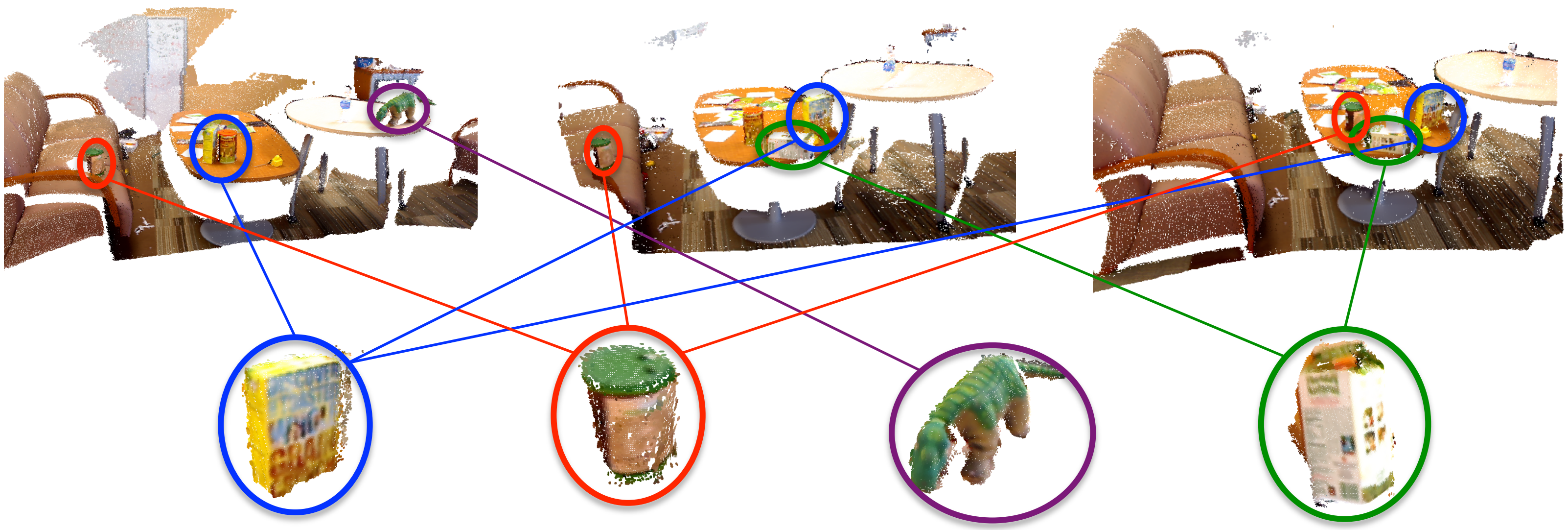
RGB-D Object Discovery via Multiscene Analysis



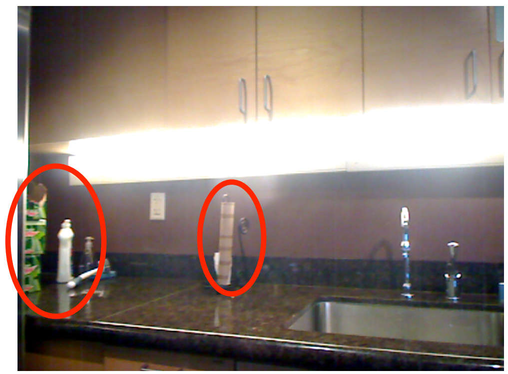
Evan Herbst, Xiaofeng Ren, Dieter Fox

University of Washington / Intel Labs Seattle

Problem Description

- **Goal:** identify and model objects using change over time

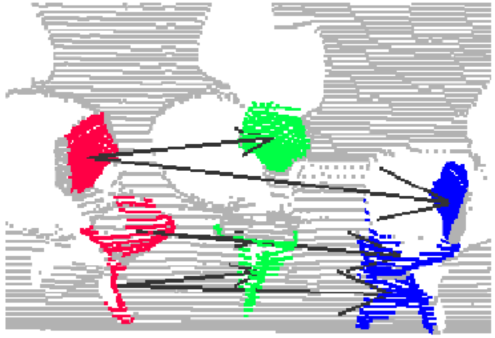


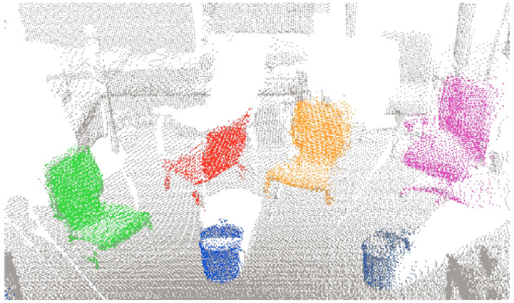


  - Handle textureless objects
  - Avoid appearance/shape priors
- Represent a location as static + dynamic parts

Object Discovery from Range Data

- [Shin et al, ICRA10, RSS10]: ICP initialized with local 3-D feature matches
- [Ruhnke et al, ICRA09]: 2-D feature matches in generated views

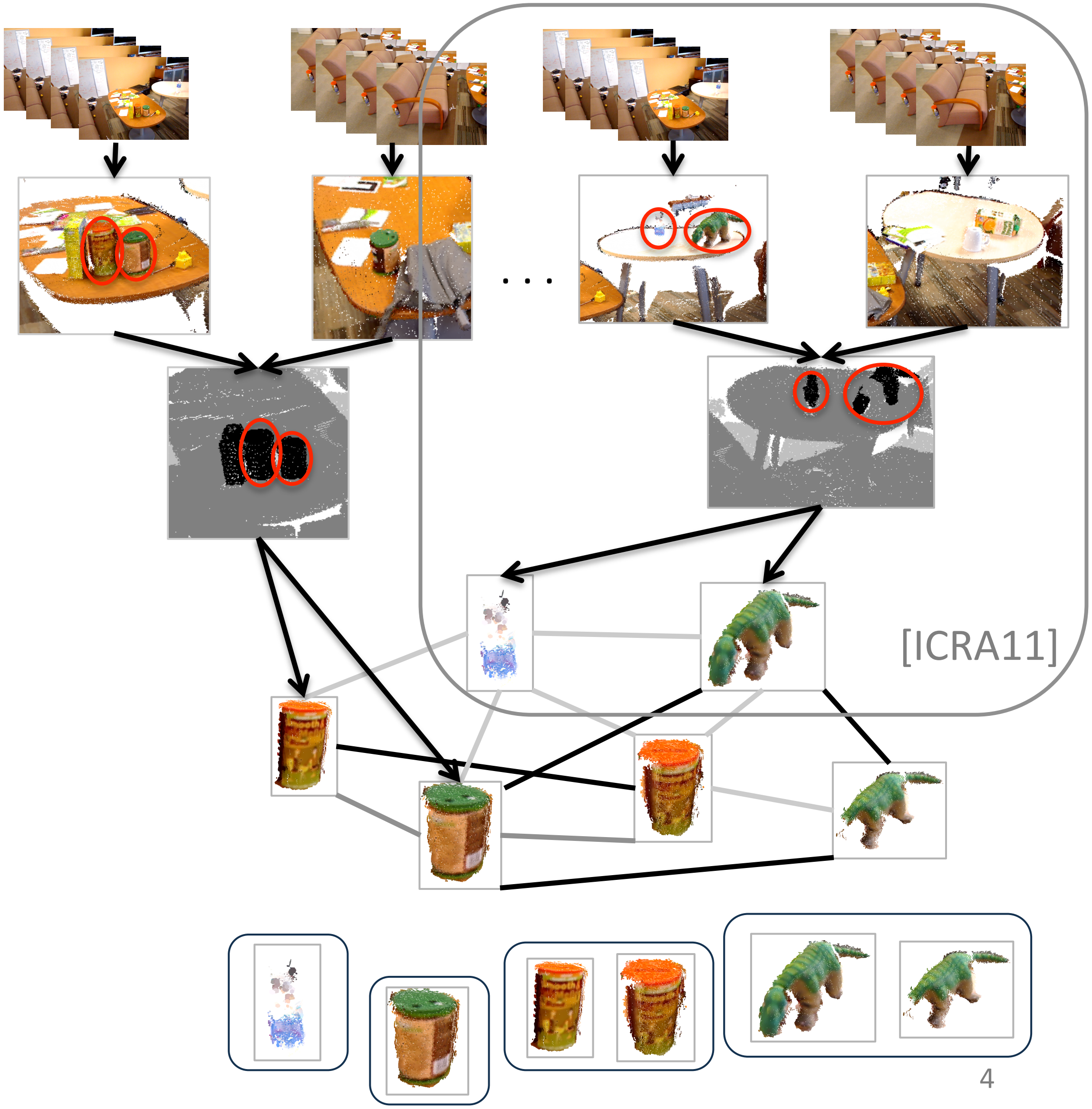




- Single scenes
- Focus on detecting repetitive structure, model merging
- No visual information

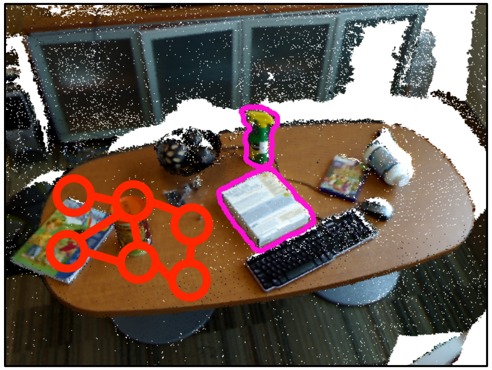
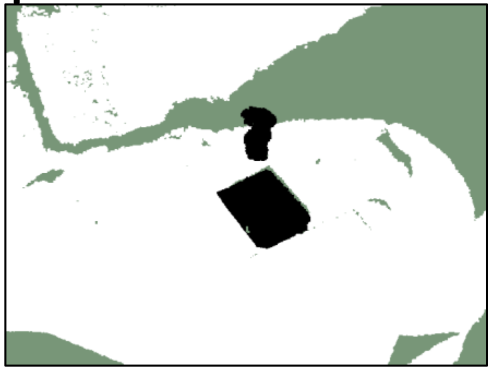
Method Overview

1. SLAM
2. Change detection
3. Segmentation
4. Similarity
5. Object discovery



Two-Scene Segmentation (ICRA11)

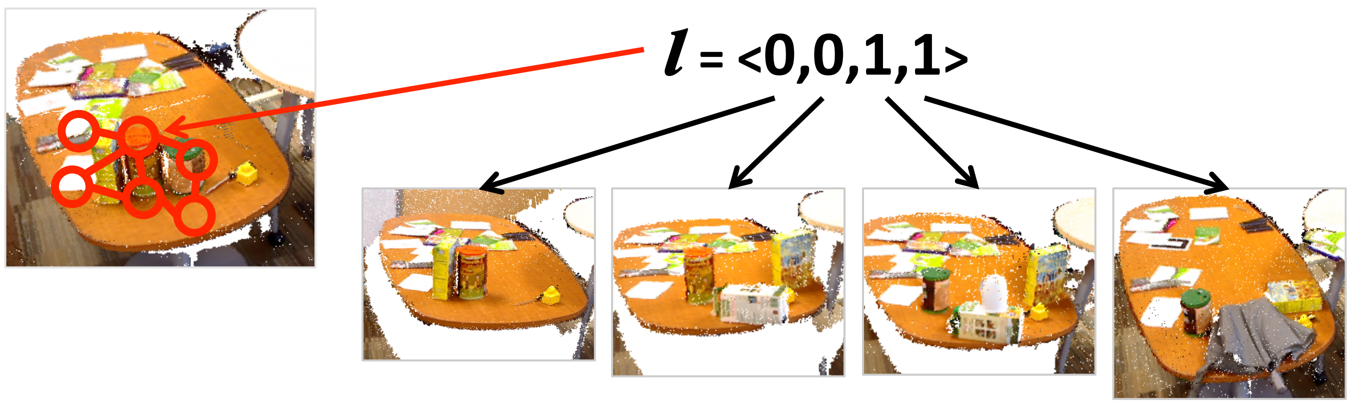
- Pointwise change detection
  - probabilistic sensor measurement model
  - Incorporate depth, color, surface orientation
  - Handle occlusion
  - View-based: accumulate evidence from different viewpoints
- MRF over scene points to reduce noise


- Limitations
  - Can only make use of two scenes
  - No notion of segment persistence

Multiscene Segmentation

- Extra scenes provide extra constraints
- MRF to assign change labels wrt all scenes at once



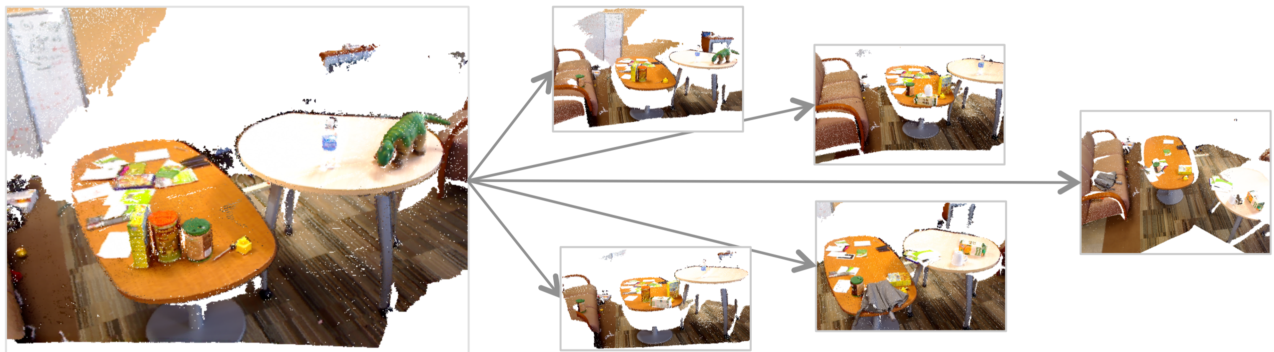
- Data cost based on change detection

$$E_D(s,l) = \sum_j \begin{cases} p(s \text{ moved in } j), & l_j = 0 \\ 1 - p(s \text{ moved in } j), & l_j = 1 \end{cases}$$

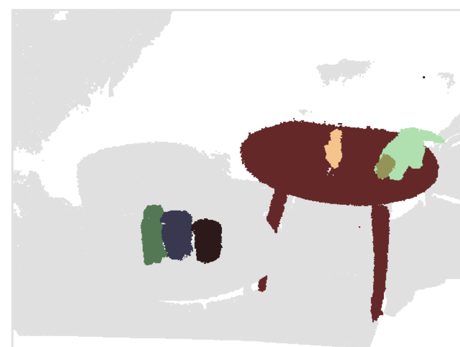
- Smoothness cost: Hamming distance weighted by curvature

Segmentation Results

- Concatenated pairwise MRFs vs. multiscene MRF



two-scene MRFs

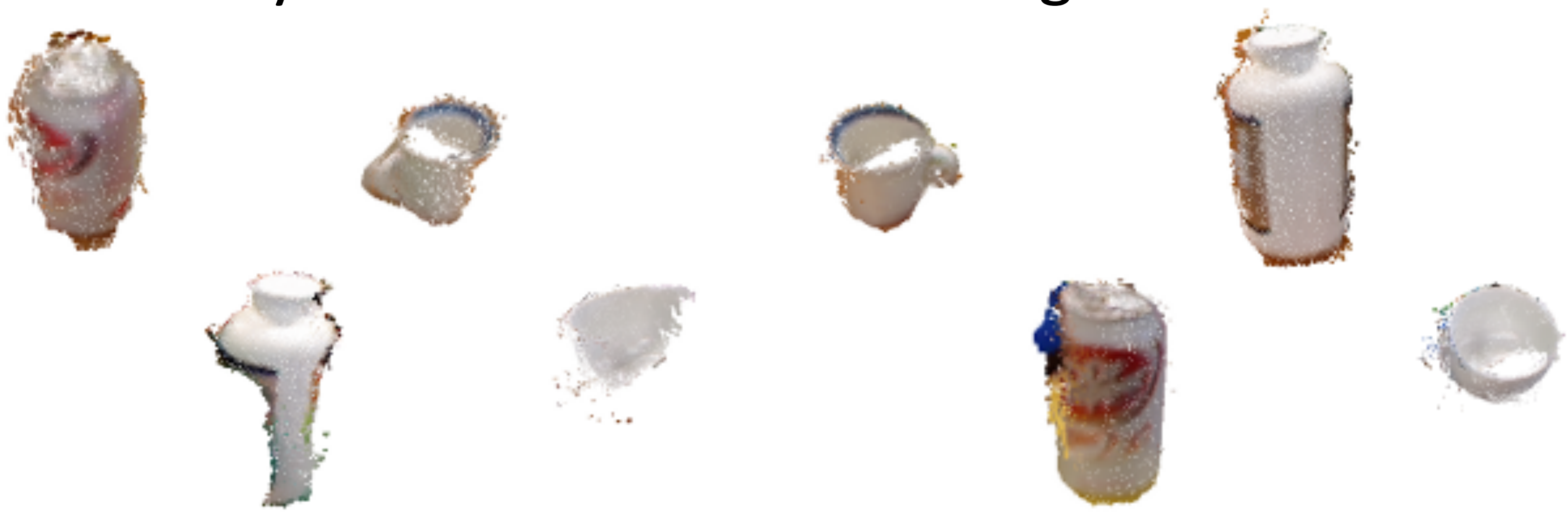


multiscene MRF

- Over 39M points (12 scenes), 25% reduction in each of type I/II errors in change detection

Object Discovery

- Output of segmentation step:
  - Partial 3-D models
  - Many 2-D views of each 3-D segment



- Goal: group these segments by object

Object Discovery: Spectral Clustering

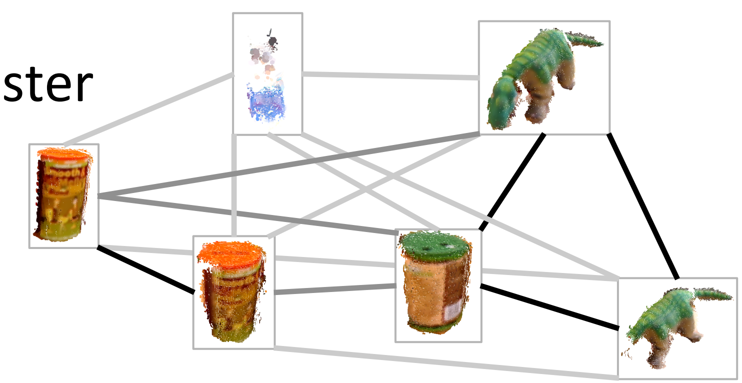
- Given pairwise similarities (edge weights in a graph whose nodes are segments)

$$[W]_{ij} = w_{ij} \quad , \quad i, j \in V$$

- Minimize the *normalized cut* criterion

$$\arg\min_{A \subset V} \left(cut(A) + cut(\bar{A}) \right) \text{, where} \quad cut(A) = \frac{\sum_{i \in A, j \notin A} w_{ij}}{\sum_{i \in A, j \in V} w_{ij}}$$

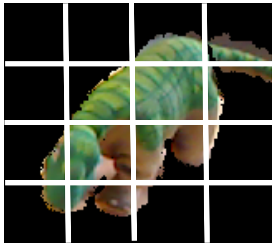
  - Minimize similarity across clusters
  - Maximize similarity within each cluster



Need a similarity measure over segments

Segment Similarity Scores

- 2-D matching with whole-segment features
  - Spatial pyramid of local descriptors
  - Score is descriptor vector distance

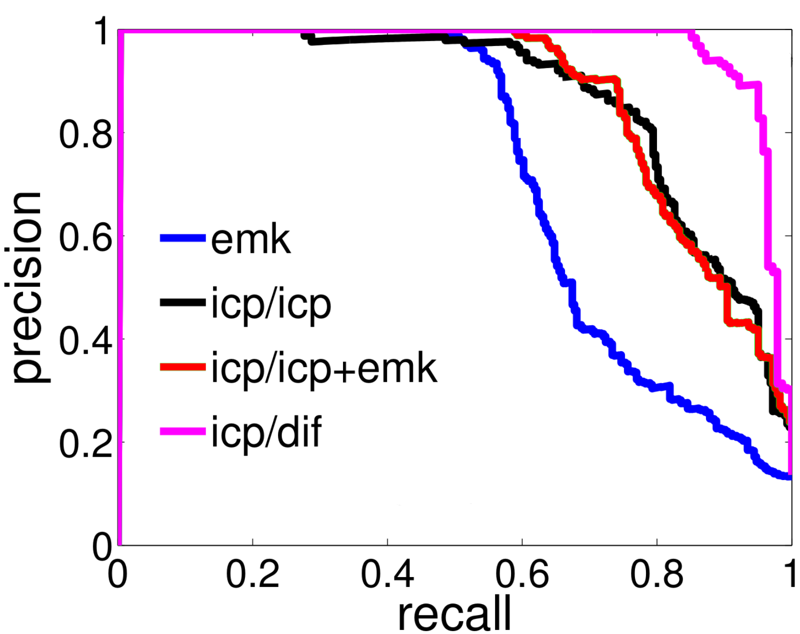


- 3-D matching with ICP
  - Use color to limit correspondences
  - Random restarts; score is lowest Euclidean error achieved
- 3-D matching using beam-based sensor model
  - Advantages: multiple cues; occlusion reasoning
  - Same ICP random restarts
  - Score based on pointwise change probabilities

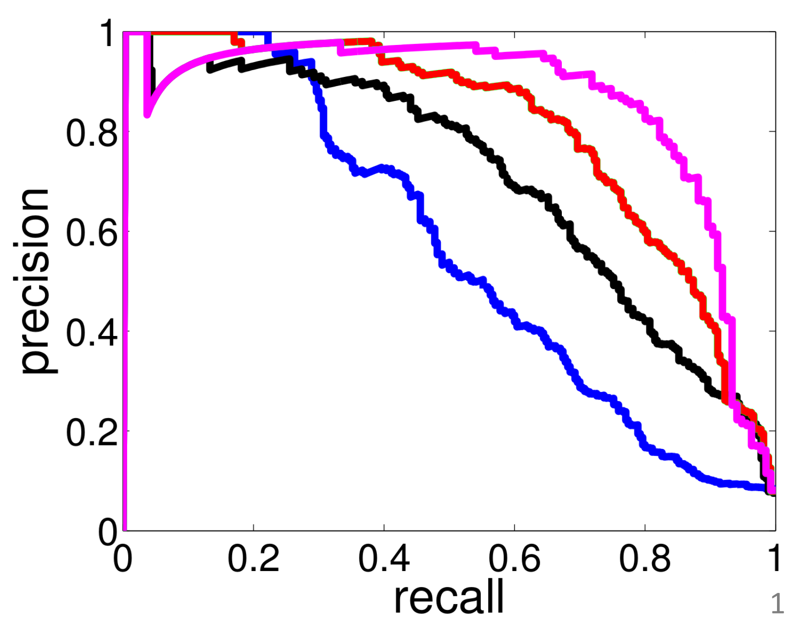
Evaluating Segment Similarity Scores

- Labeled which segments are same object
- Precision/recall wrt same-object pairs
- Sliding threshold to get P/R curve

Dataset A: 48 segments; 8 objects



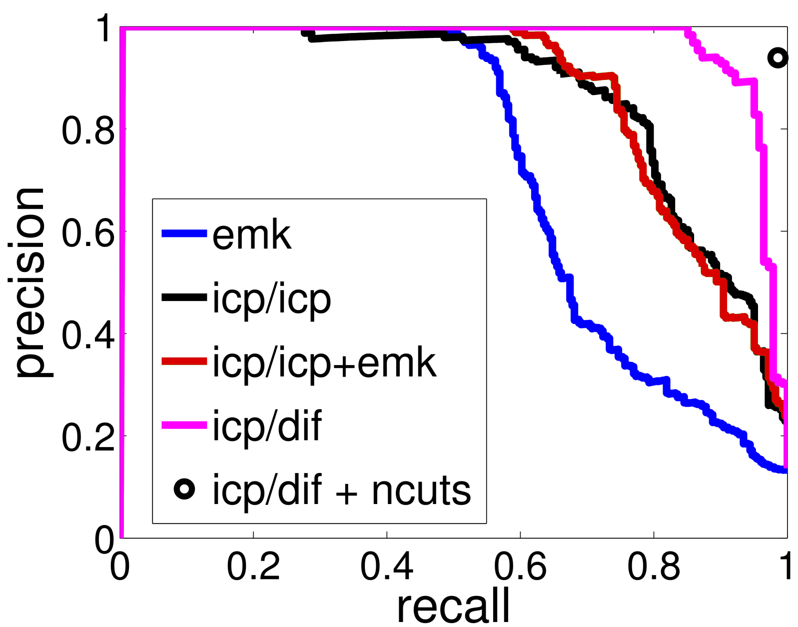
Dataset B: 64 segments; 15 objects



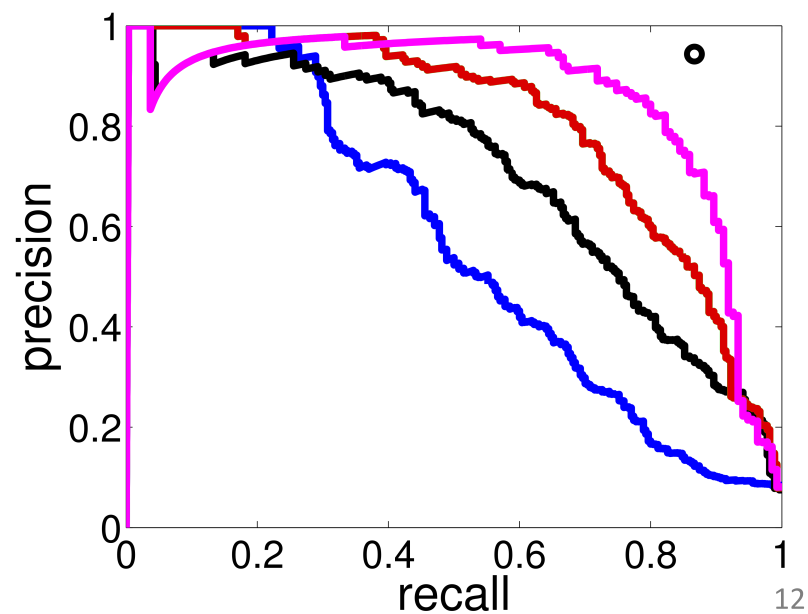
Evaluating Object Discovery

- Labeled which segments are same object
- Precision/recall wrt same-object pairs
- Sliding threshold to get P/R curve

Dataset A: 48 segments; 8 objects



Dataset B: 64 segments; 15 objects



Object Discovery Result

Object Discovery Result

Recap + Future Work

- RGB-D object discovery from many scenes
- Both segmentation and discovery based on probabilistic sensor model
- Segmentation using changes across all scenes
- Next step: active vision to diagnose and fix matching problems in real time



Thank you