Semantic Modeling of Places using an Object-based Representation

Frank Dellaert

Georgia Institute of Technology

Joint work with **Ananth Ranganathan**

now at Honda Research Labs

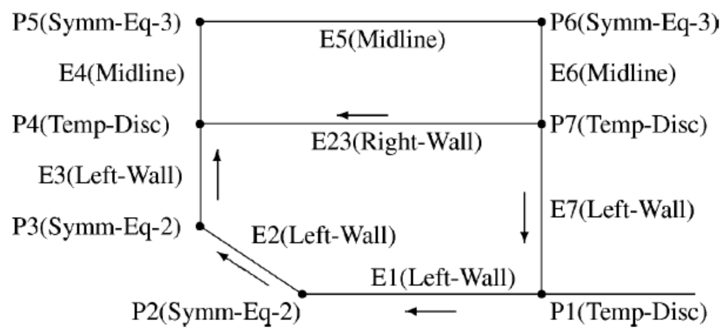
RSS 2007, Ananth's PhD Thesis 2008, ICRA 2009

Robotic Mapping

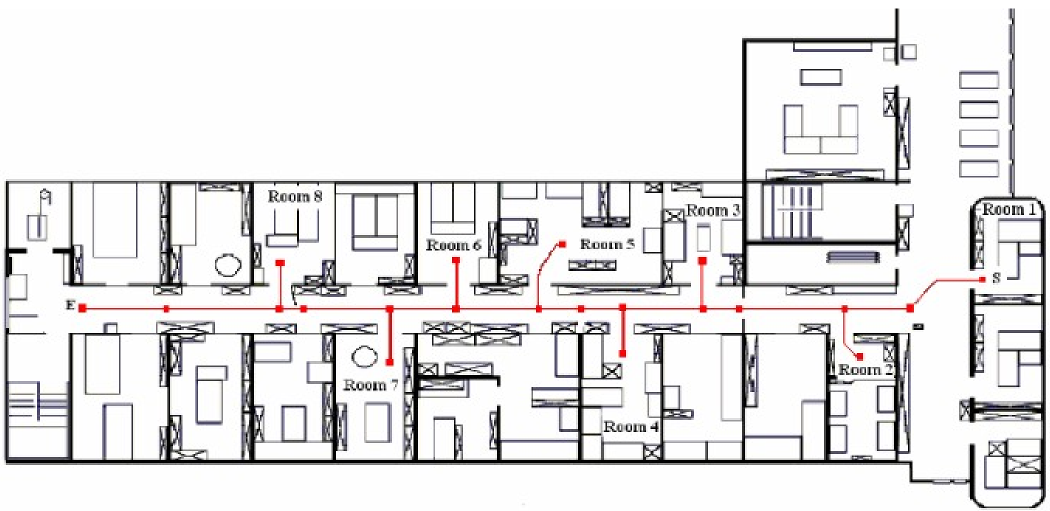
- Maps are essential for most autonomous tasks
- Care-giving
- Transport
- Domestic chores

Topological Maps

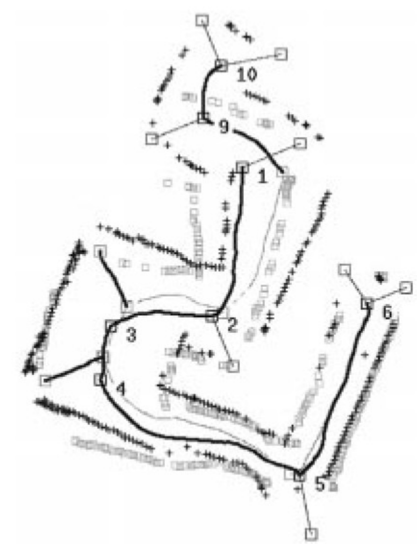
- Graphical representation
- Nodes denote places or regions
- Edges denote connectivity wrt the robot



Kuipers, 2000



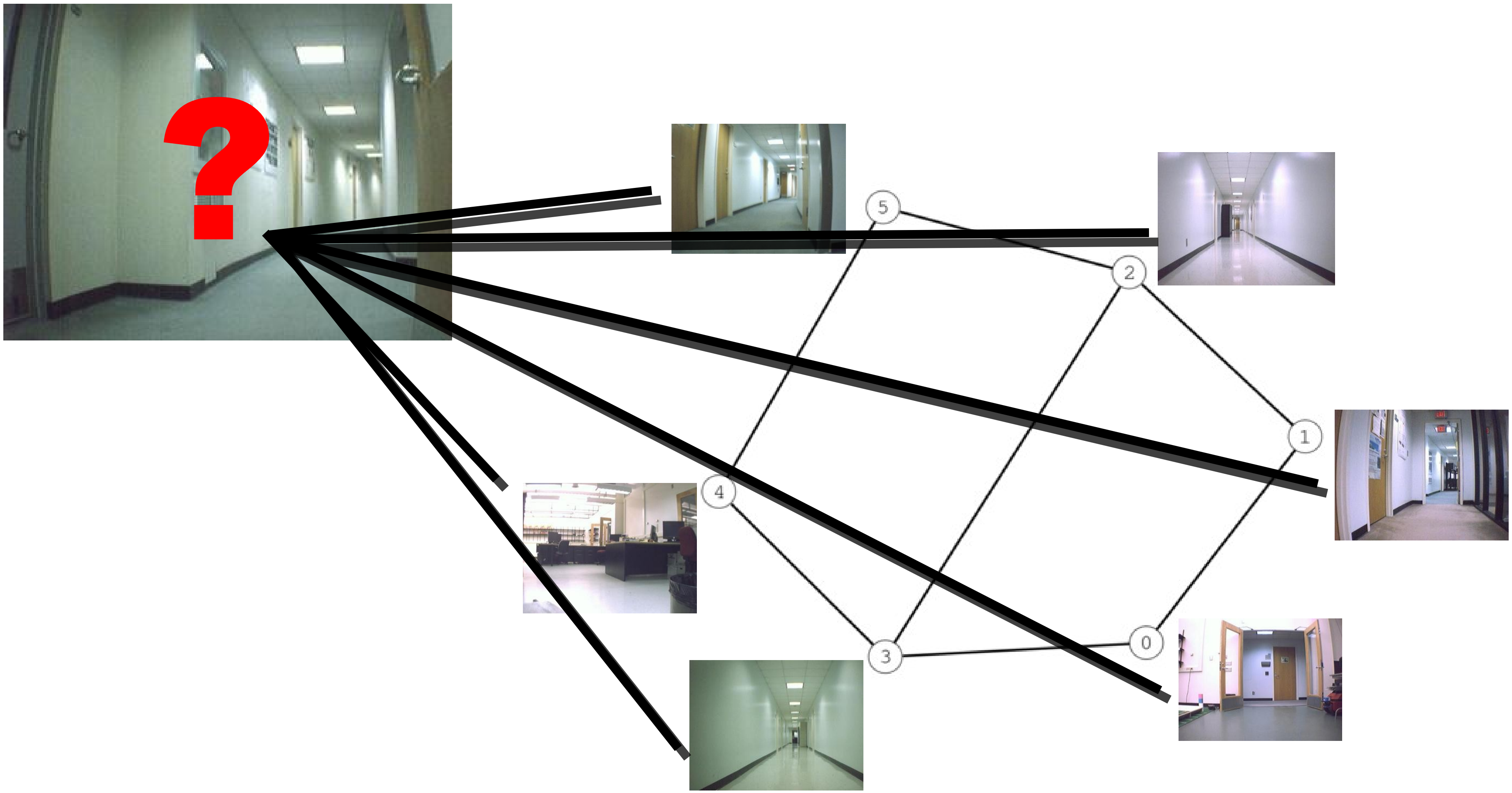
Tapus et. al., 2006



Choset & Nagatani, 2001

So: Main Problem in Topological Mapping

Have I visited this place before?



Aliasing and Variability




Place Recognition vs. Place Categorization

- Robots need something in between
- Specific Locations are fairly fluid as well
  - Office chair location changes all the time
  - Desks shift in appearance

Objects as semantic information

Lets go to the lounge with the coffee machine

The paper is in the printer room

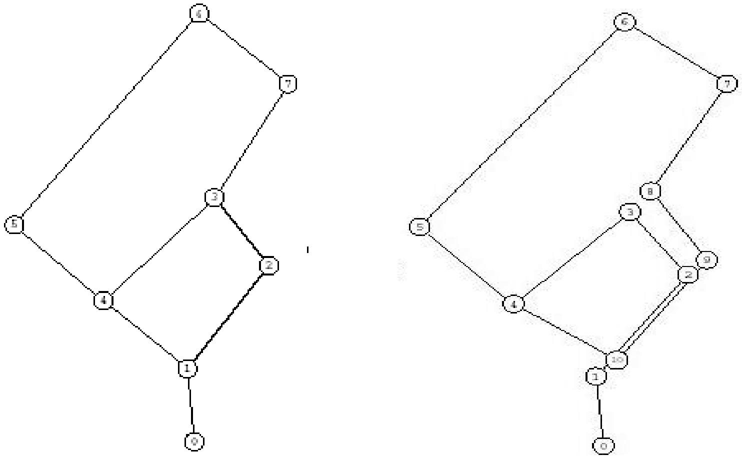
- Objects often define the places they are in
- Many canonical robot tasks involve identifying and manipulating objects

Objects are a good basis for semantic representations

Topological Mapping

No systematic **probabilistic** framework for topological mapping exists

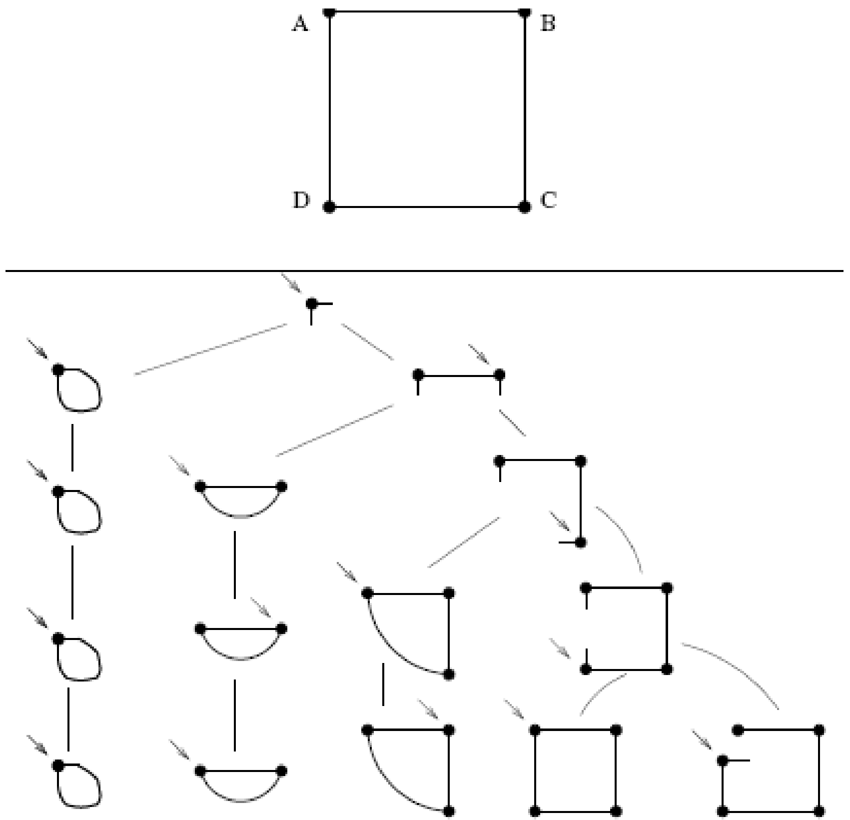
- Some Problems
- How to reason about discrete objects such as graphs ?
- How to deal with measurement noise ?



State of the Art

Maximum Likelihood is the preferred technique

- At each step, match current measurement with previous landmarks, and pick the best one

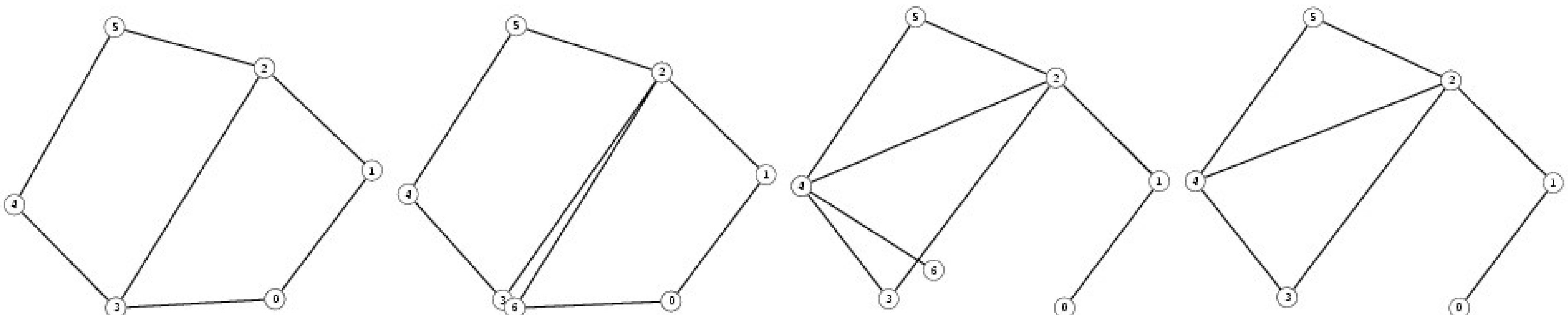


- Shatkay & Kaelbling, 1997
- Kuipers & Byun, 1991
- Choset & Nagatani, 2001
- Tomatis et al., 2003
- Tapus, 2005
- Savelli, 2005

Probabilistic Topological Mapping

Solution:

Define a probability distribution over the space of topologies



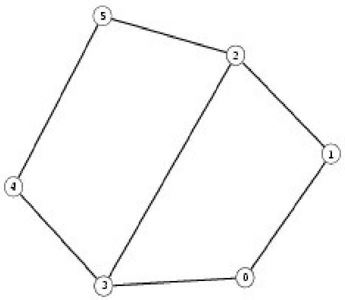
97% 2% 0.5% 0.5%

A Probabilistic Topological Map (PTM) is the Bayesian posterior on the space of topologies for a given the set of measurements

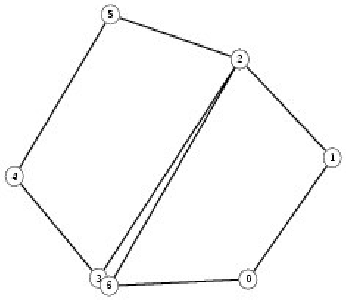
$$P(T|Z) \propto P(Z|T)P(T)$$

Overcoming Ambiguity

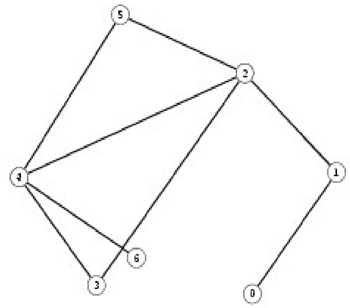
PTMs provide an estimate of ambiguity



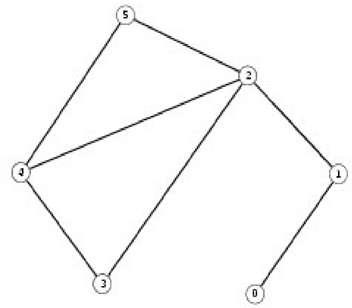
97%



2%

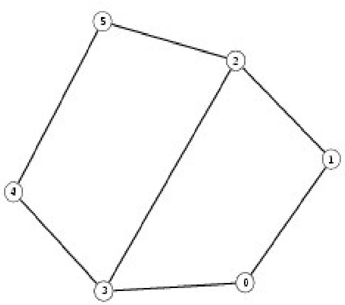


0.5%

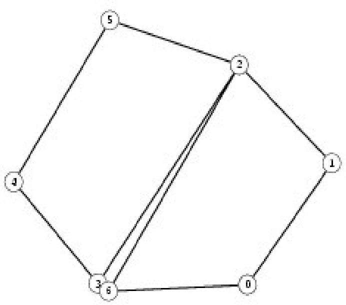


0.5%

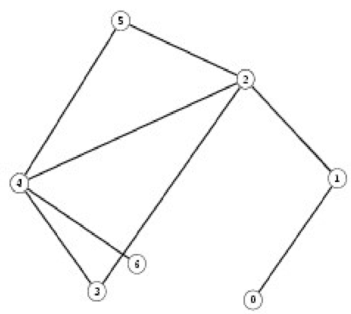
Low Ambiguity



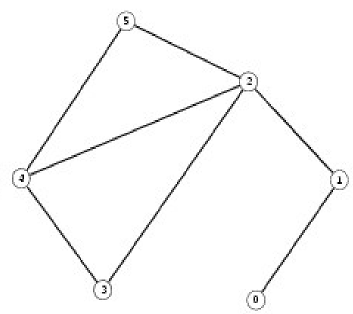
50%



49%



0.5%

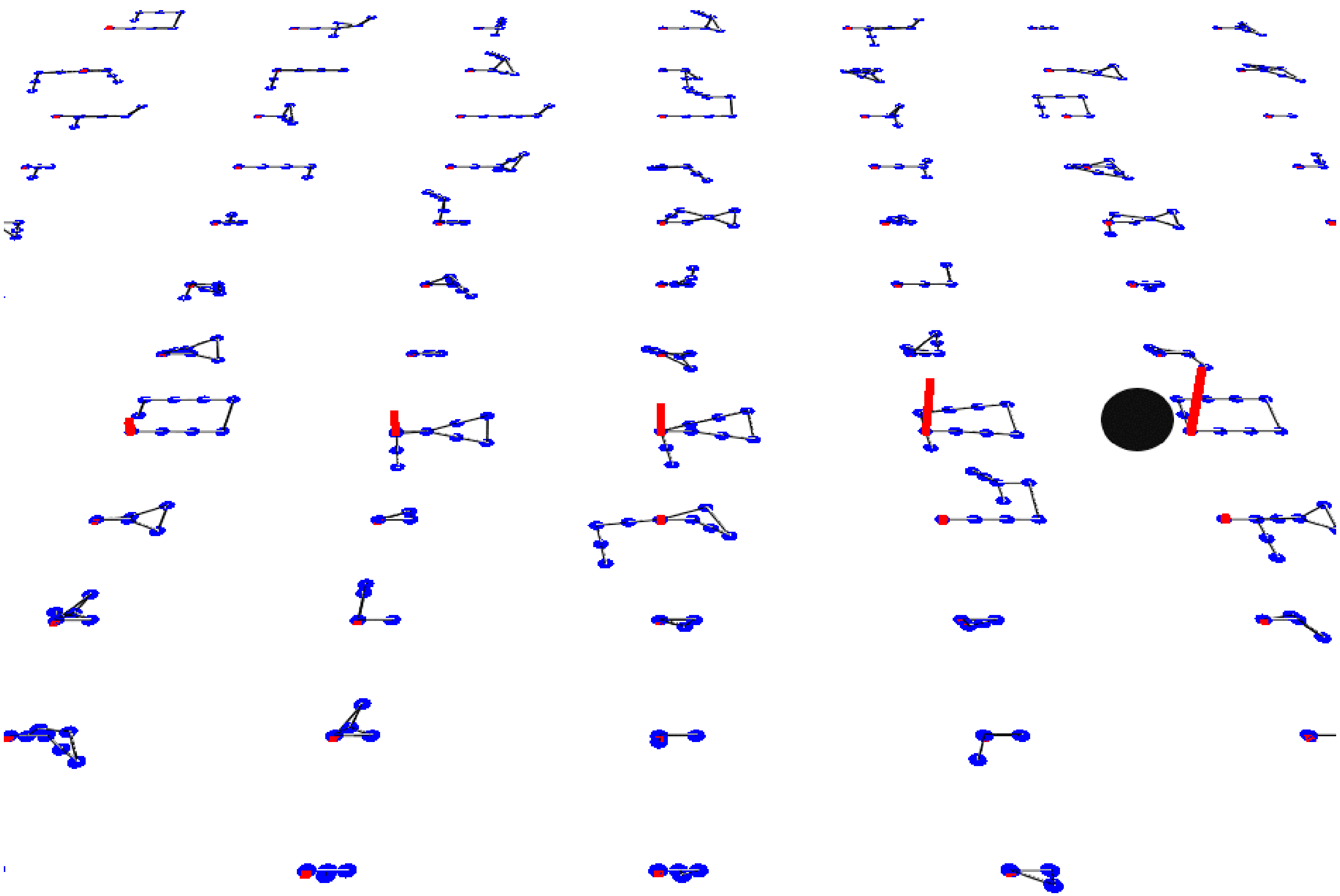


0.5%

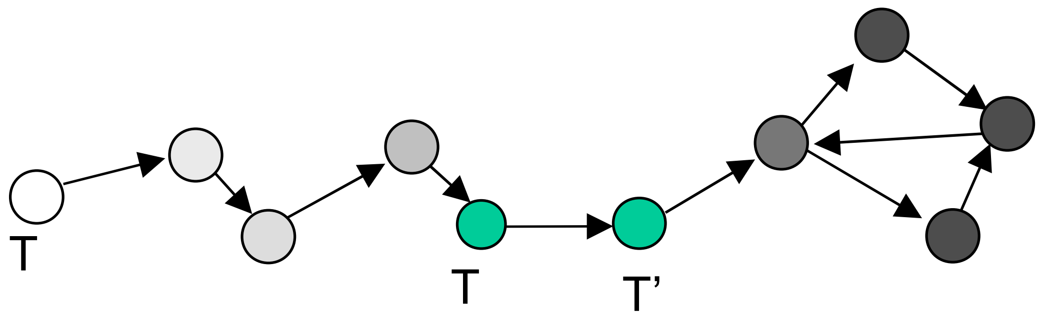
High Ambiguity

Markov Chain Monte Carlo for PTMs

MCMC can sample from arbitrary distributions in large spaces



Markov Chain Monte Carlo



- Start at T^0
- Propose a new topology $Q(T'|T^{(t)})$
- Accept the move according to the acceptance ratio

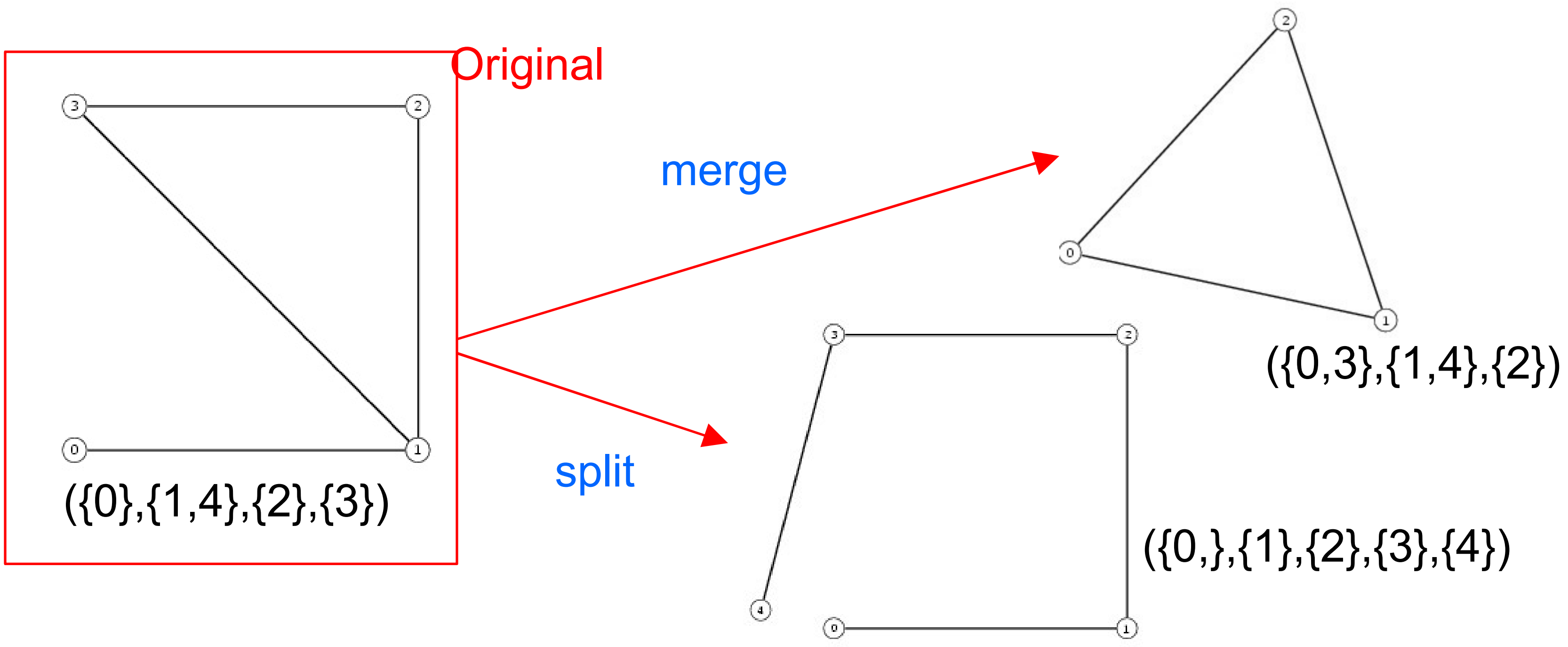
$$a = \frac{P(T'_t|Z^t)}{P(T_t|Z^t)} \frac{Q(T'_t \rightarrow T_t)}{Q(T_t \rightarrow T'_t)}$$

Target

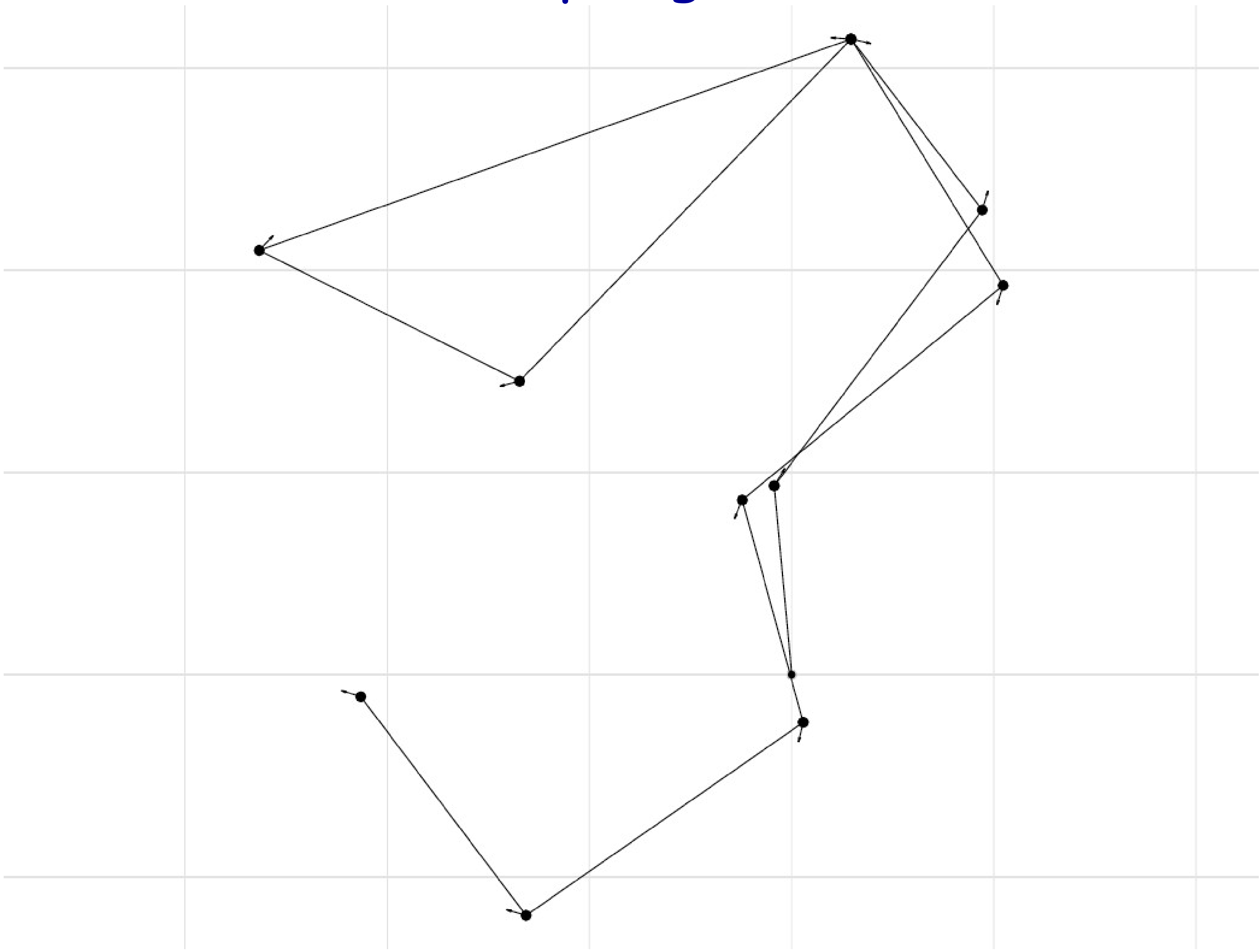
Proposal

MCMC Proposal

- **Split-merge proposal distribution**
- Split step – split a random set
- Merge step - merge two random sets

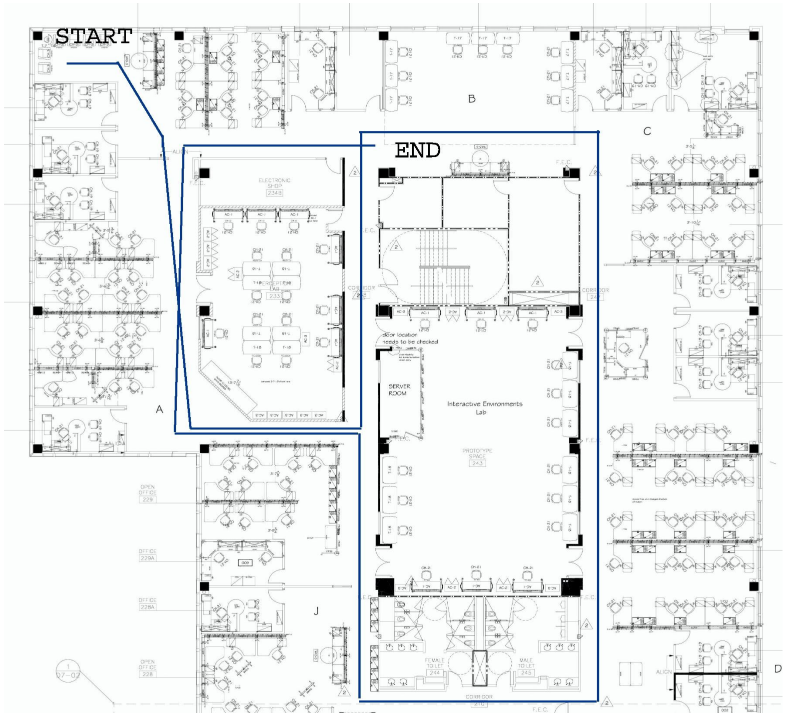


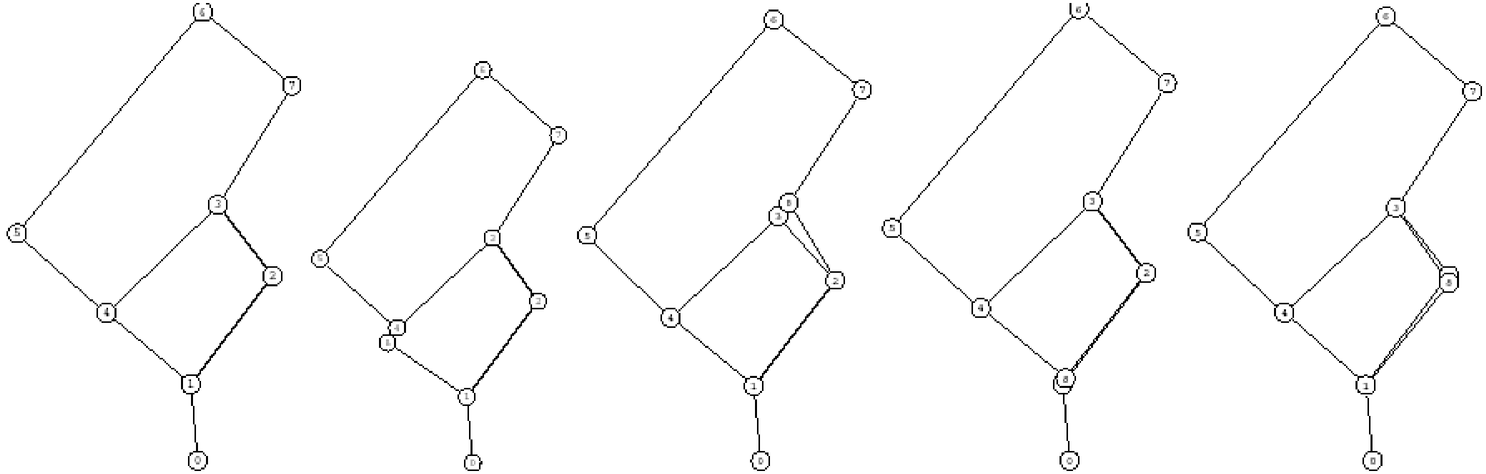
A Markov Chain on Topologies



Sample MCMC Result

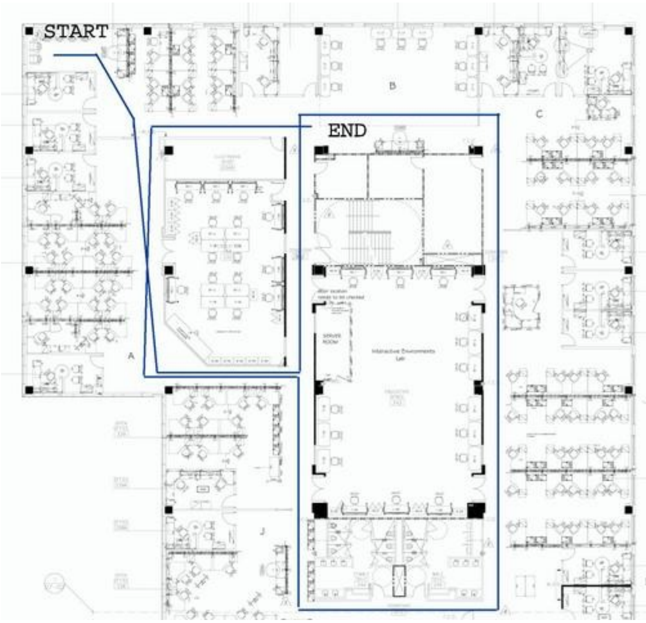
- TSRB dataset
- 12 hand-picked landmarks

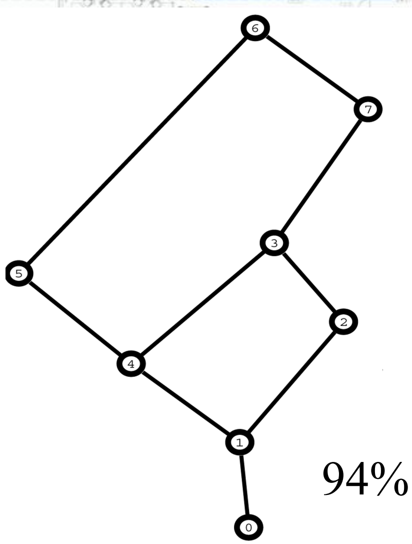


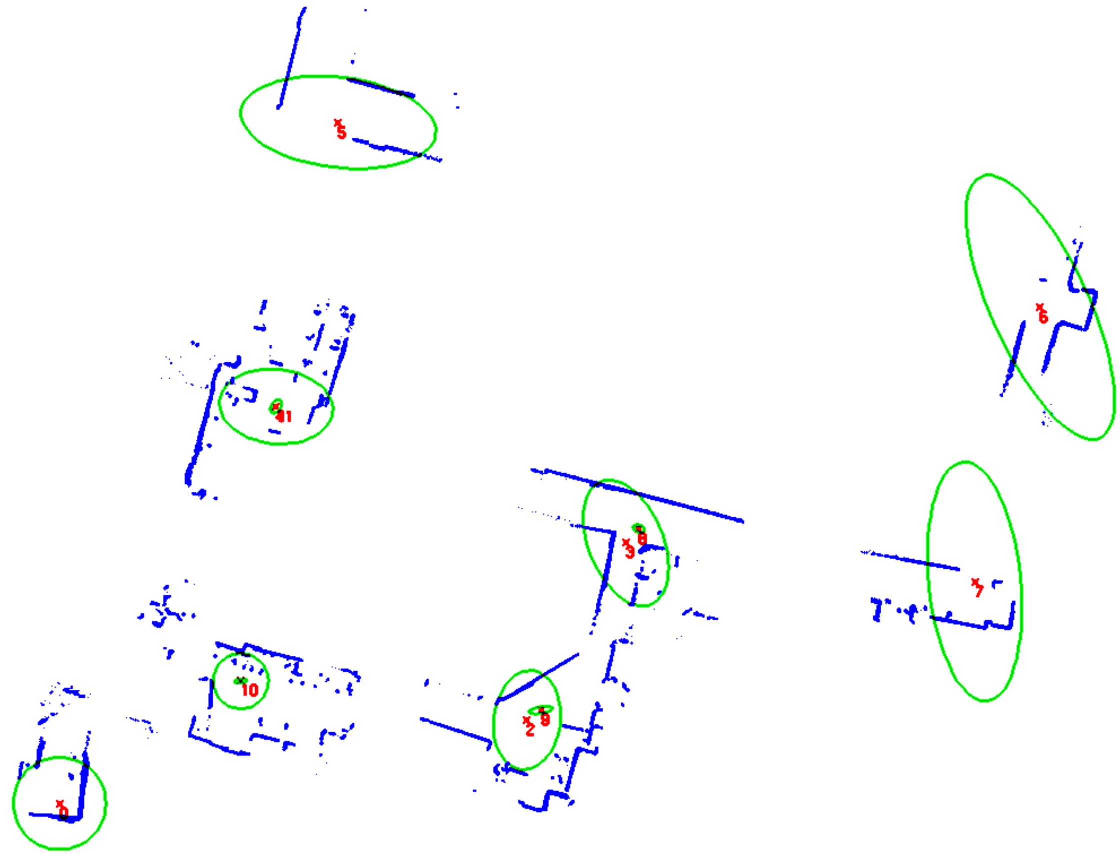


99.5% 0.25% 0.14% 0.12% 0.01%

Result - TSRB Dataset

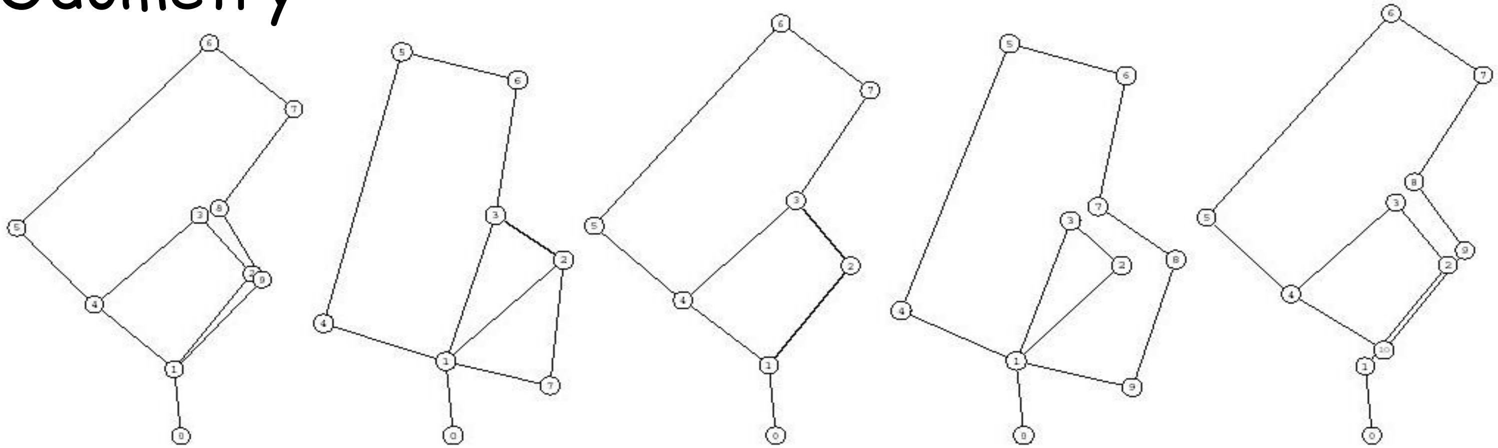






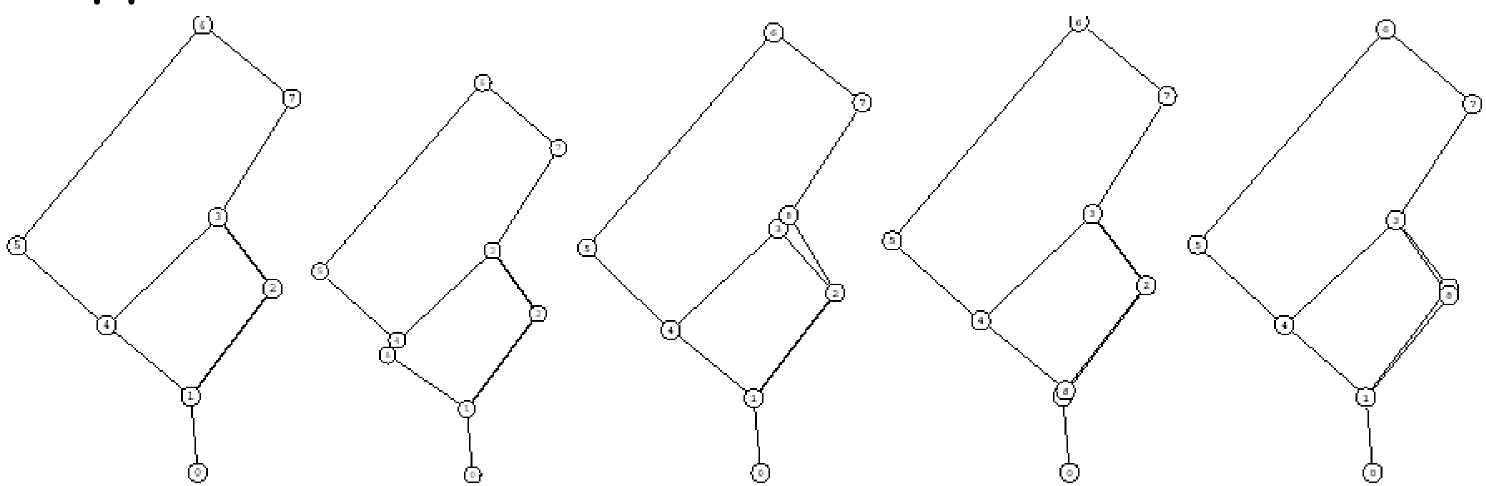
Can be done *without* sensors :-)

Only Odometry



43% 24% 7% 3% 2%

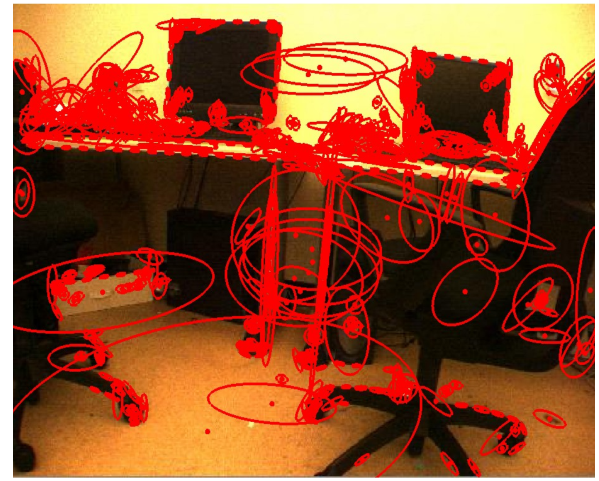
With Appearance

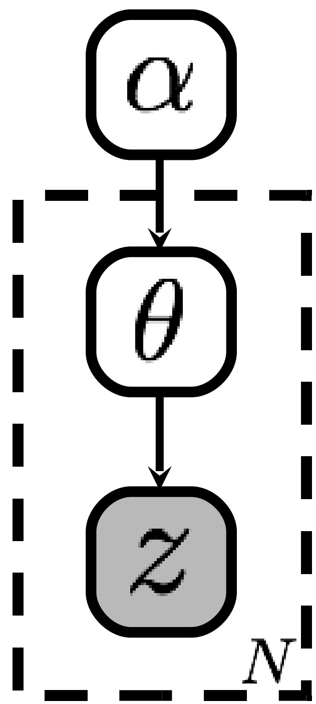


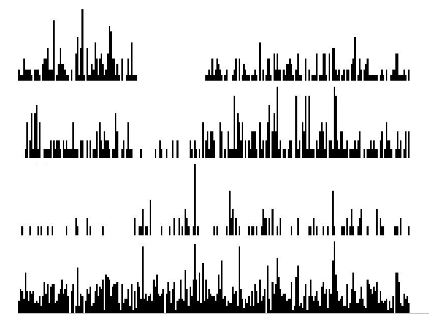
99.5% 0.25% 0.14% 0.12% 0.01%

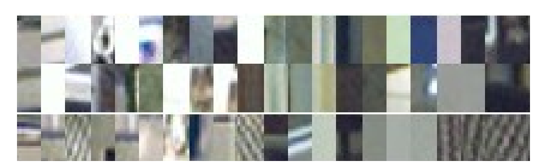
Visual SIFT Features

- Vector quantized SIFT features for appearance
- Feature histogram is measurement
- Multivariate Polya model for measurements





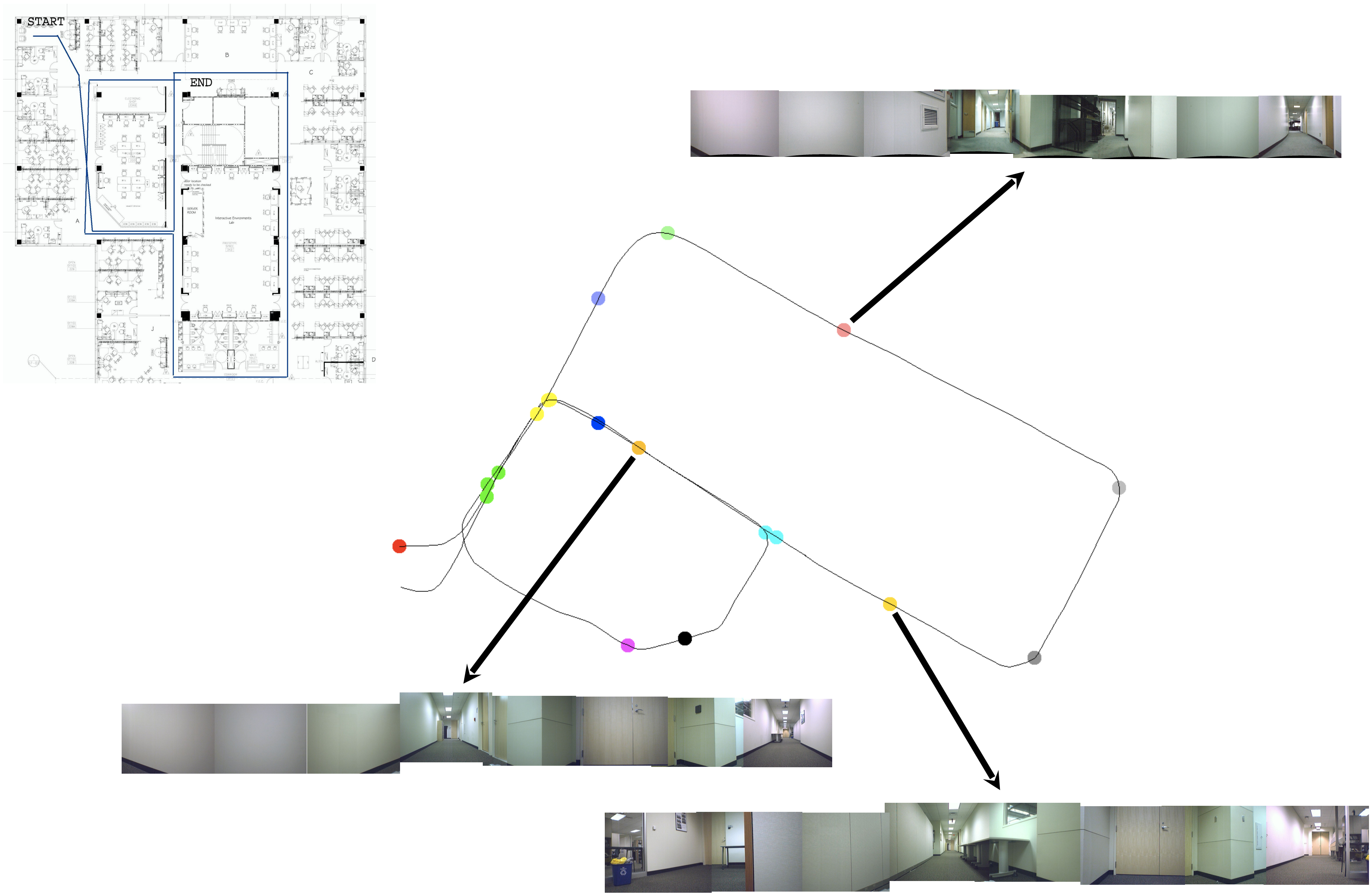




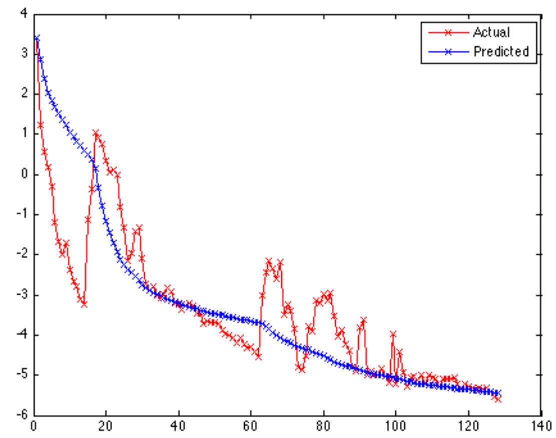
Landmark Detection: Bayesian Surprise

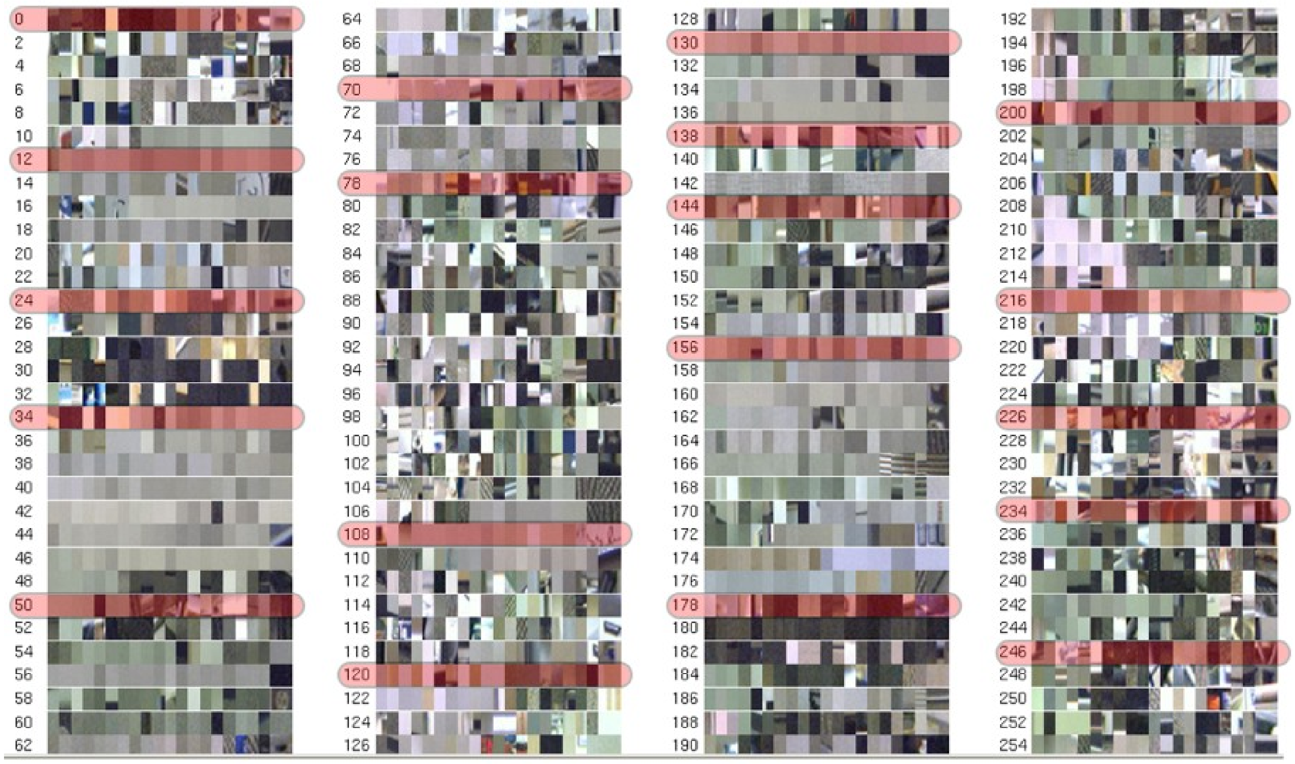
$$S(z) = \int P(M) \log \frac{P(M)}{P(M|z)}$$

TSRB Dataset



TSRB Dataset





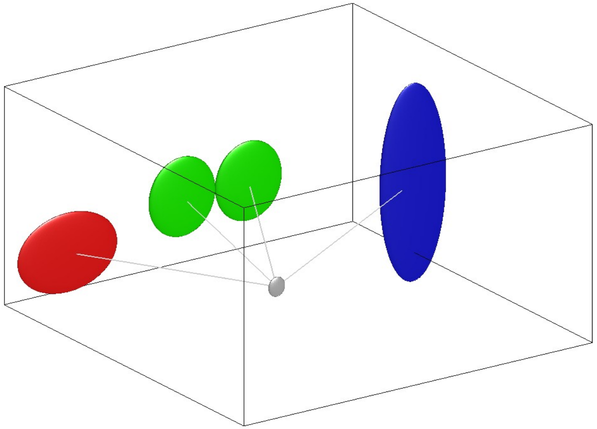
Place modeling using objects

Representation

- Represent a place as a set of objects along with their local 3D location

Given

- Fixed, specified object vocabulary
- Measurements – Stereo image pairs

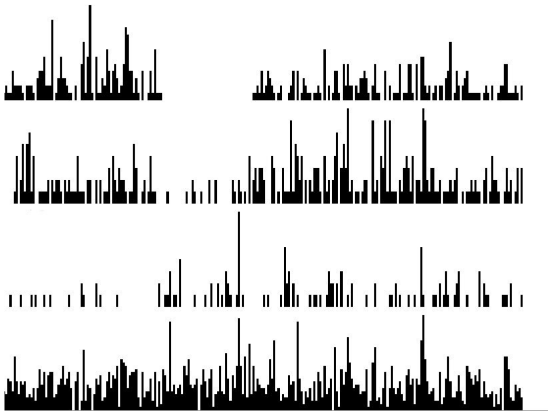


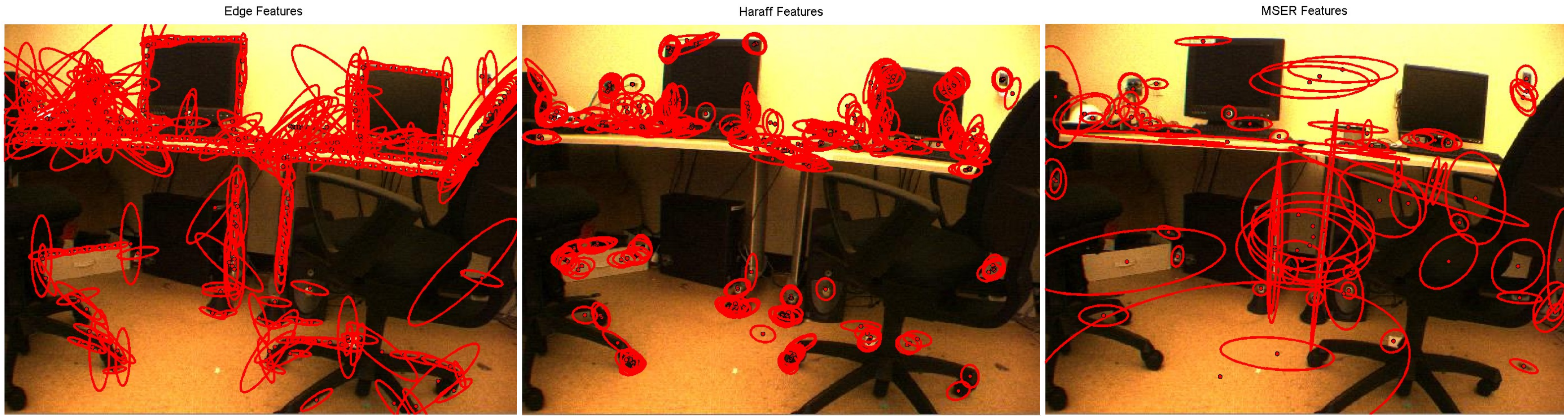
Semantic information is useful for solving tasks

- Office Assistant – "Where is the copier located?"
- Domestic robot – "When should the laundry be removed from the washing machine?"
- Transportation robot – "How much more fuel will be used if this diversion is taken?"

Object Modeling

- Object shape is modeled as a 3D Gaussian
- Appearance histogram on quantized feature descriptors
- 3 types of complementary features
  - Canny edges, Harris corners, Maximally Stable Extremal Regions (MSER)





Related work

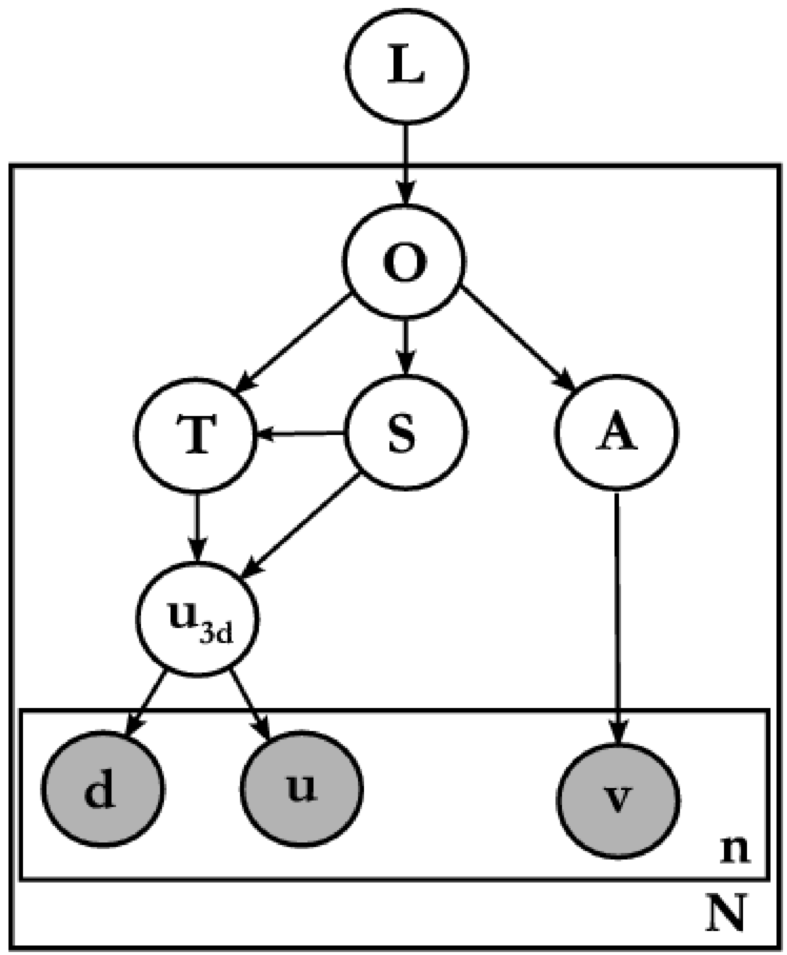
- Previous work in robotics has mainly modeled small objects using SIFT matching
  - Ekvall et. al., Image and Vision Understanding, 2005
  - Vasudevan et. al., RAS, 2007
- Approaches from vision
  - Sivic et. al., ICCV 2005
  - Sudderth et. al., CVPR 2006.
- Joint modeling of object shape and appearance has proven powerful
  - Fergus et. al., CVPR 2005
  - Crandall et. al., CVPR 2007

Object Representation for Place Modeling

- Objects need to be modeled before places can be modeled
- Objects have a **shape** and an **appearance**
- Object models are learnt during a training phase

- Object models are used to infer place models from stereo images

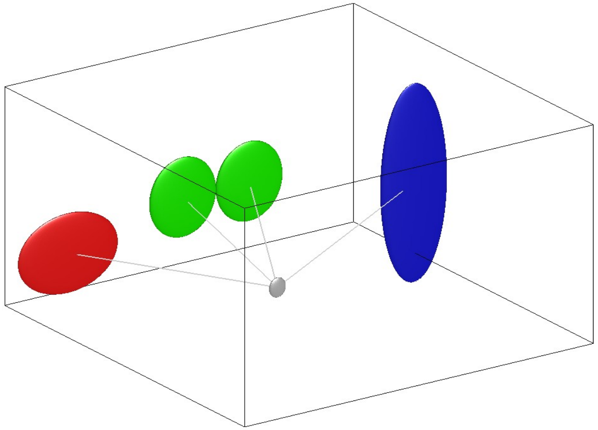
A Generative Model for Places



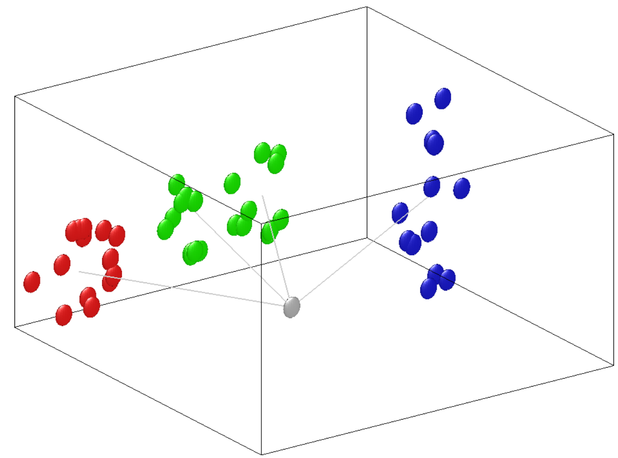
A Generative Model for Places

Place Label L

Objects
3D Location T
Appearance
Shape

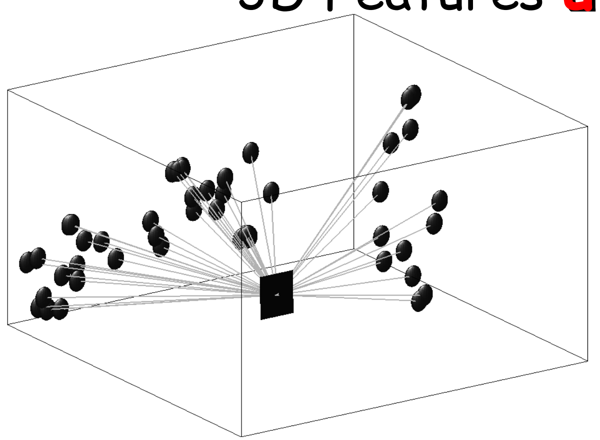


3d



3D Features u

V Feat. Appearance
Pixel Location U
Depth

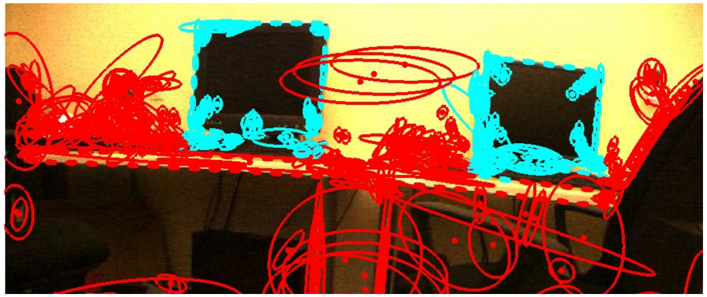


3d

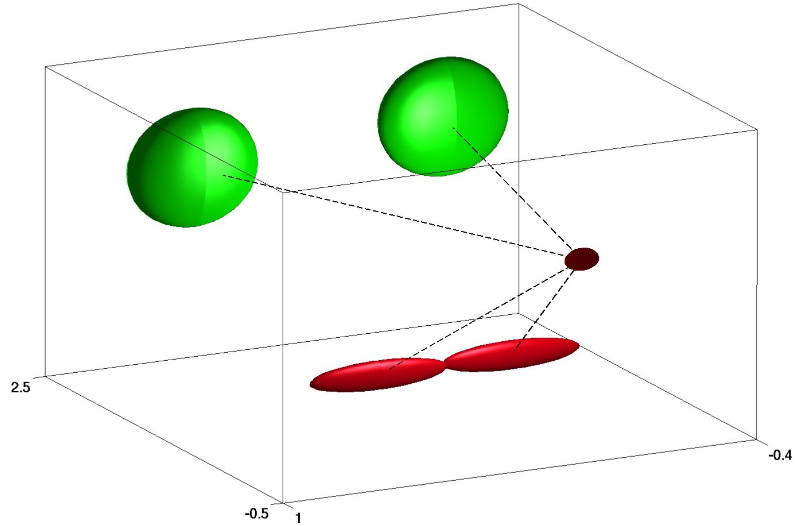
$P(L,O,T,A,S,u,D,U,V)$

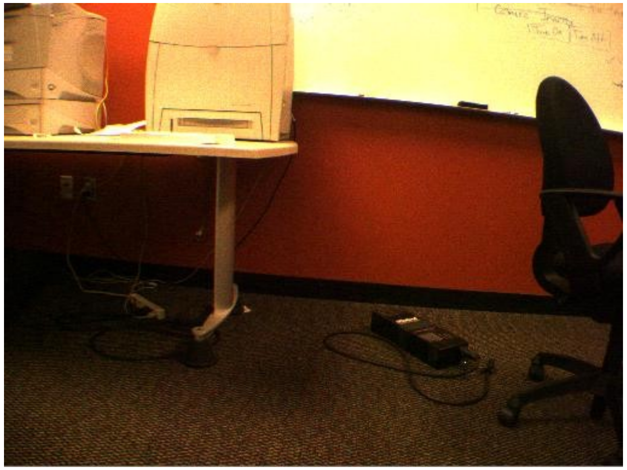
Supervised Learning of Object Models

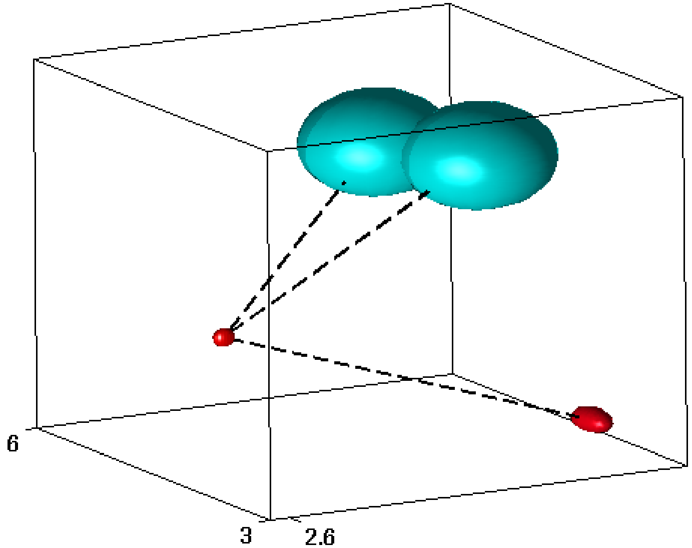
- Objects are roughly segmented out in training
- Appearance is learnt from features
- Shape Gaussian is learnt from 3D feature locations
- Vocabulary of 5 objects
  - Monitor, drawer, chair, printer, cupboard
- Training set of 68 images



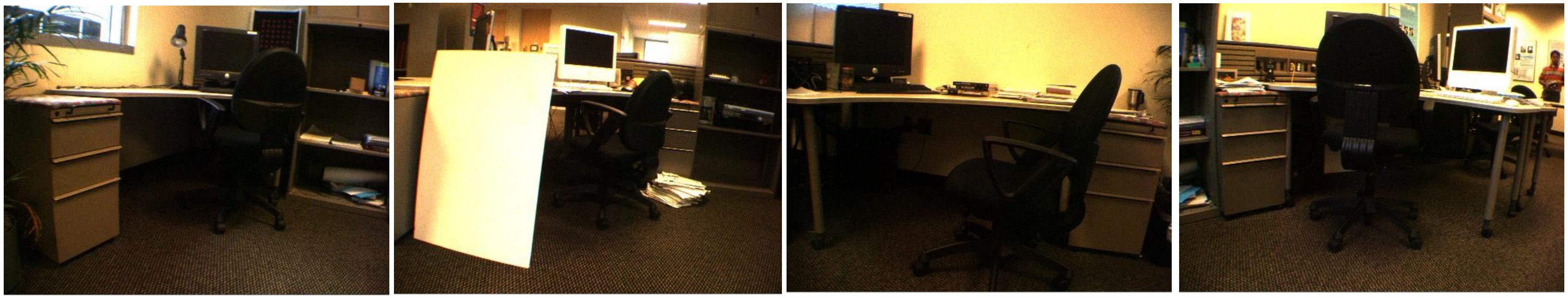
Place Models

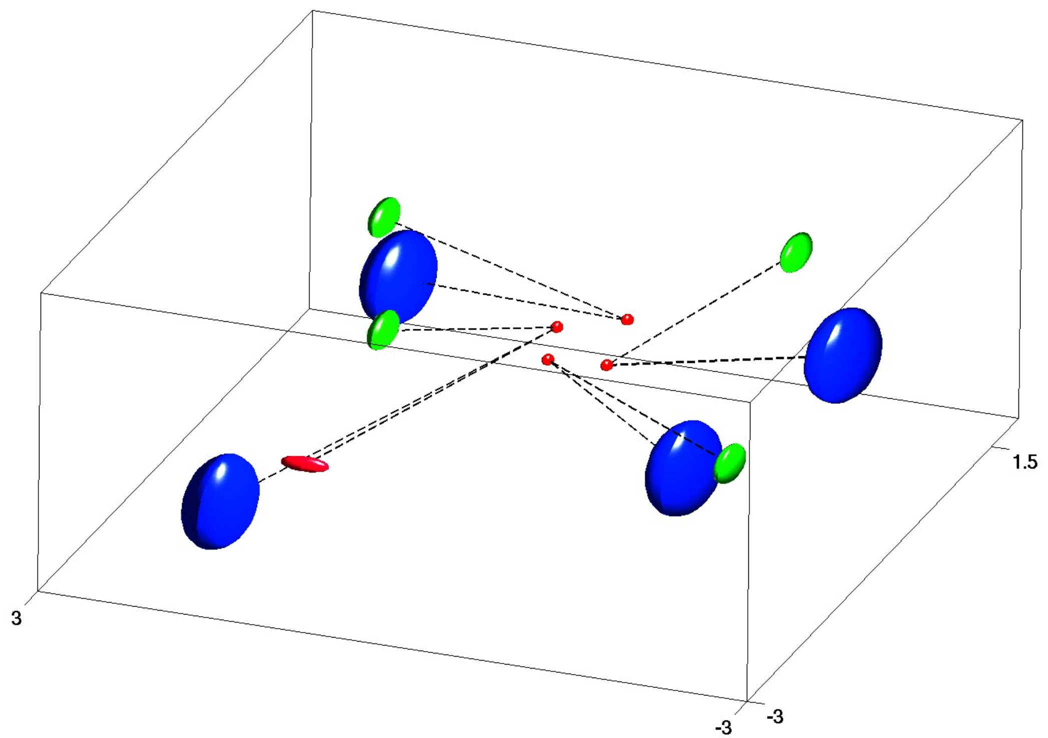






Panoramic Place Models

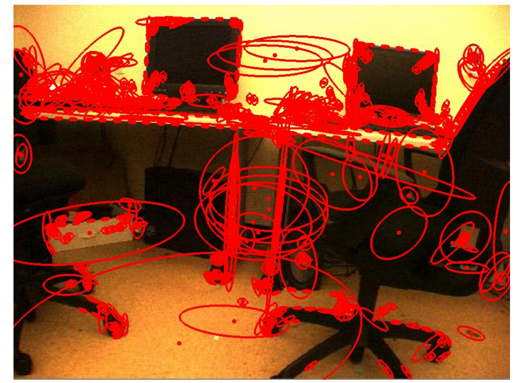




Inference for Place Models

Inference is the inverse of generative process

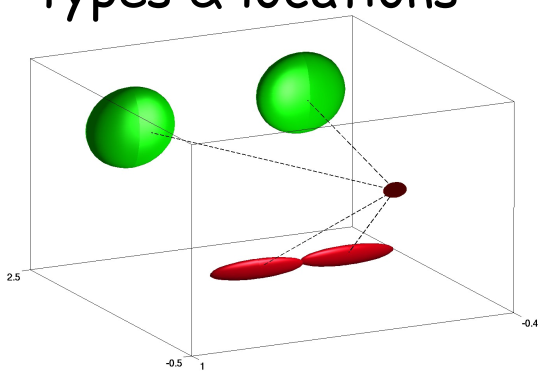
Stereo image pair



Detect features

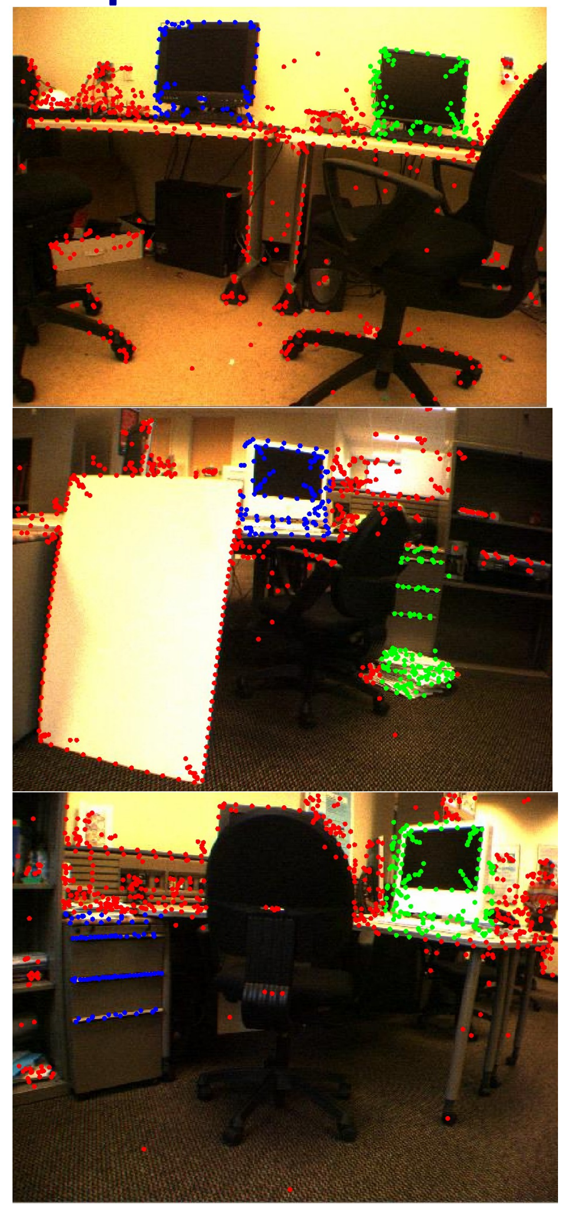
Segment objects

Compute object types & locations



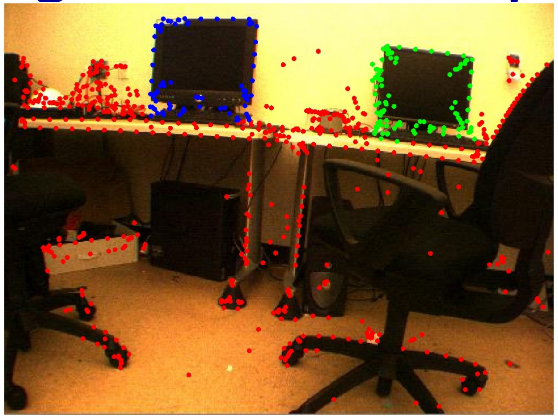
Object Segmentation / Correspondence

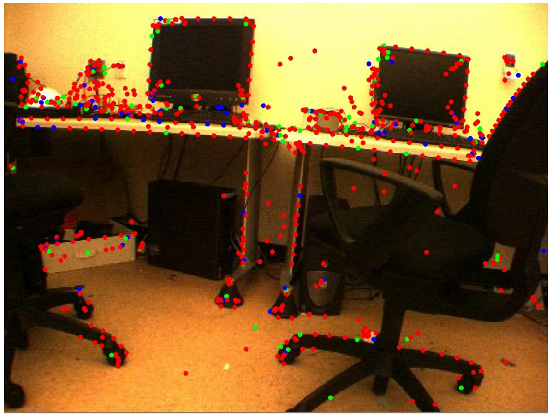
- How to find the correct segmentation without knowing which objects are in the image?
- This is a correspondence problem



Answer : Sample over segmentations

Sampling over Correspondences



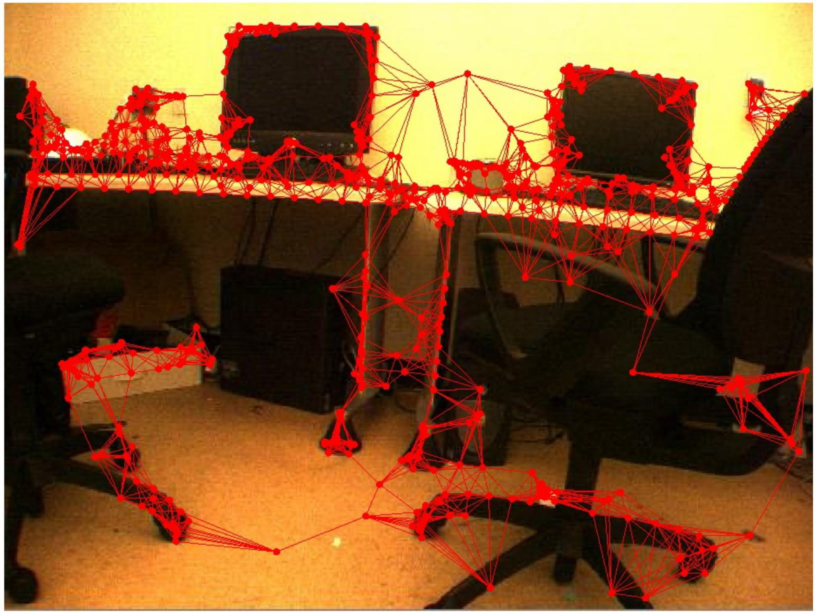


- Features belonging to objects appear in clusters
- Objects have distinct appearance distribution
- Object boundaries display sharp depth discontinuities

Idea

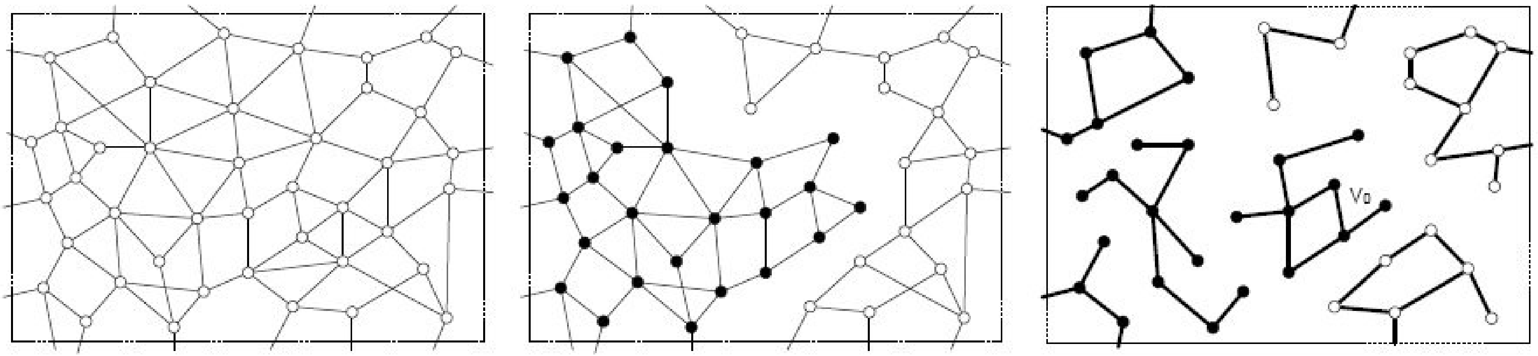
Place features in a Markov Random Field that encodes this information

Feature-based Markov Random Fields



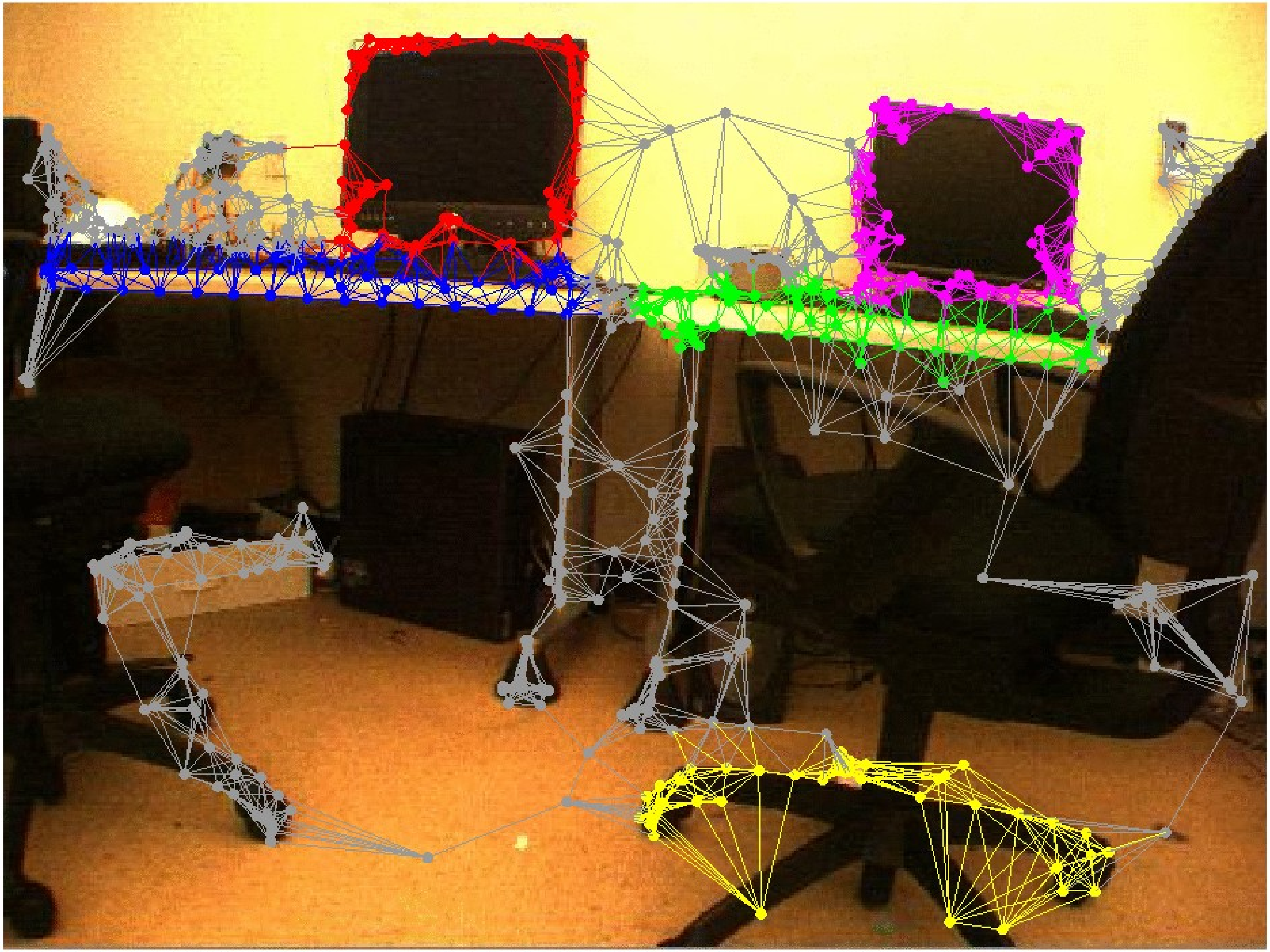
- Create MRF by connecting each feature to k nearest neighbors in the image
- Discriminative probability on each edge
- Probabilities encode
  - Differences in appearance
  - Differences in 3D location
  - Image distance

Sampling over segmentations using Swendsen-Wang cuts



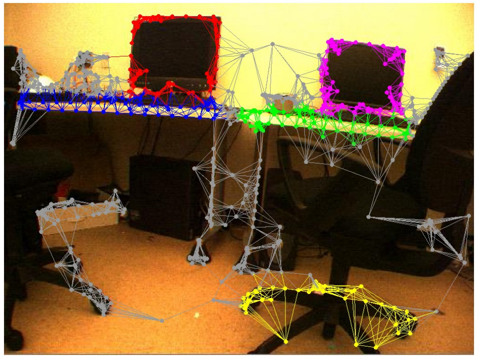
- For each sample
- Take the MRF and turn each edge on/off according to the discriminative edge probability
- Obtain connected components of the graph (the segmentation)
- Assign objects to the components of the segmentation and compute its likelihood

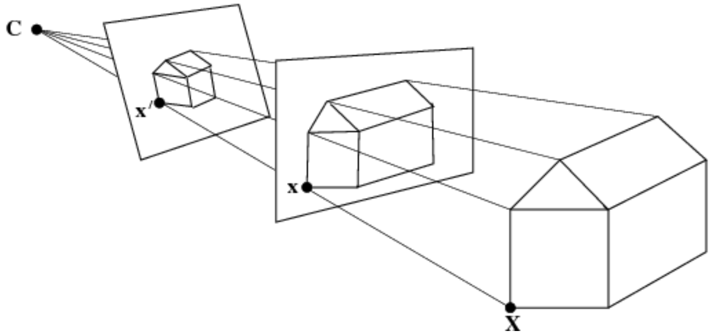
Sampling over segmentations

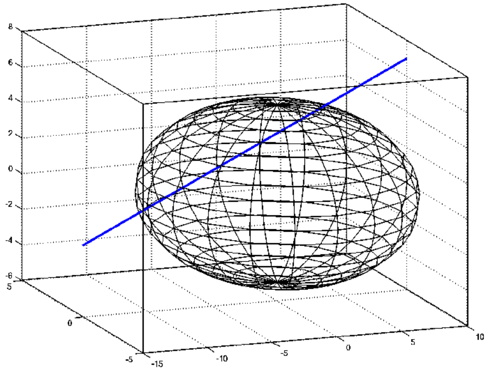


Sampling object types

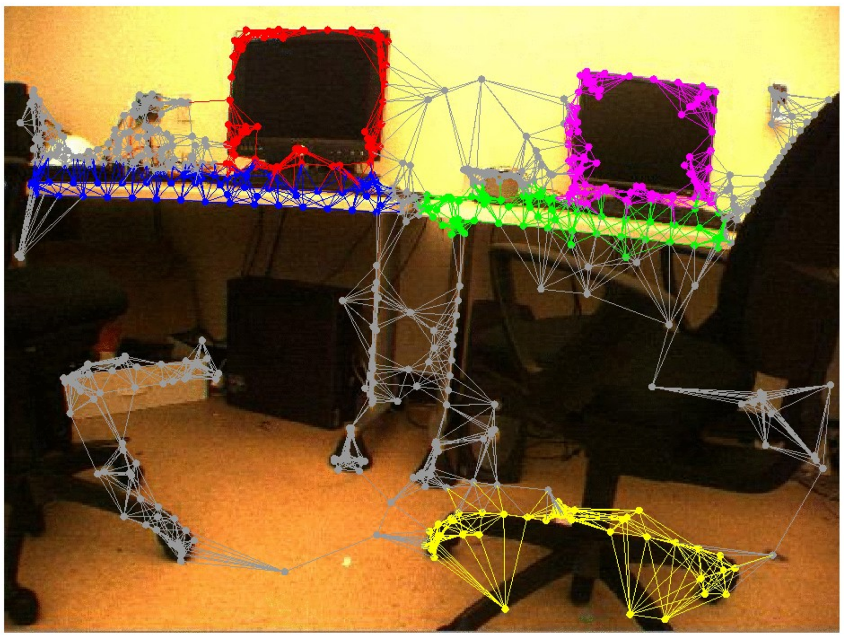
- Assign objects to each component by sampling from a prior distribution
- Evaluate the likelihood for the assignment
- **Appearance likelihood** for each component
- **Scale of object** provides 3D location
- **Stereo depth likelihood** provides location

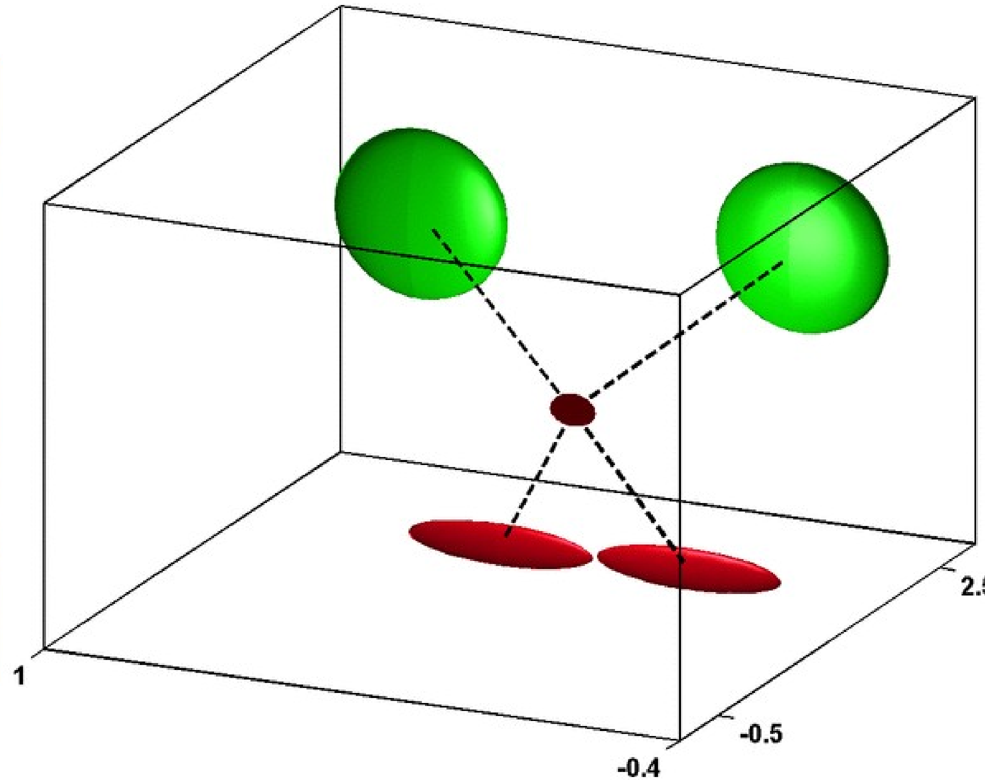






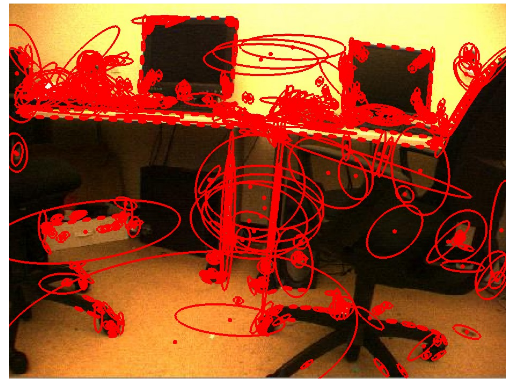
Fitting objects to a segmentation



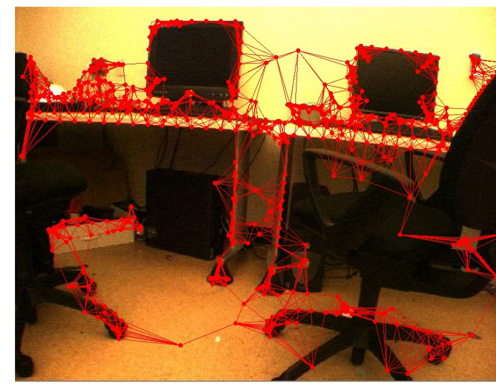


Summary

Stereo image pair

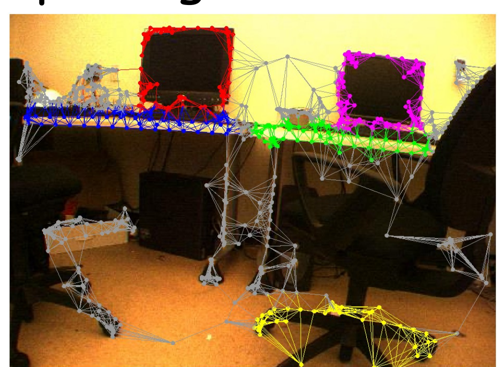


Detect features

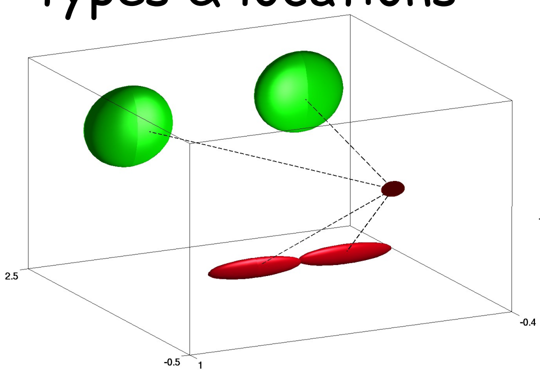


Create MRF

Sample segmentations

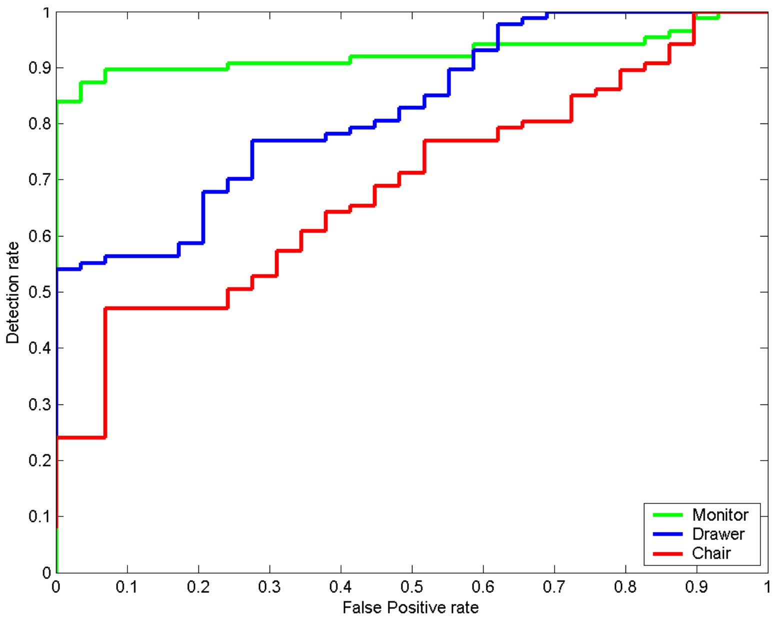


Compute object types & locations



Results – Object Recognition

- Tested on images containing only a single object
- Criterion – MAP result should contain the correct object type
  - Location was not considered
- Results compare favorably to PLSA recognition rates



Results - Place Recognition

	1	2	3	4	5	6
1	0.98	0	0	0	0.02	0
2		0.58	0.42	0	0	0
3			0.58	0	0	0
4				1	0	0
5					0.98	0
6						1

Future Directions

- Better modeling
  - Inhomogeneous objects, large viewpoint changes
- Faster computation
  - Graph cuts, subgraph matching
- What objects to model?
- Objects may be moved about
- Semantic mapping with objects