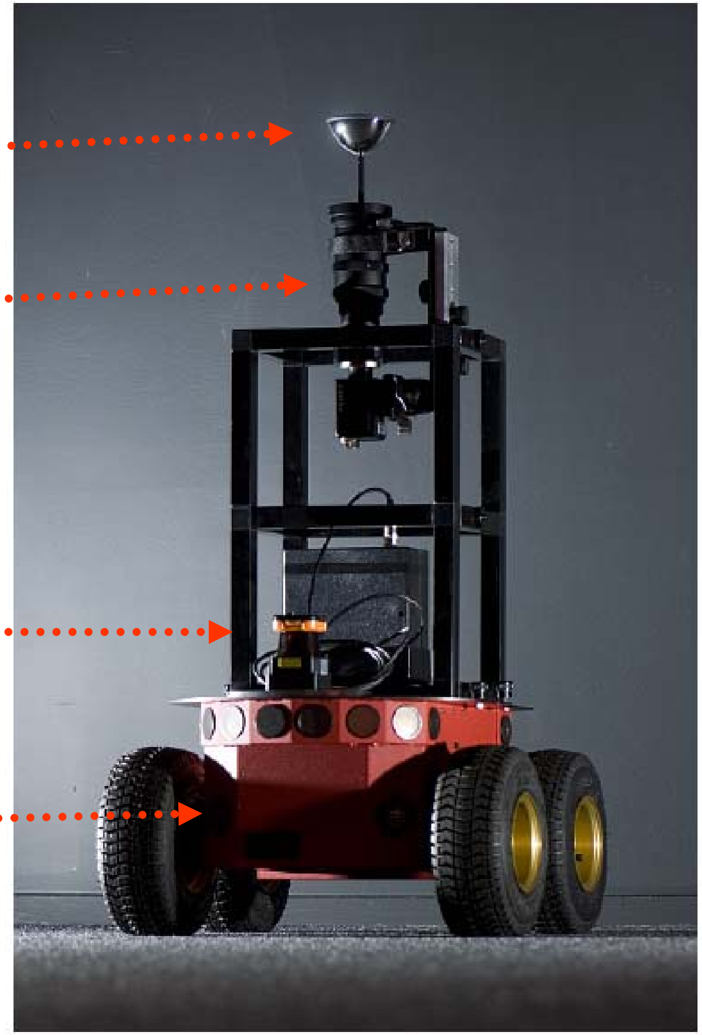
Experiments

Omnidirectional Mirror

4.2 megapixels camera
Jai TMC-4100GE.

Hokuyo UTM-30LX Laser

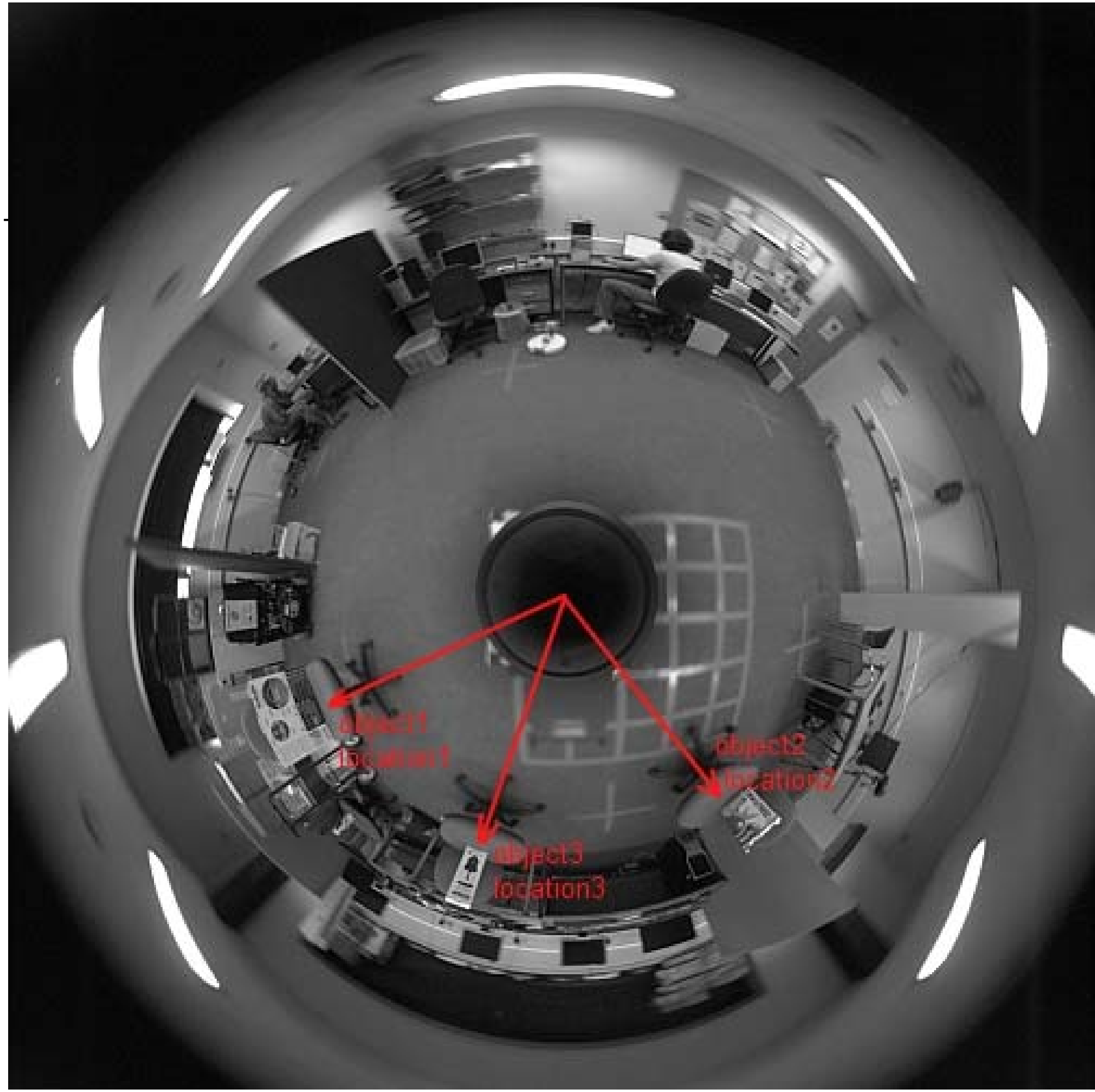
PIONEER 3-AT Robot

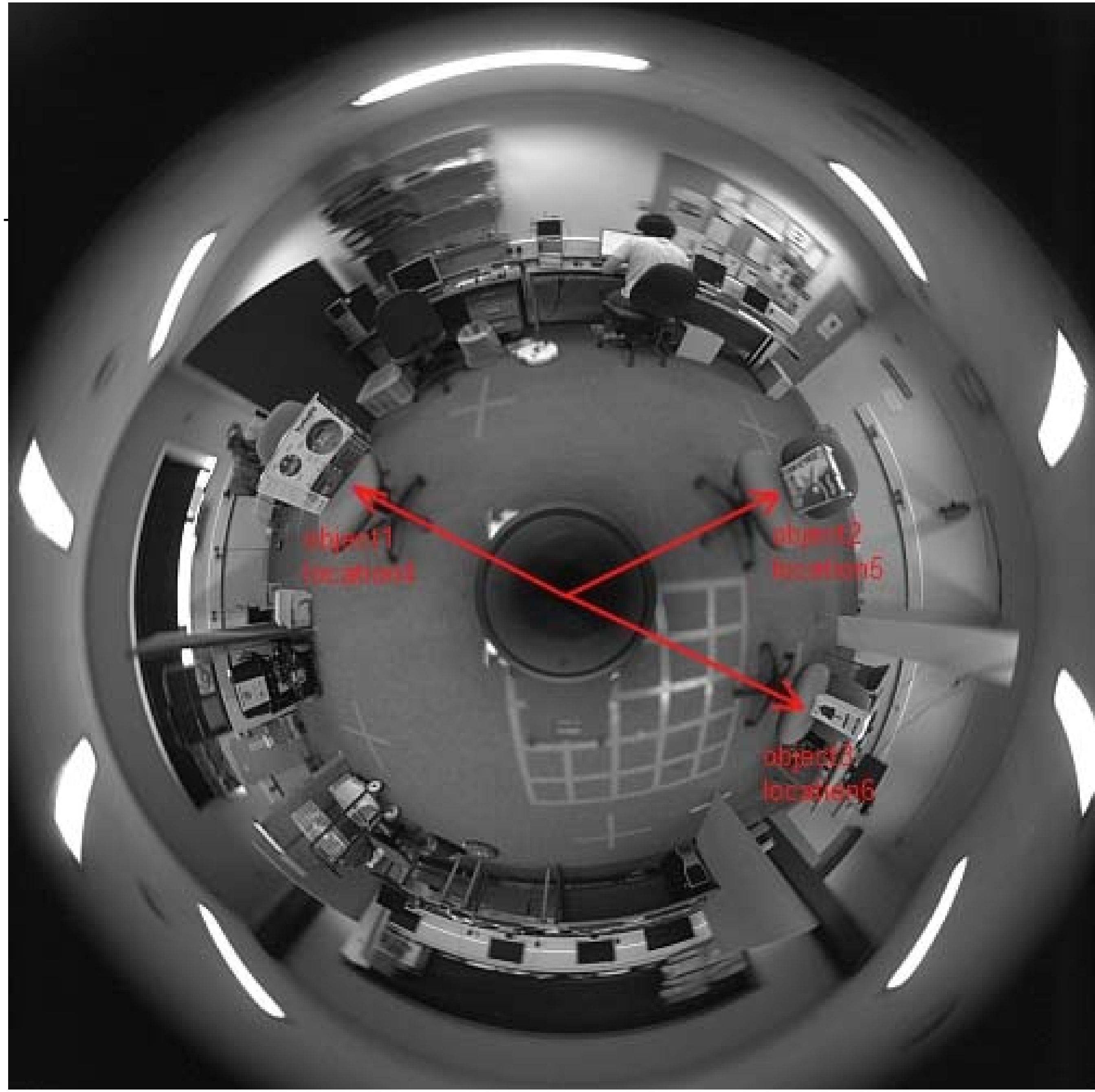


Experiments

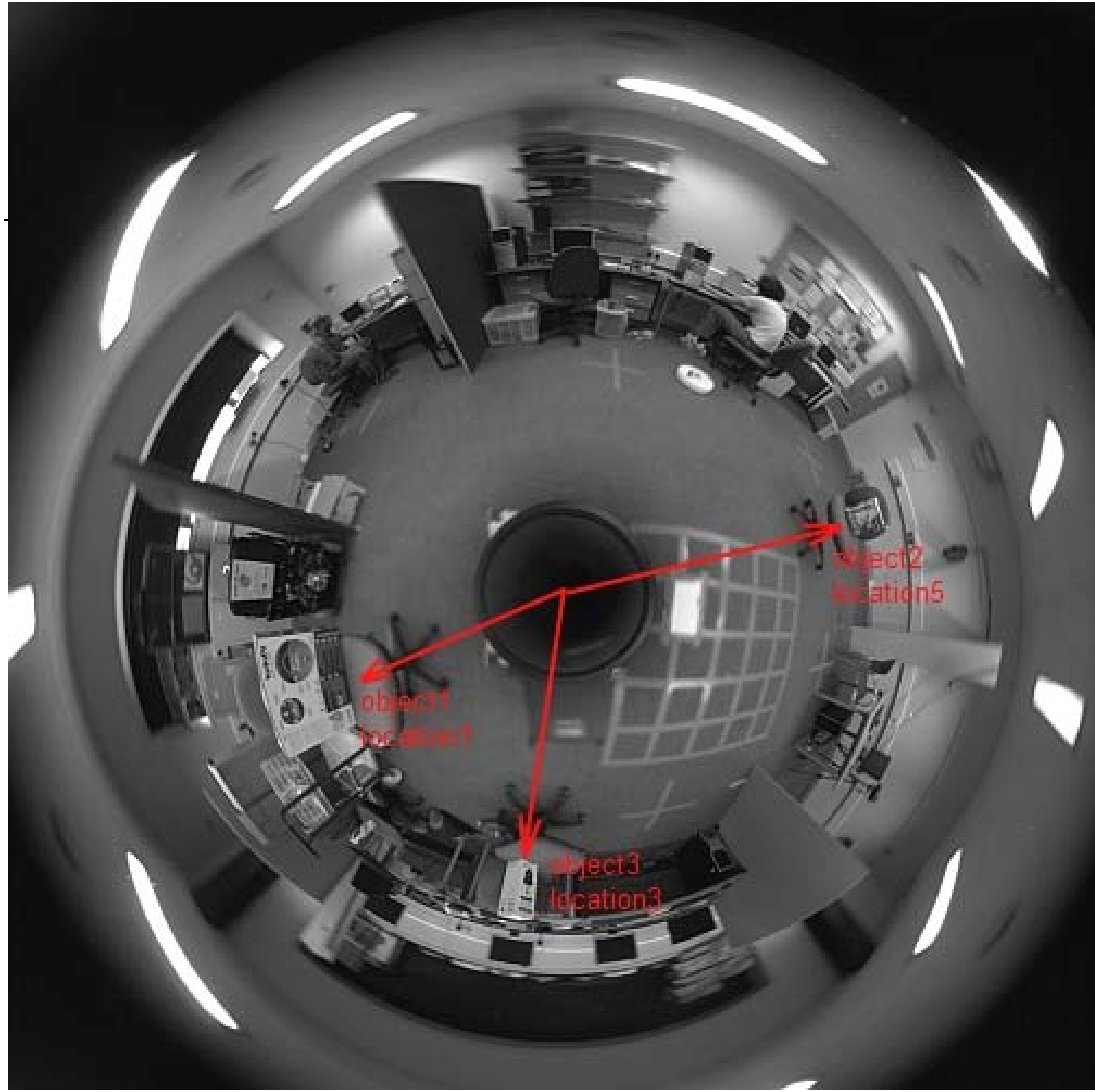
- The test data consisted of 88 images recorded from random locations of the robot in our lab.
- We manually changed the locations of some objects, sometimes temporarily removing the object from the room.
- The memory model was tested with 5 stages for STM and 15 for LTM. Sometimes the robot was not able to detect the objects due to occlusion or a low number of matched features.

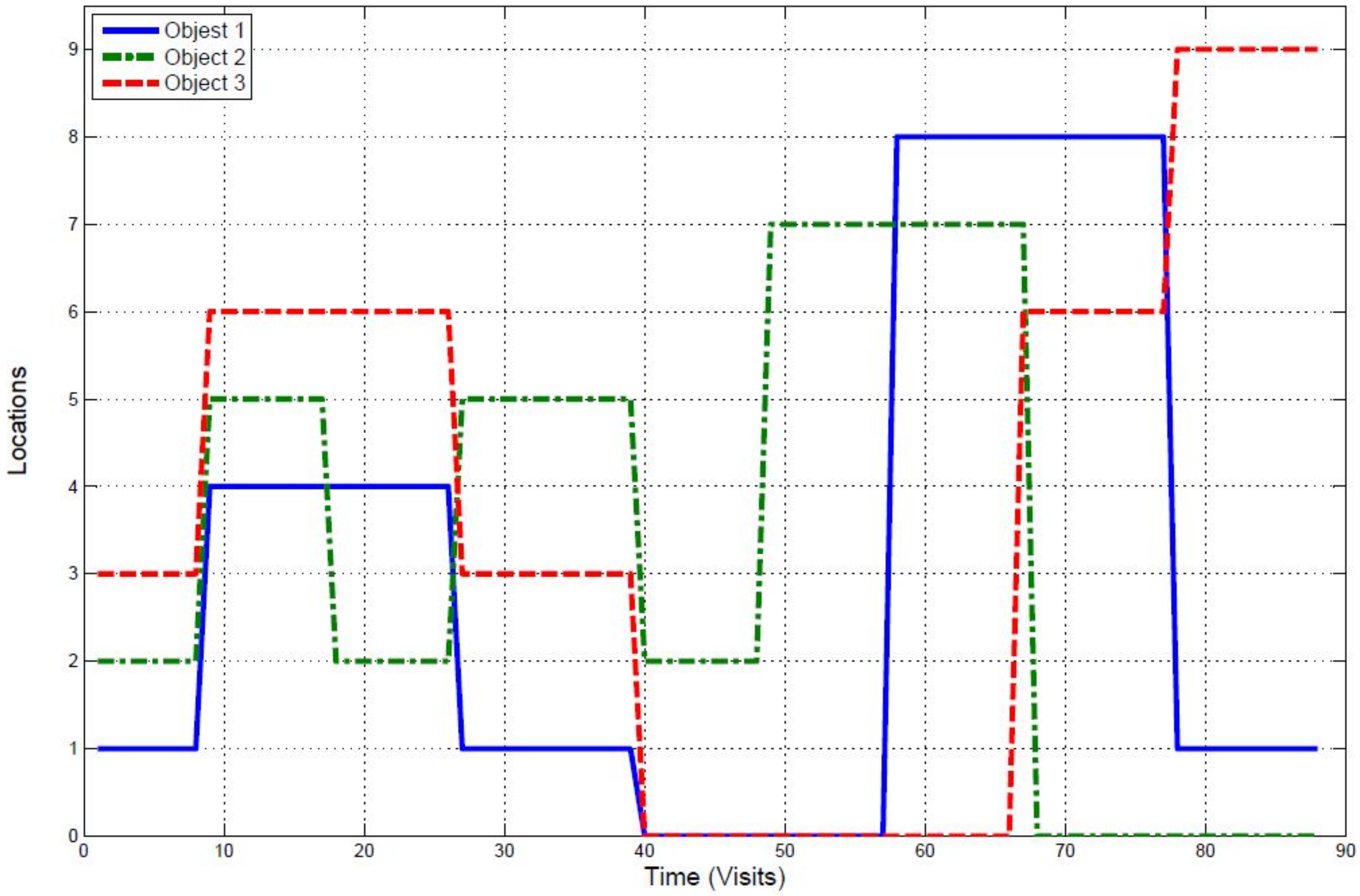
LINCOLN ROBOTICS



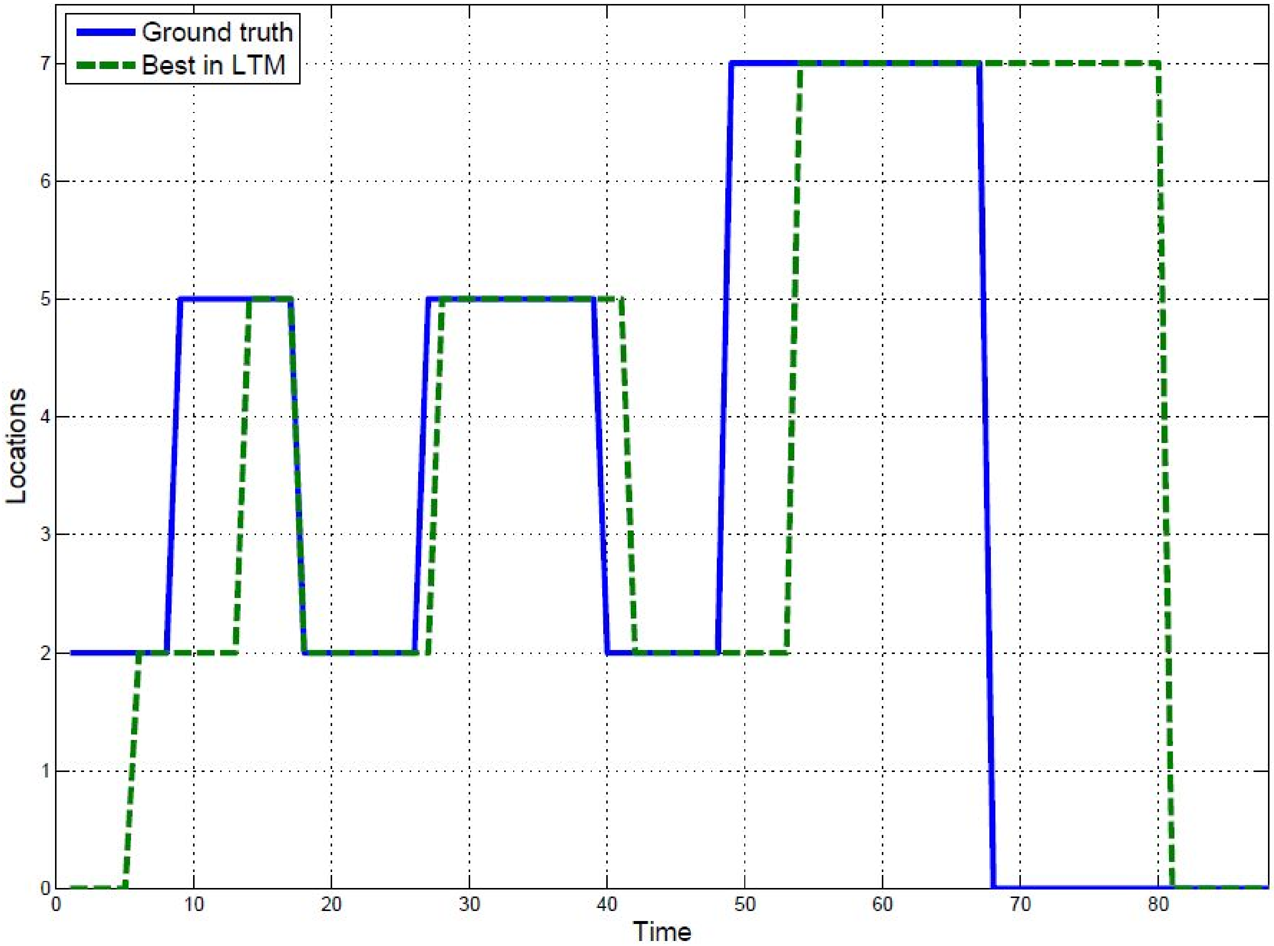


LINCOLN ROBOTICS


object2
location5
object1
location1
object3
location3

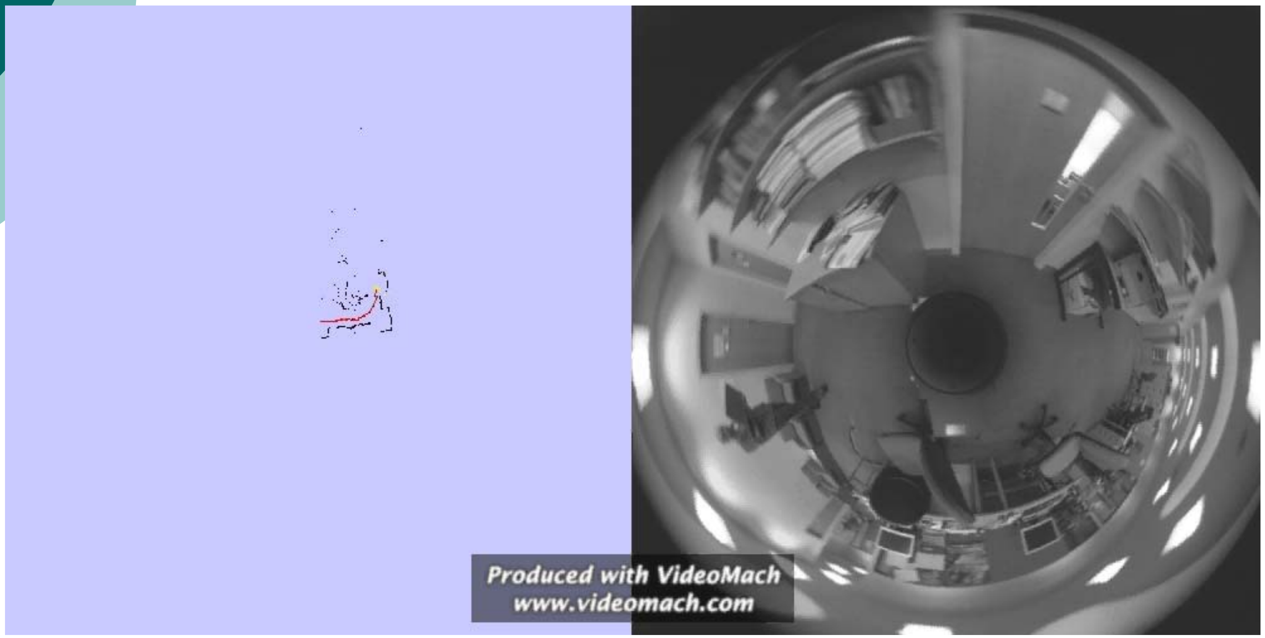


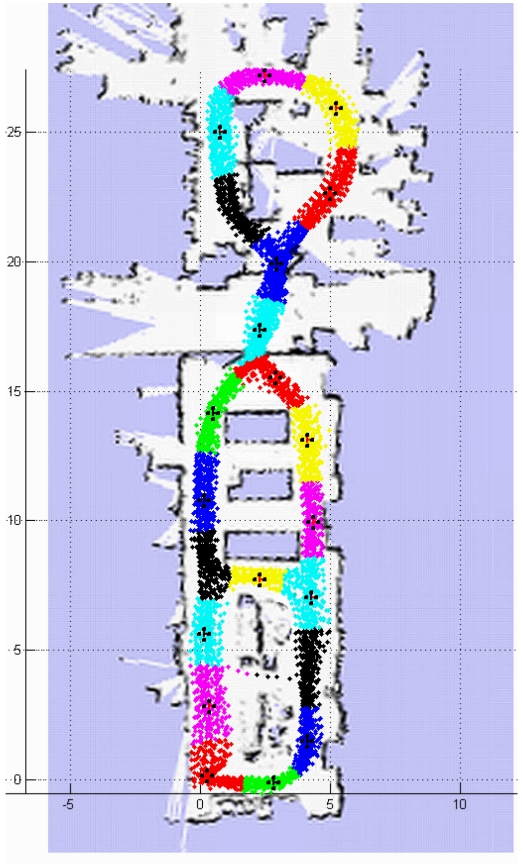
Ground truth information for the experiment



“Where was object X the last time you have seen it?”
“What are the most likely locations to find object X in the map?”

Large-Scale Experiments





F. Dayoub, G. Cielniak, and T. Duckett, “Long-term experiments with an adaptive spherical view representation for navigation in changing environments,” Robotics and Autonomous Systems (to appear), 2010.

Conclusions and Future Work

- An object-based semantic memory for mobile service robots, augmenting our previous results in long-term operation for the topological and metric levels of the robot's map.
- The results show that the semantic memory follows the ground truth location of objects in the test environment with a small time lag, with some variation due to noise in perception.
- Work in progress:
  - Temporal calibration
    e.g. learning rates, how many stages in STM & LTM
  - Large-scale experiments
  - Represent episodes involving people as well as objects & places → generalization to semantic memory