Object-aware Guidance for Autonomous Scene Reconstruction

Ligang Liu, **Xi Xia**, Han Sun, Qi Shen,

Juzhan Xu, Bin Chen, Hui Huang, Kai Xu

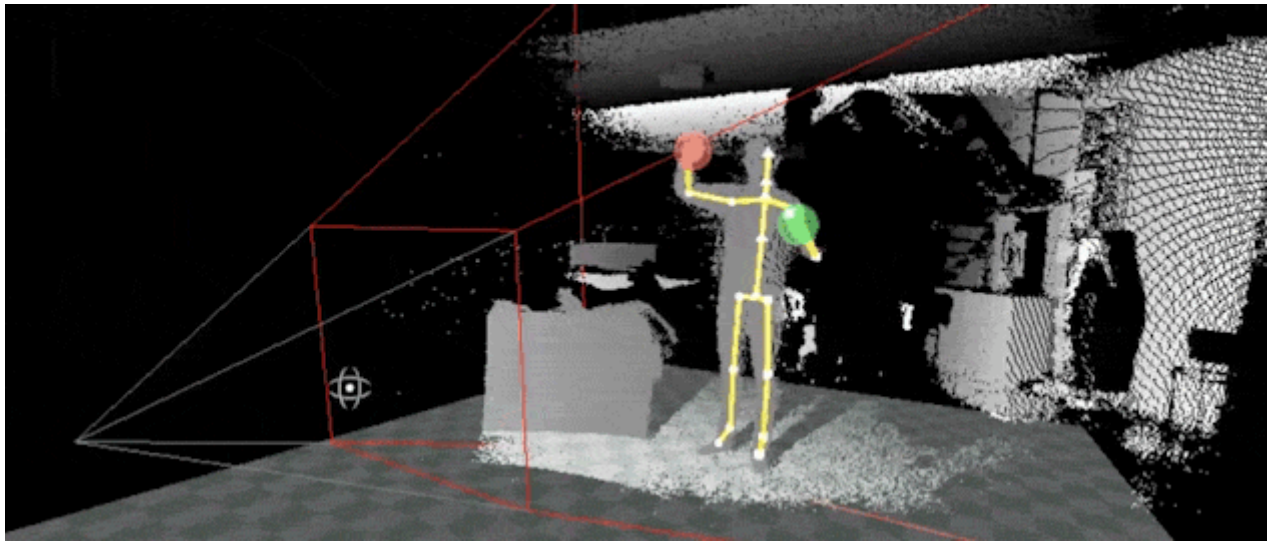
University of Science and Technology of China

Shenzhen University

National University of Defense Technology

Background

- Commodity RGB-D sensors



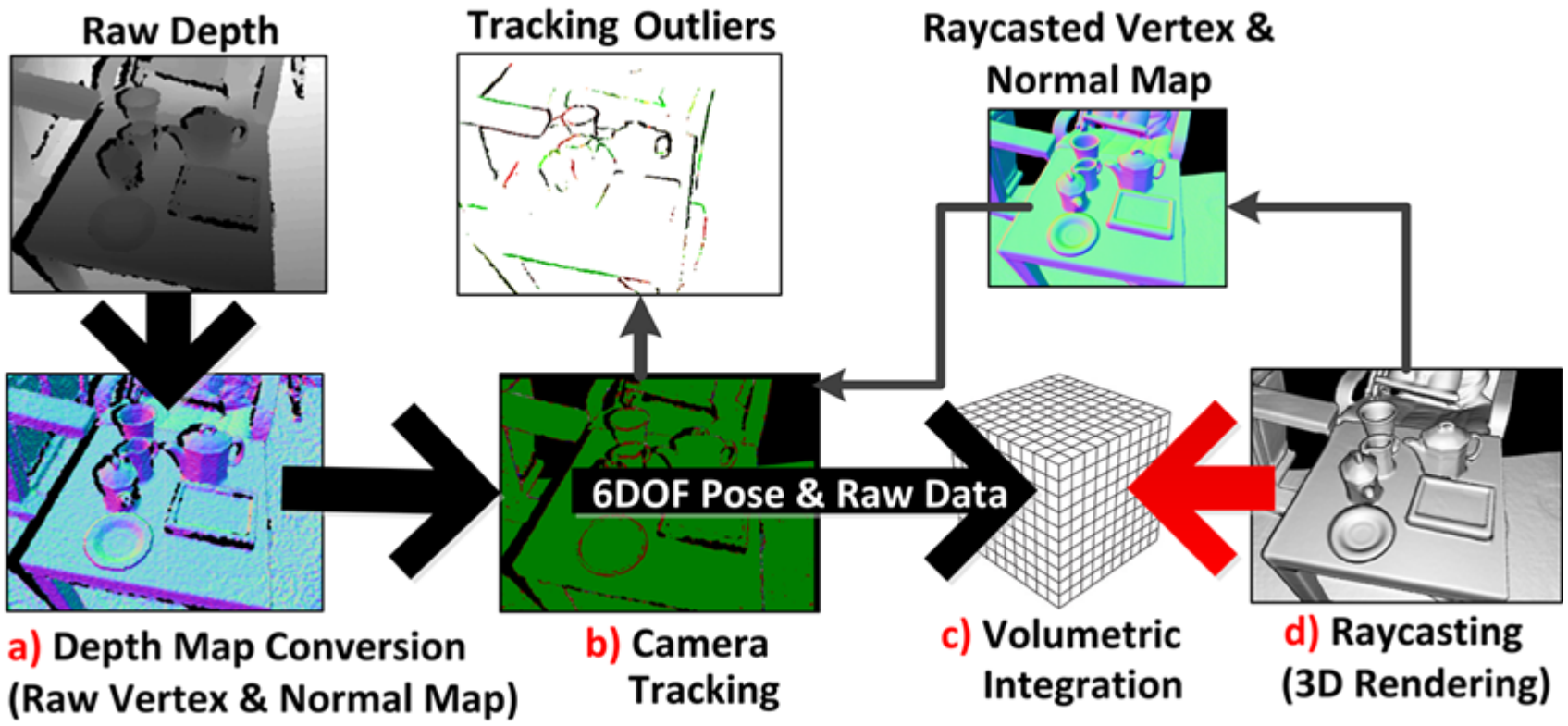
Microsoft Kinect

PrimeSense

Intel RealSense

Background

- RGB-D sensor allows real-time reconstruction

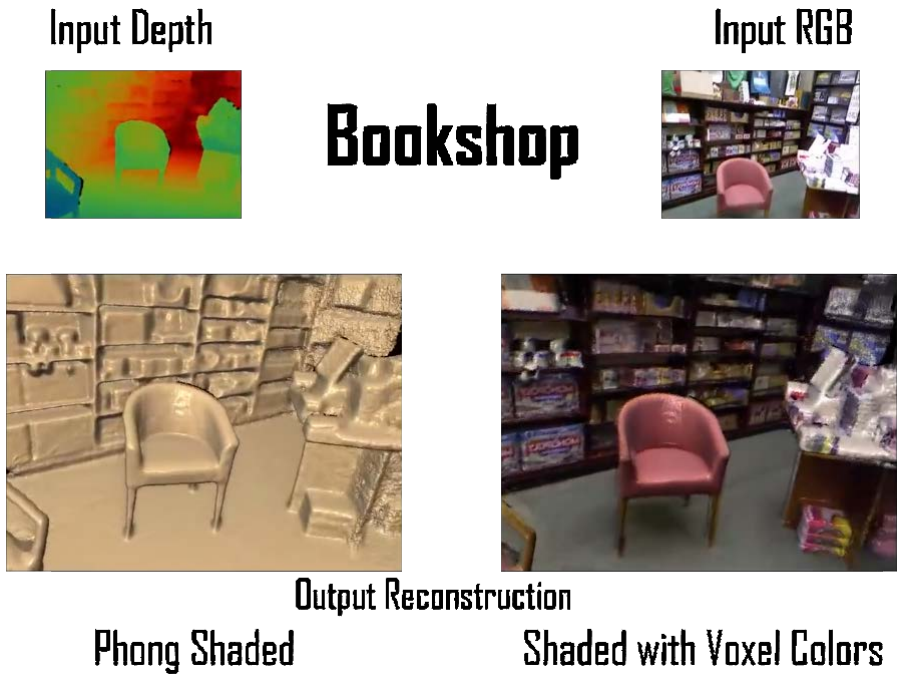


KinectFusion

[Izadi et al. 2011]

Background

- Other real-time reconstruction methods



Voxel Hashing

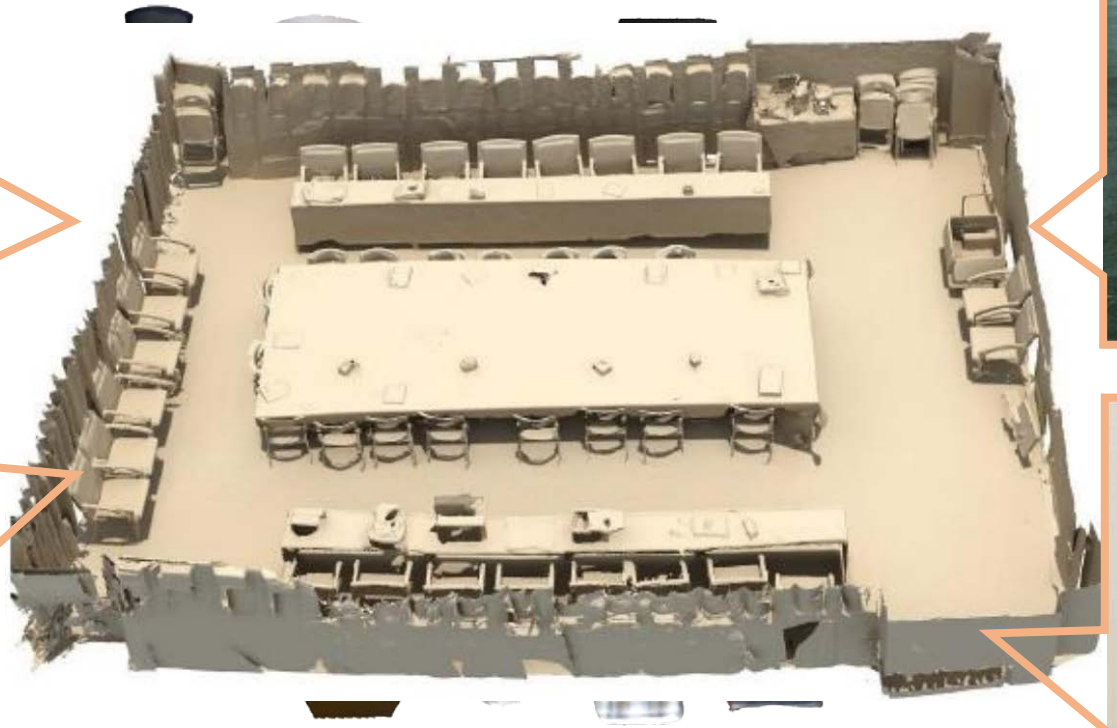
[Nießner et al. 2013]

ElasticFusion

[Whelan et al. 2015]

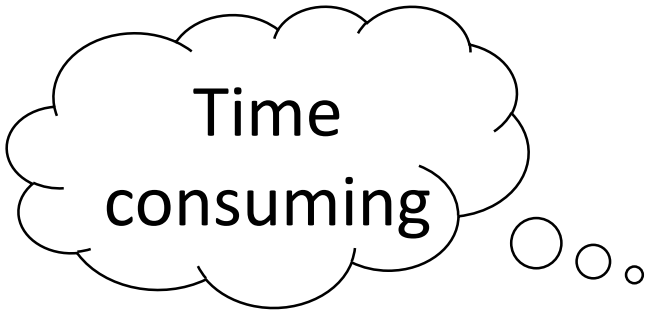
Background

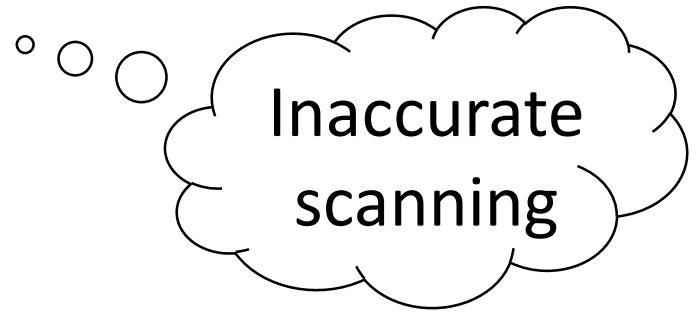
- Indoor scene reconstruction -> **3D object models**



Background

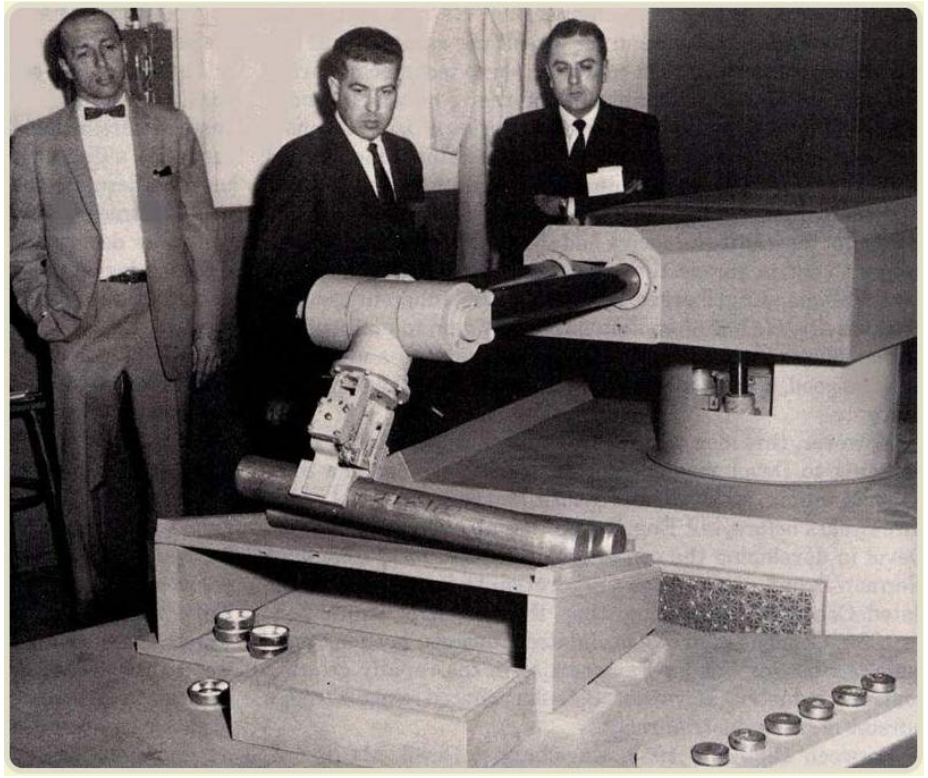
- Human scanning is a laborious task [Kim et al. 2013]





Background

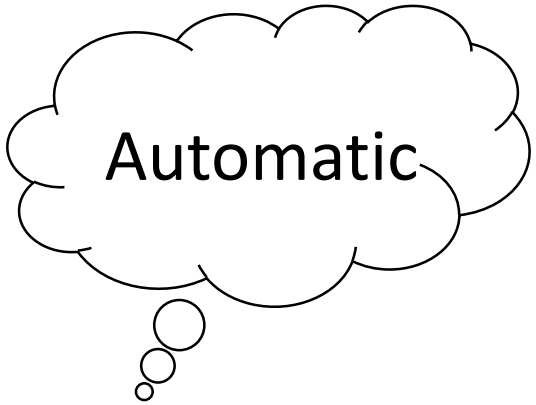
- Modern robots are more and more reliable and controllable.

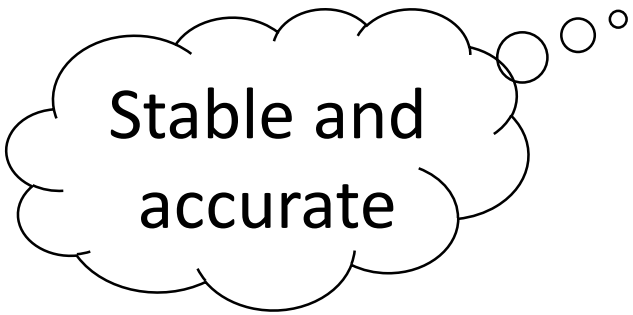


Unimation, 1958

Fetch, 2015

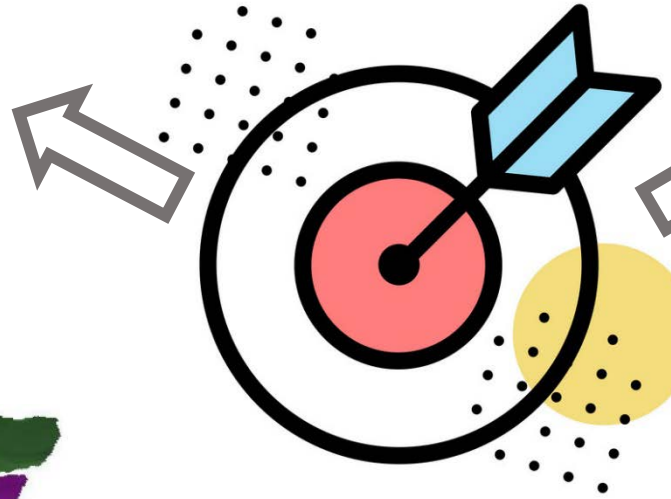
Motivation



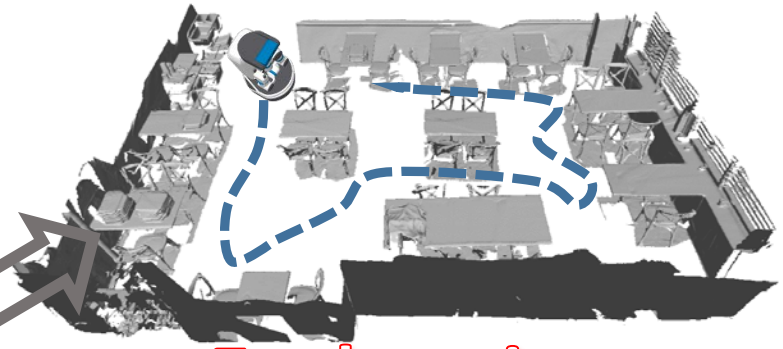


Goal

Reconstruction

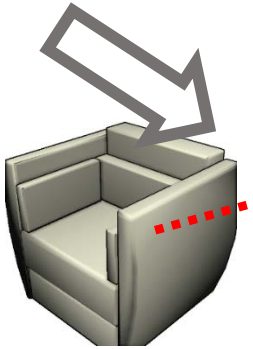


GOAL



Exploration

Segmentation



Recognition

Existing Works

- High quality scanning and reconstruction of single object [Wu et al. 2014]

Existing Works

- Global path planning and exploration [Xu et al. 2017]

Existing Works

- Active reconstruction and segmentation [Xu et al. 2015]

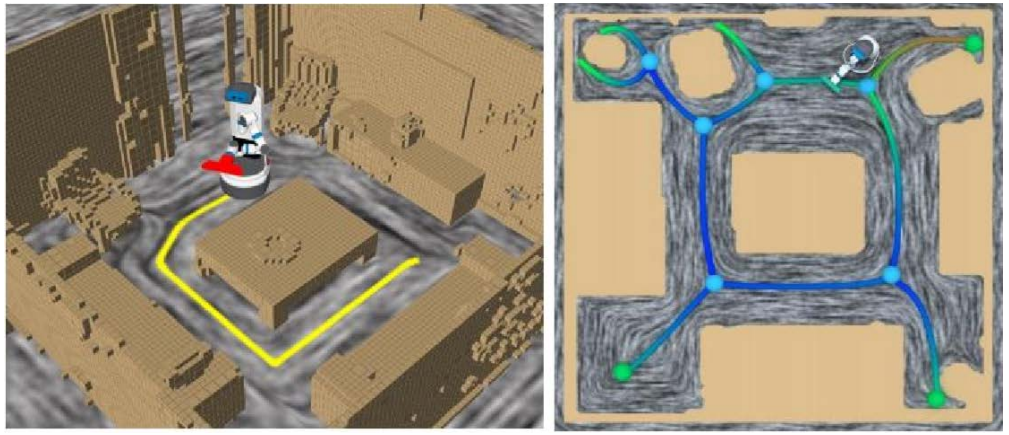
Existing Works

- Local view planning for recognition [Xu et al. 2016]

Conclusion of Existing Works

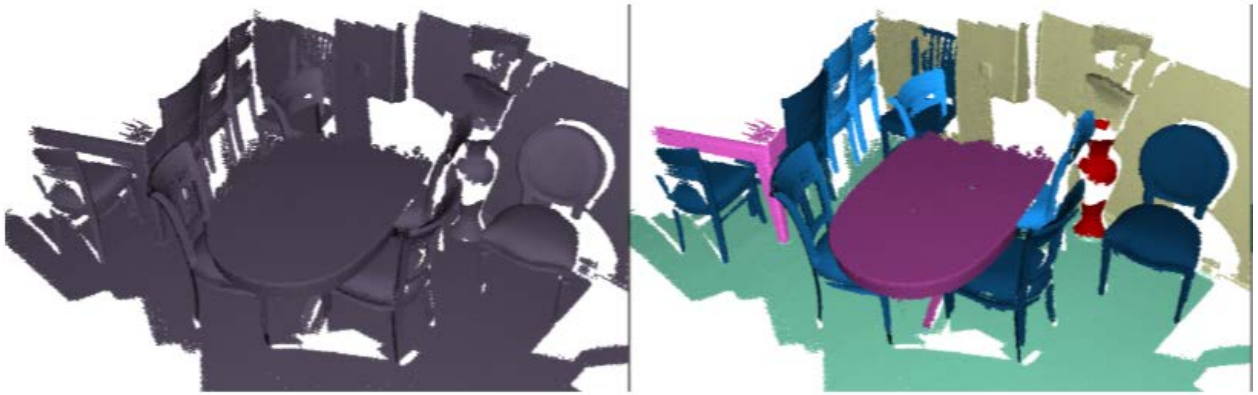
- Two pass scene reconstruction and understanding.
- Can only use **low-level** information in first exploration pass.

First Pass



exploration & reconstruction
[Xu et al. 2017]

Second Pass



segmentation & recognition
[Nan et al. 2012]

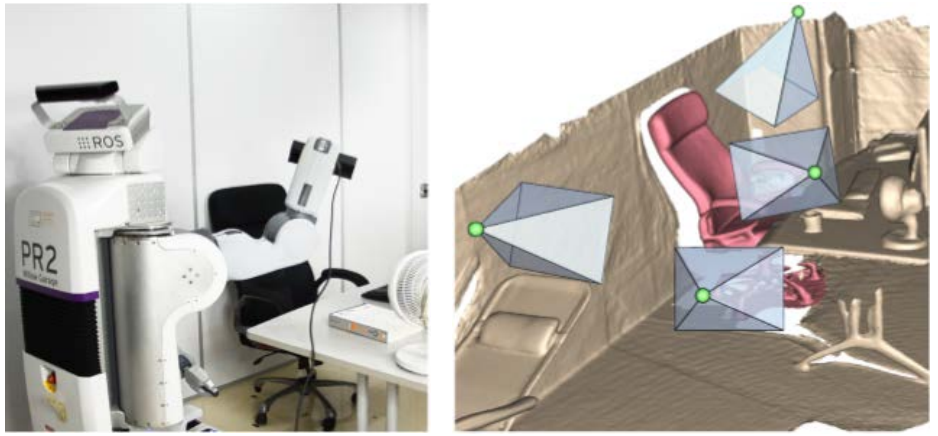
Conclusion of Existing Works

- Two pass scene reconstruction and understanding.
- Can only use **low-level** information in first exploration pass.

First Pass

reconstruction & segmentation
[Xu et al. 2015]

Second Pass



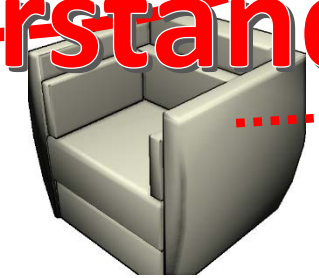
object recognition
[Xu et al. 2016]

The Main Challenge

One navigation

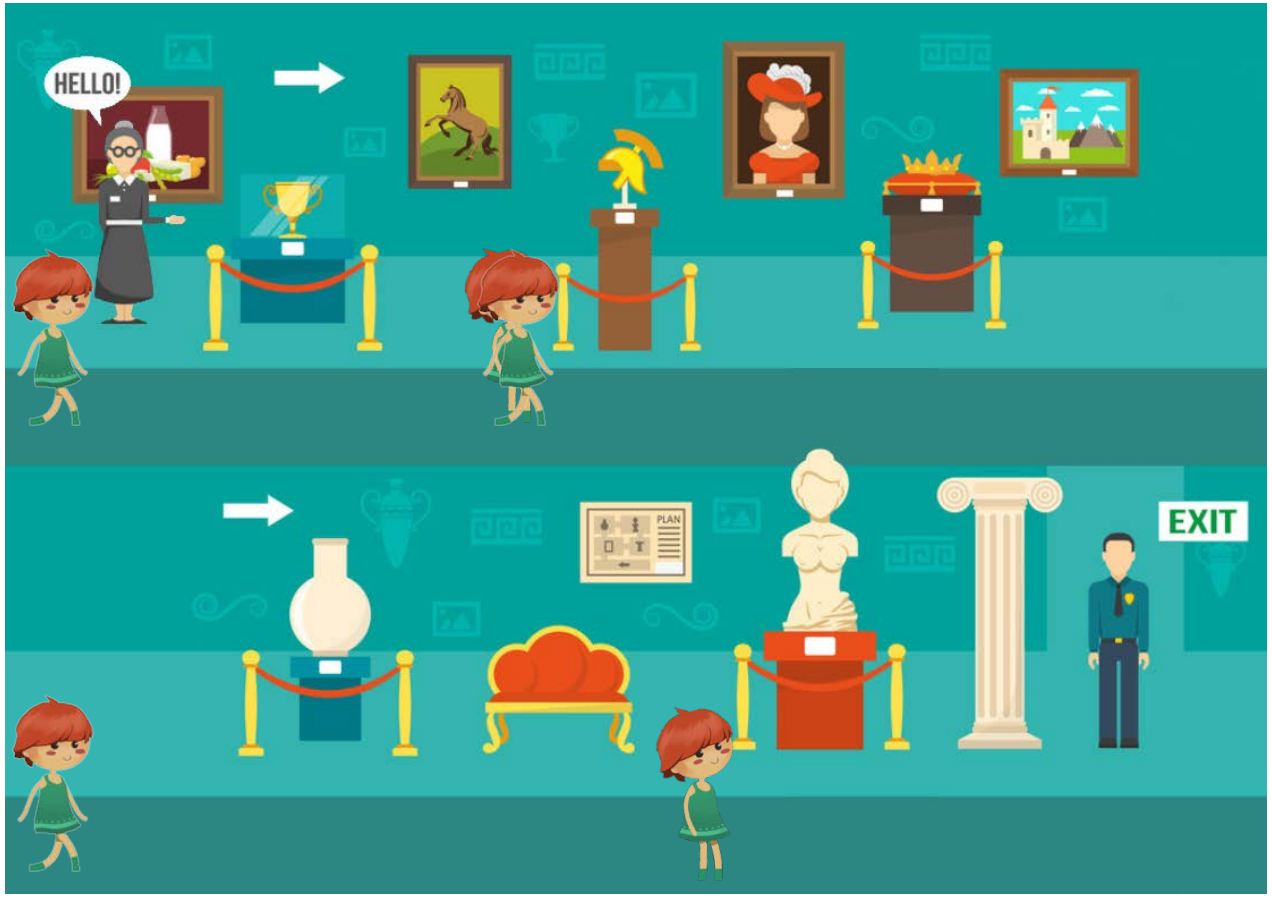
Automatic

Scene Understanding



Motivation

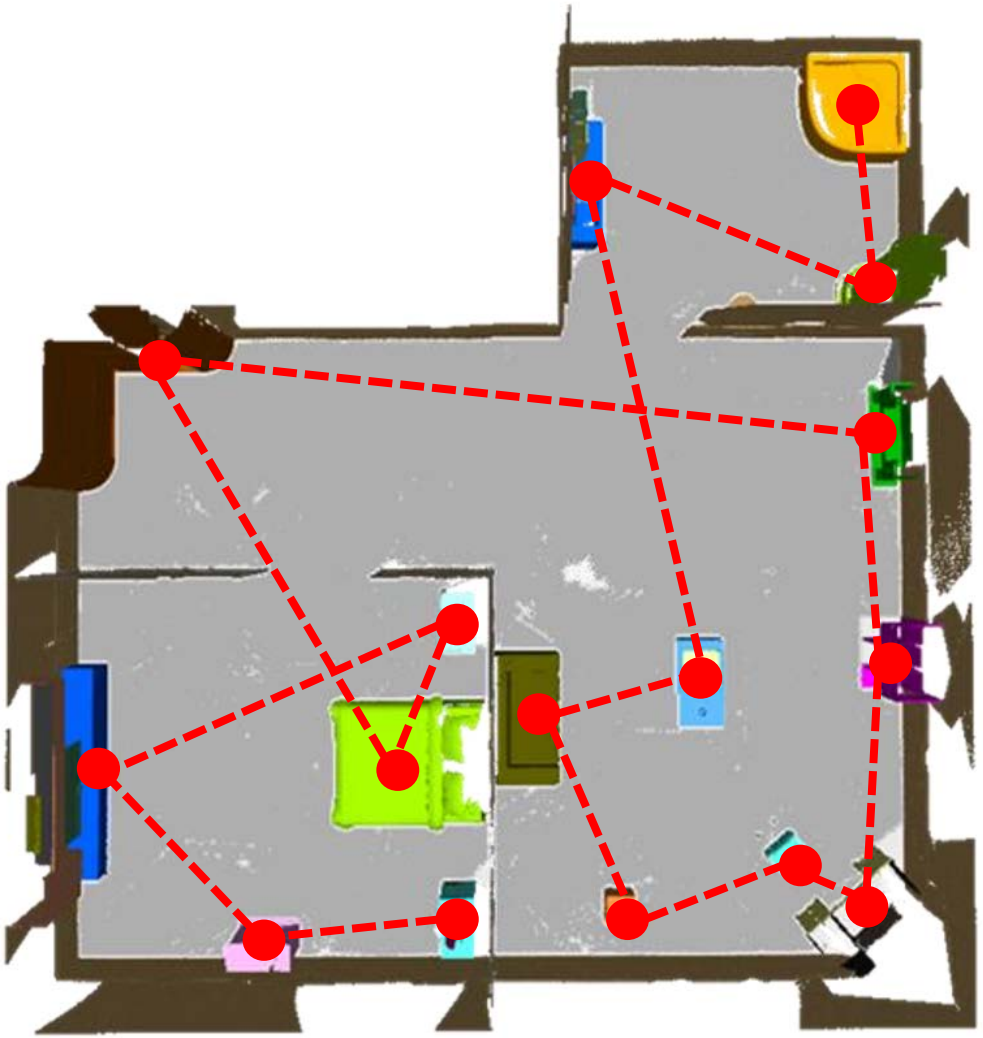
- Human explore unknown scenes **object by object**!



Motivation

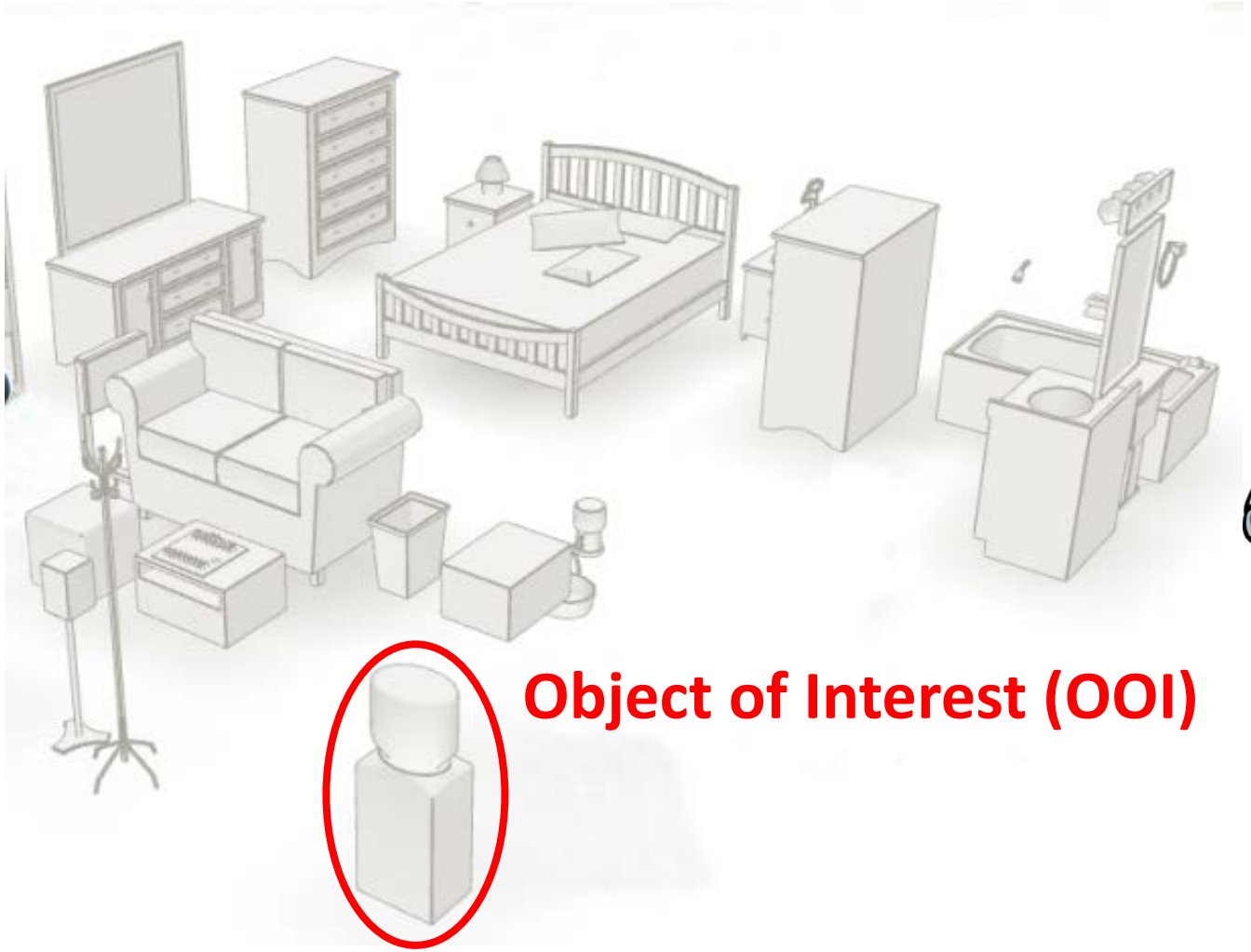
- Human tend to scan **object by object**!

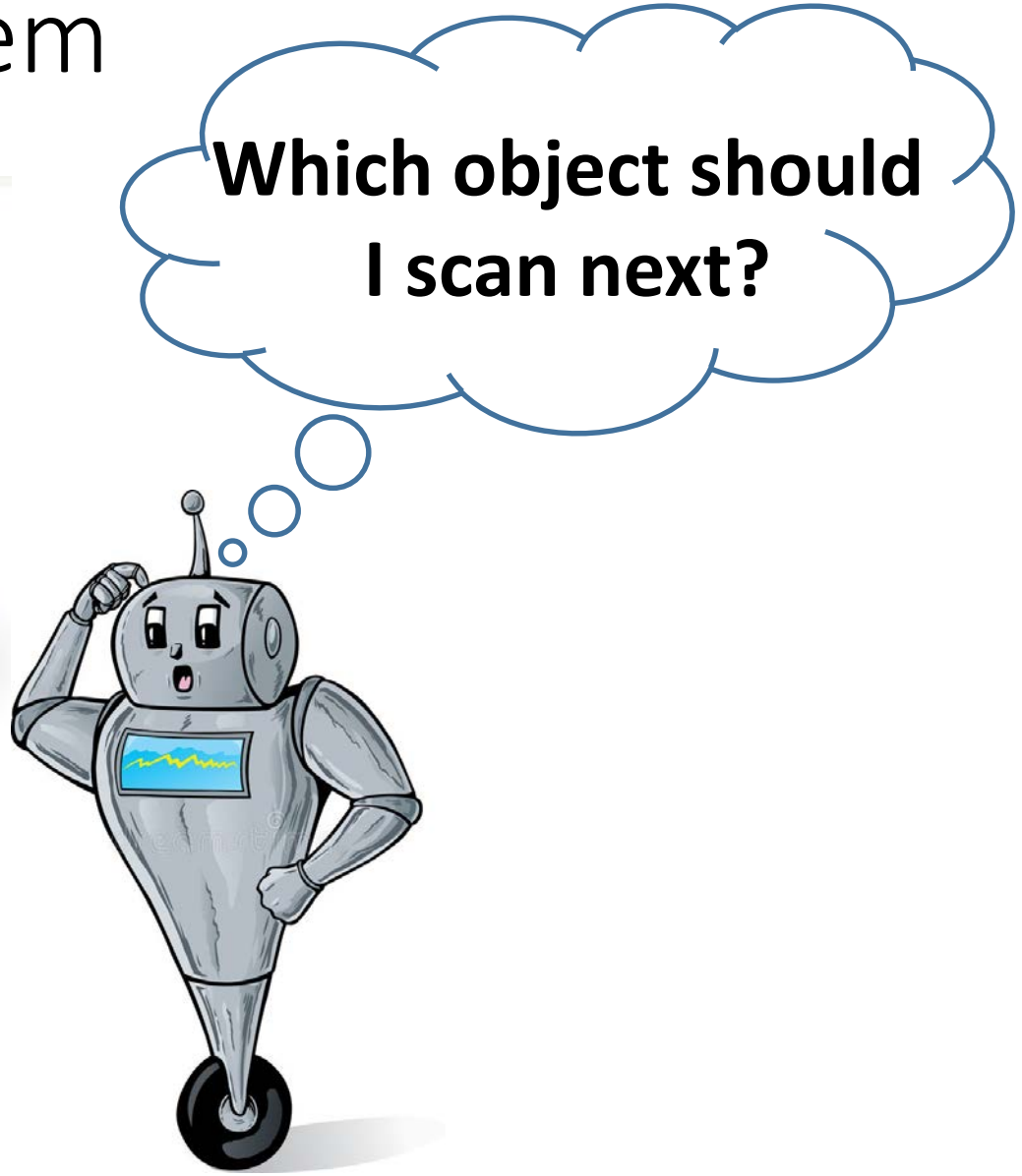
Our Solution



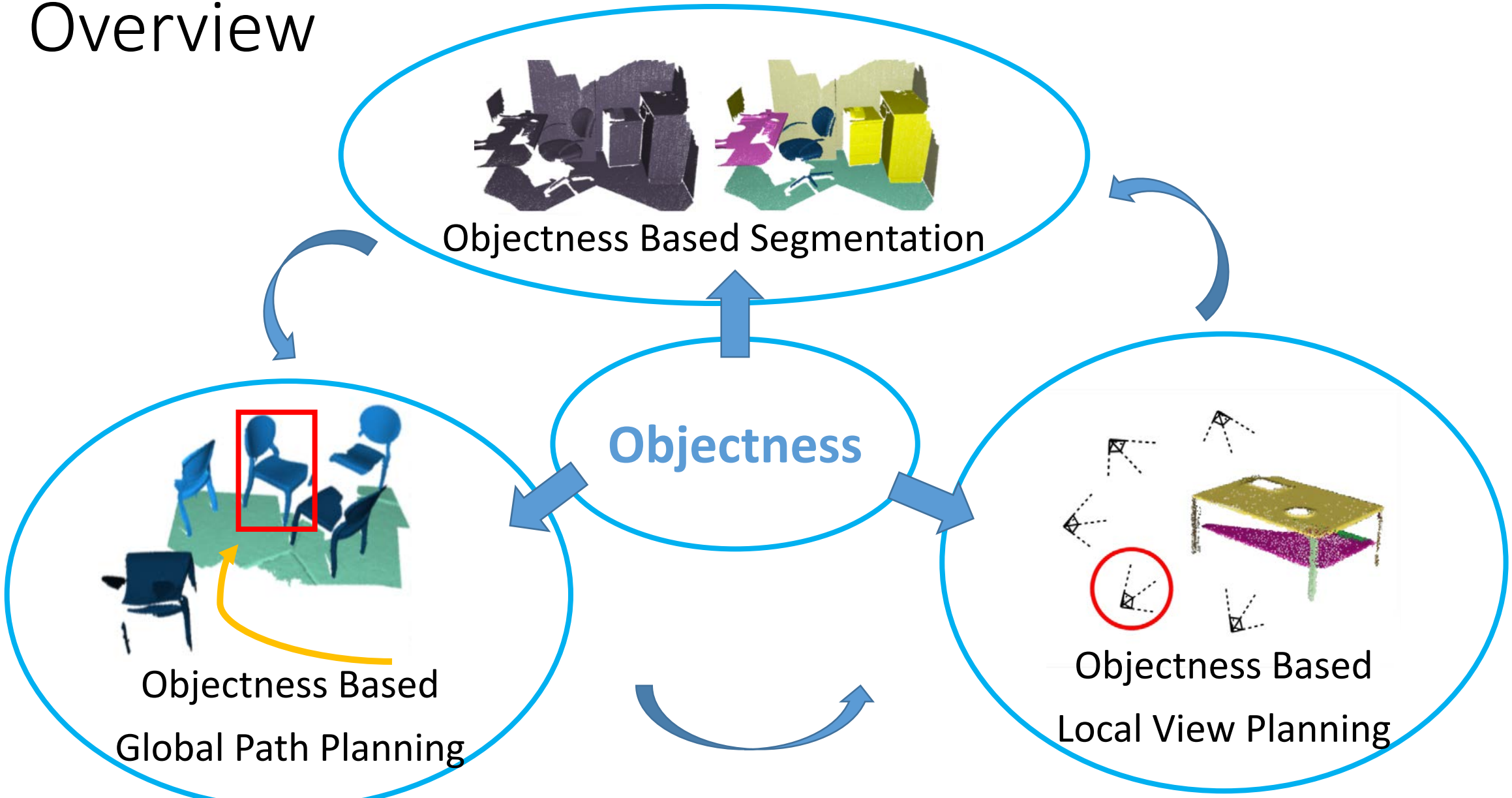
- **Key idea**: Online recognized **objects** serve as an important **guidance** map for planning the robot scanning.

The Next Best Object Problem



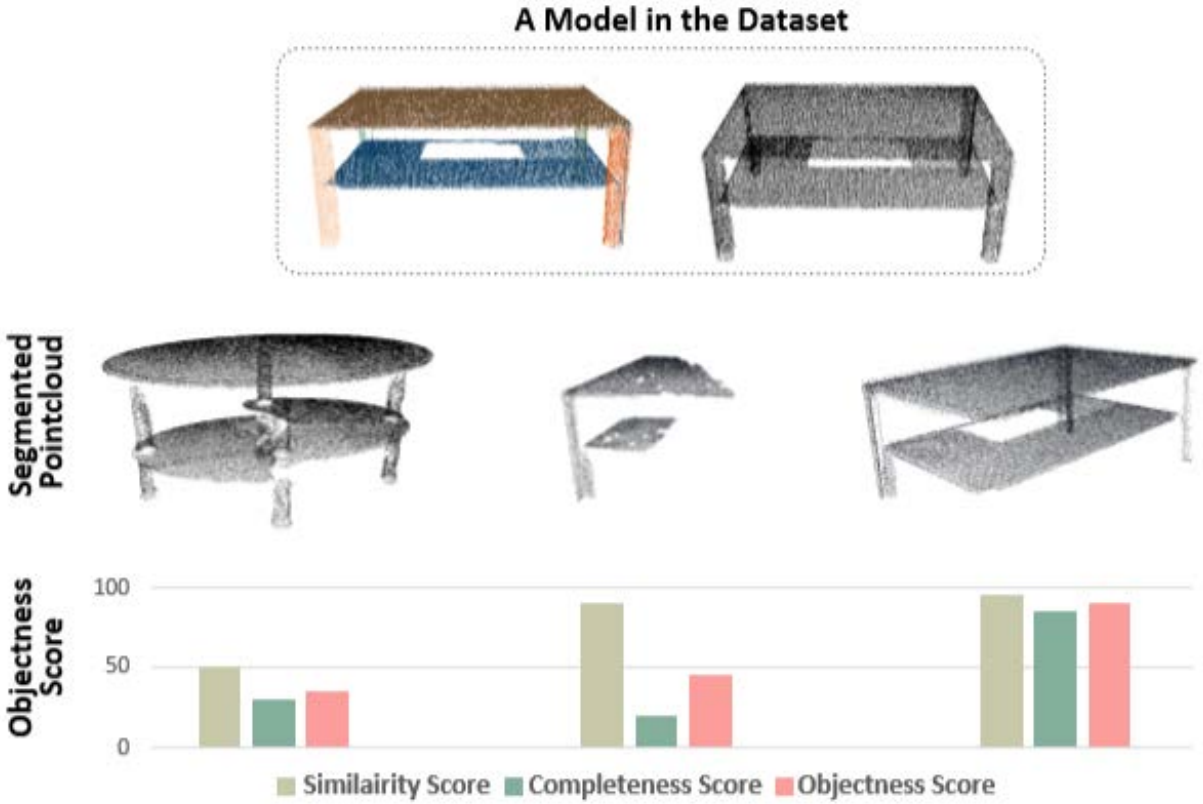


Overview

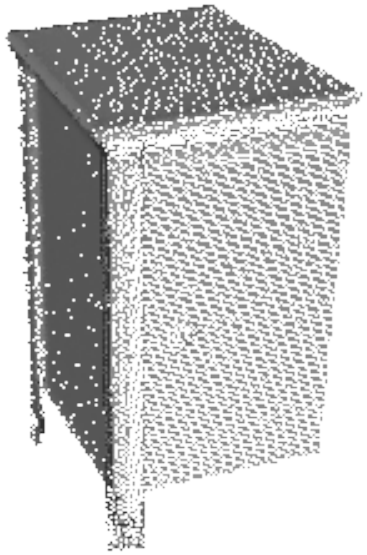

Objectness Based Segmentation
Objectness
Objectness Based
Global Path Planning
Objectness Based
Local View Planning

Model-Driven Objectness

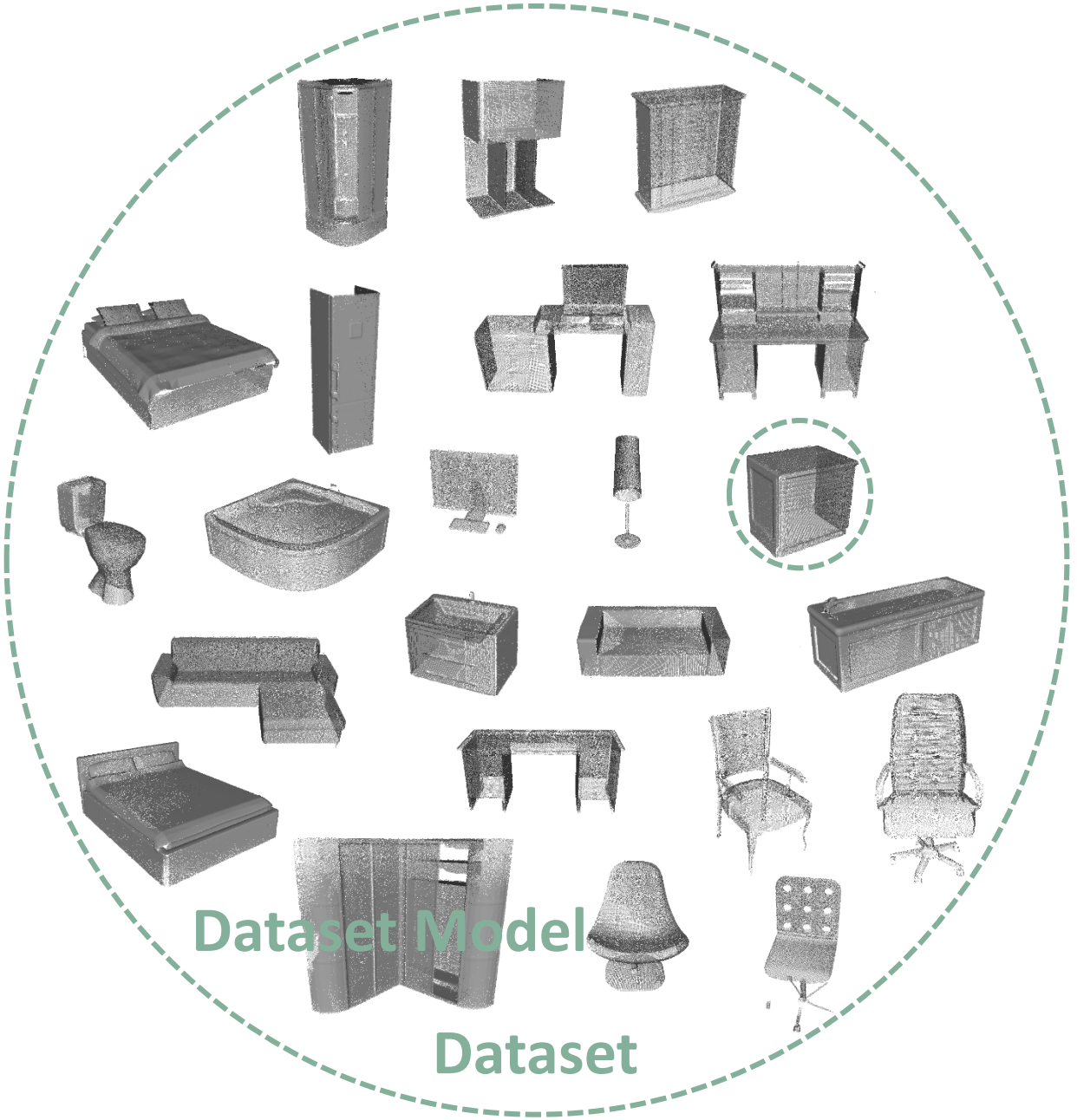
- Objectness should measure both similarity and completeness



Partial Matching



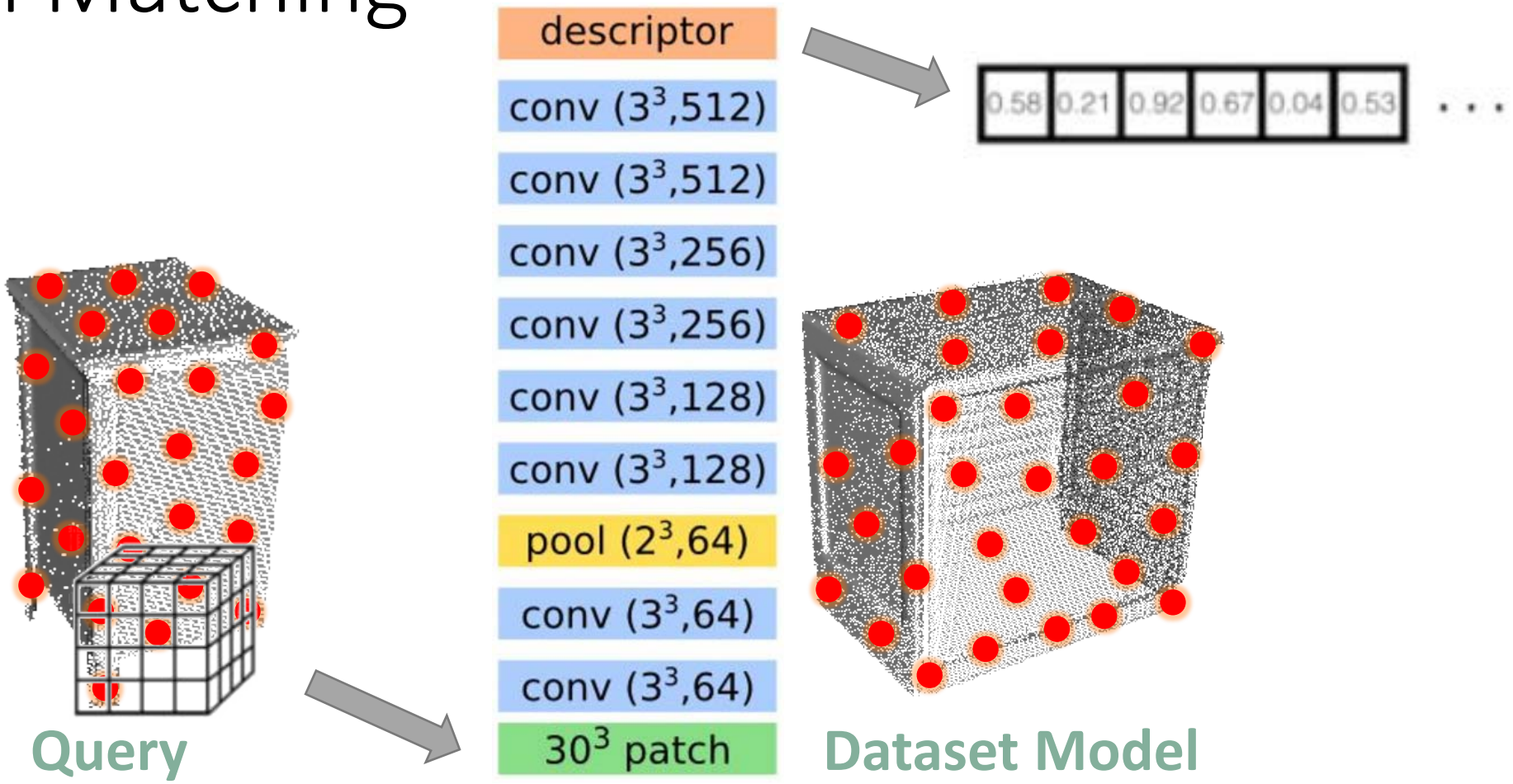
Query



Dataset Model

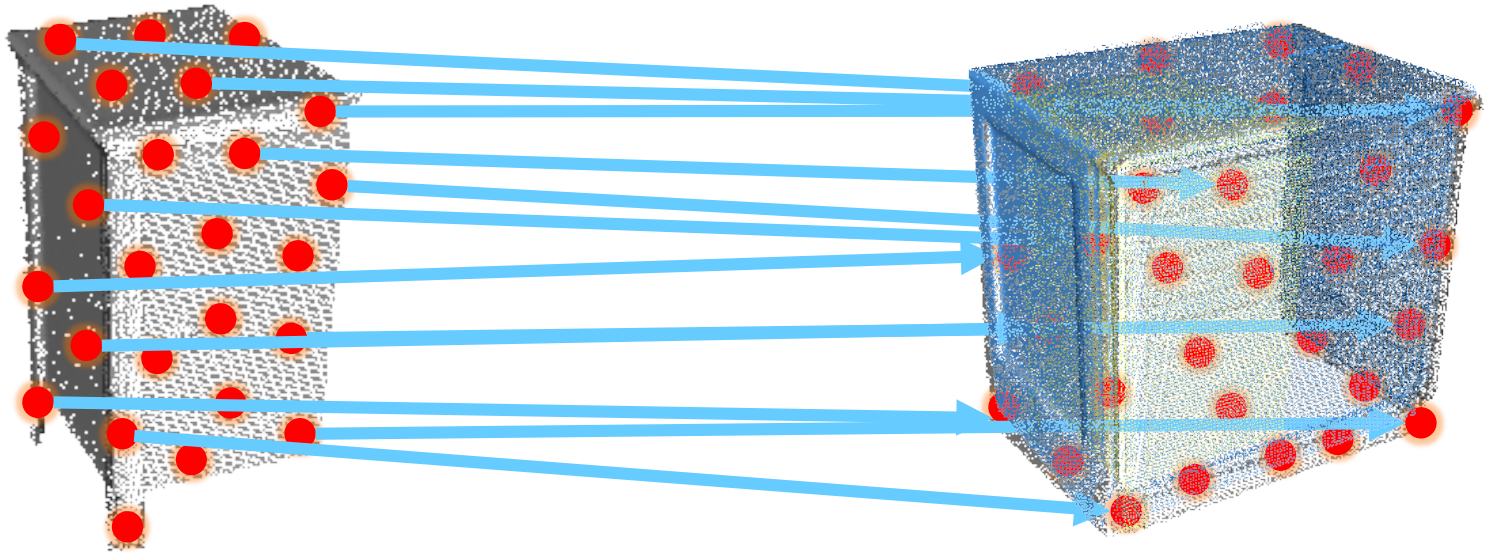
Dataset

Partial Matching



3DMatch [Zeng et al. 2016]

Partial Matching



Query

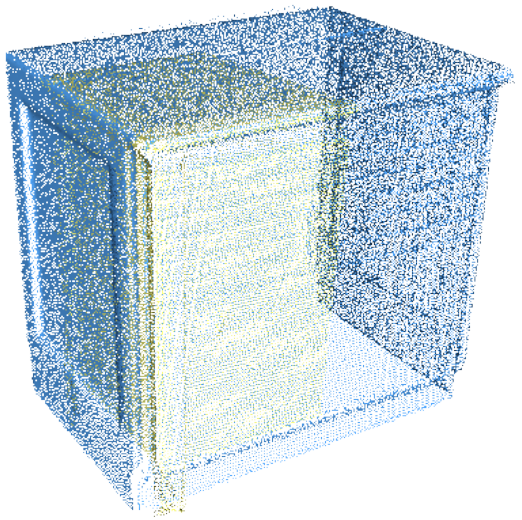
Dataset Model

Model-Driven Objectness

$$d(X, Y) = \frac{1}{n_p} \sum_{i=1}^{n_p} d(x_i, Y)$$

$$d(x_i, Y) = \min_{j=1,\cdots,n_p} \|x_i - y_j\|^2$$

$$O(c, m) = \exp\left[-\frac{1}{Diag(c)}(d(c, m) + d(m, c))^{\frac{1}{2}}\right]$$



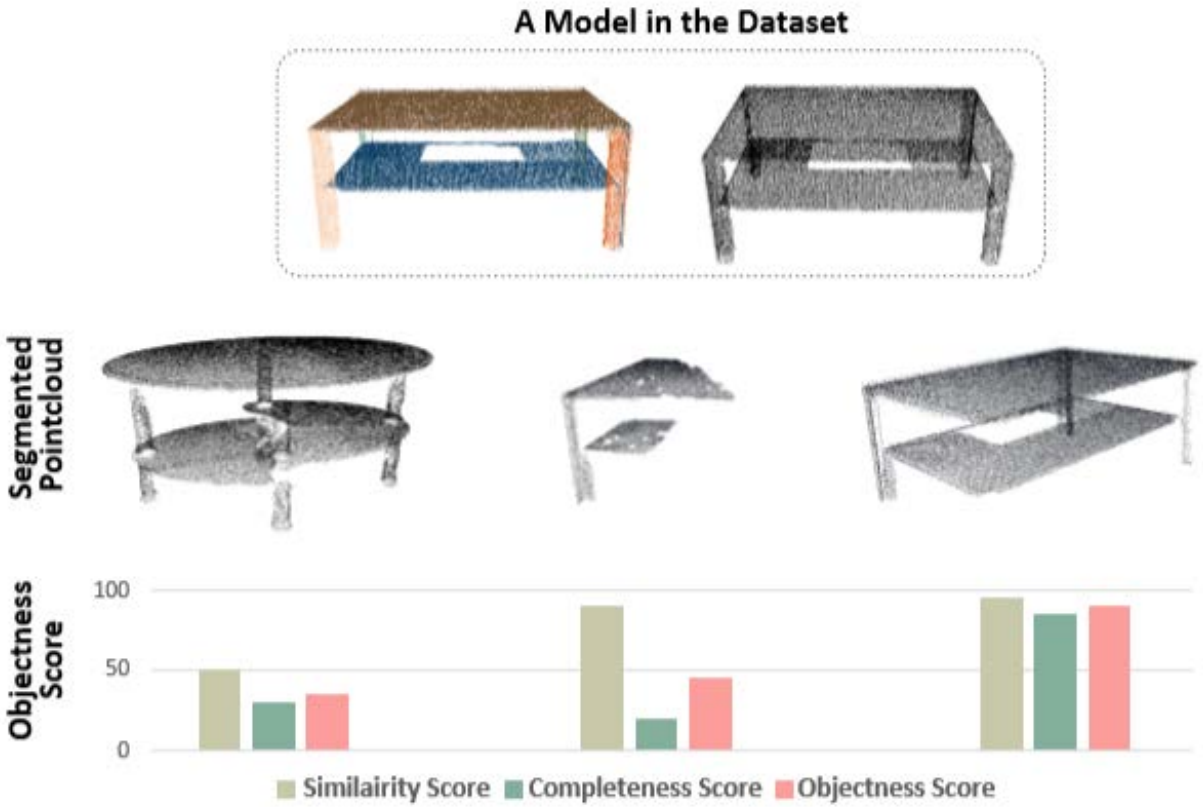
Model-Driven Objectness

- Objectness should measure both similarity and completeness

$$d(X, Y) = \frac{1}{n_p} \sum_{i=1}^{n_p} d(x_i, Y)$$

$$d(x_i, Y) = \min_{j=1,\cdots,n_p} \|x_i - y_j\|^2$$

$$O(c, m) = \exp\left[-\frac{1}{Diag(c)}(d(c, m) + d(m, c))^{\frac{1}{2}}\right]$$



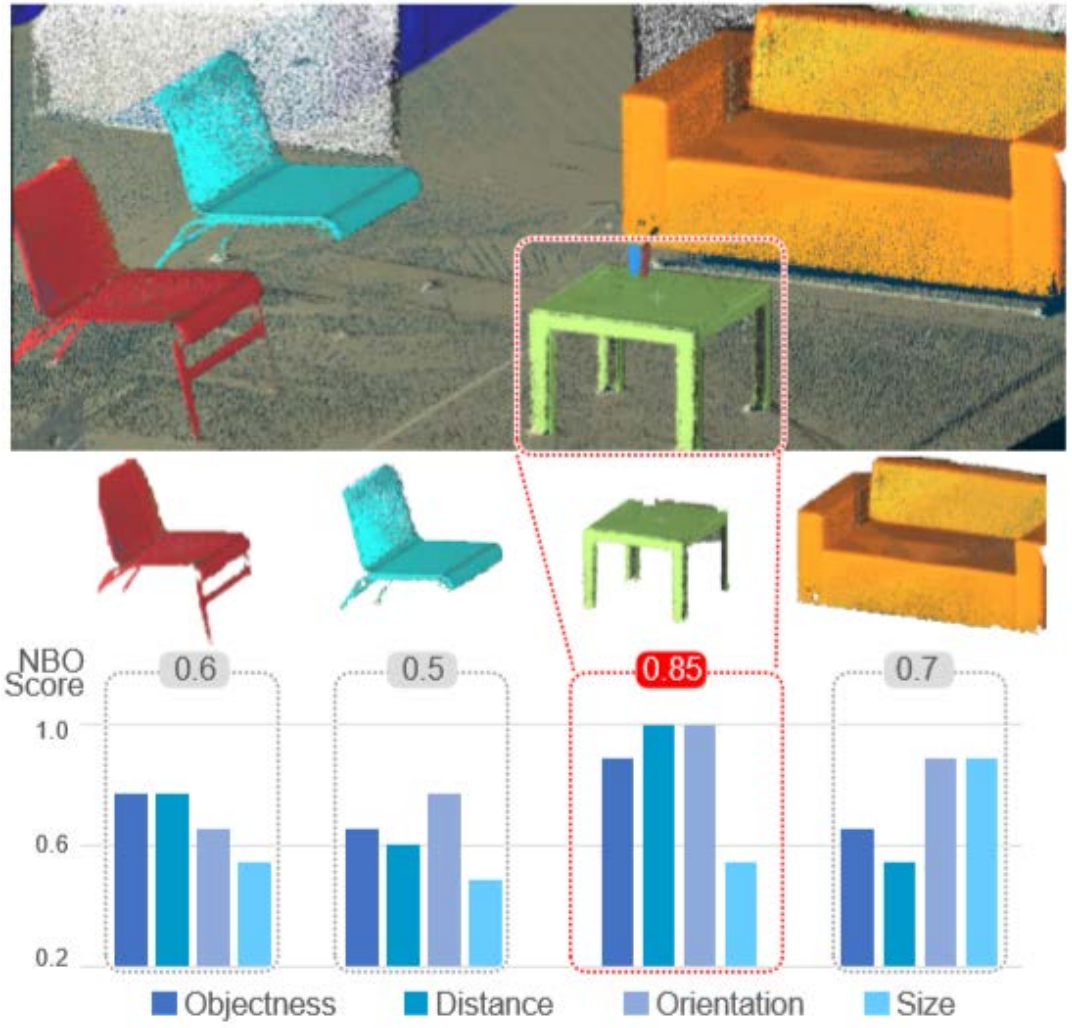
Next Best Object

Objectness

$$\gamma = \arg\max_{r \in \mathcal{R}} O(r) + S(r)$$

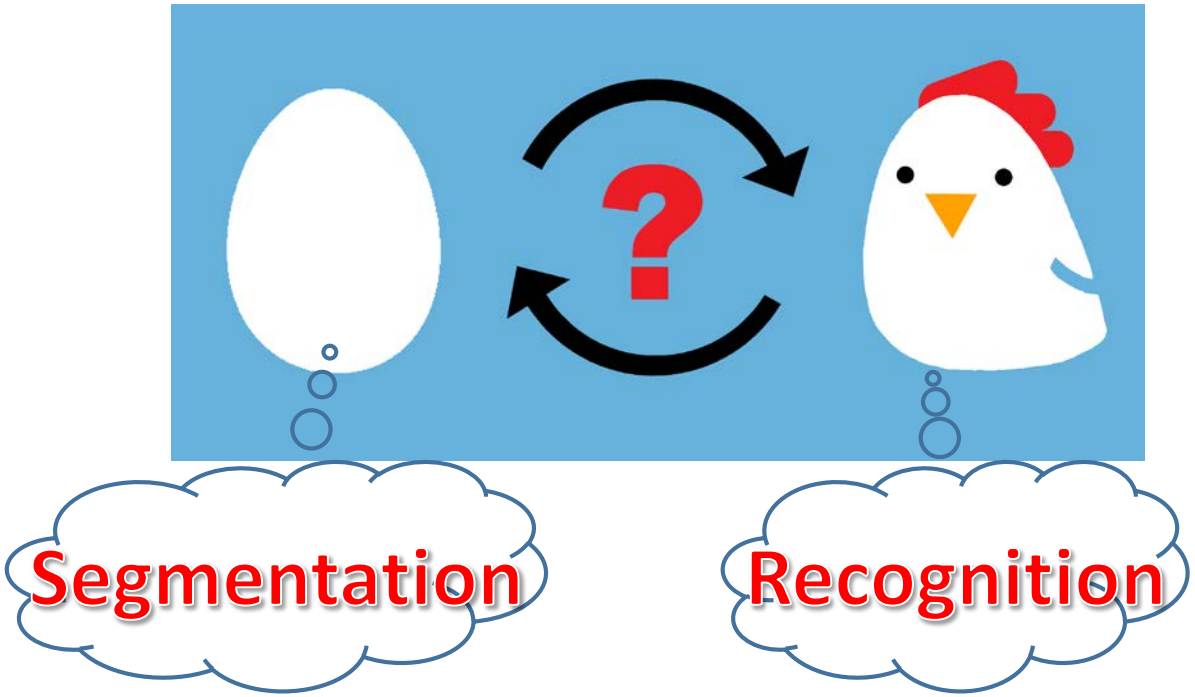
$$S(r) = w_z S_z(r) + w_e S_e(r) + w_d S_d(r)$$

Distance **Orientation** **Size**



Technical Challenge

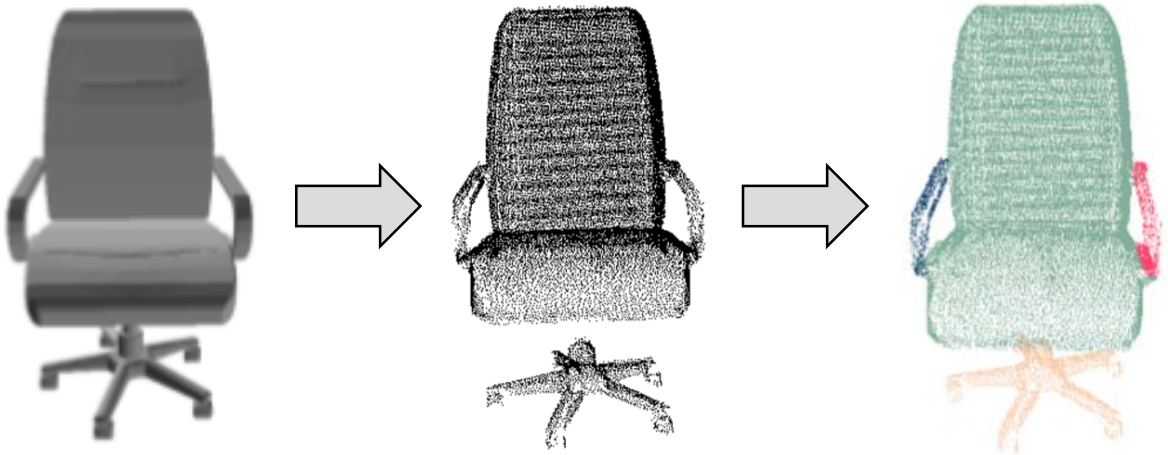
- How to segment and recognize objects during reconstruction?



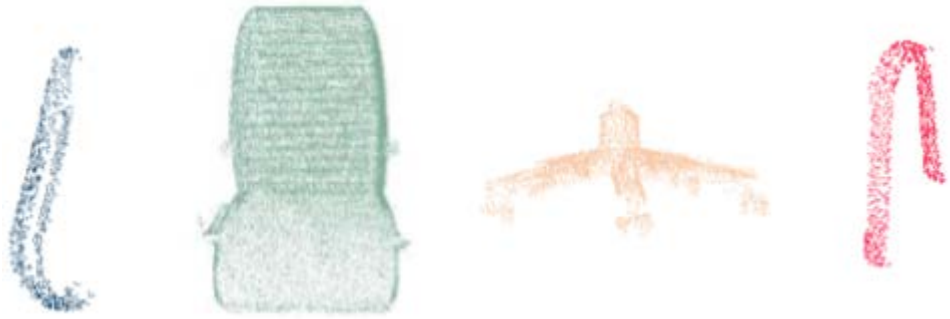
Recognition and segmentation constitute a ***chicken-egg*** problem

Pre-segmentation

[Whelan et al. 2015] [Tateno et al. 2015]



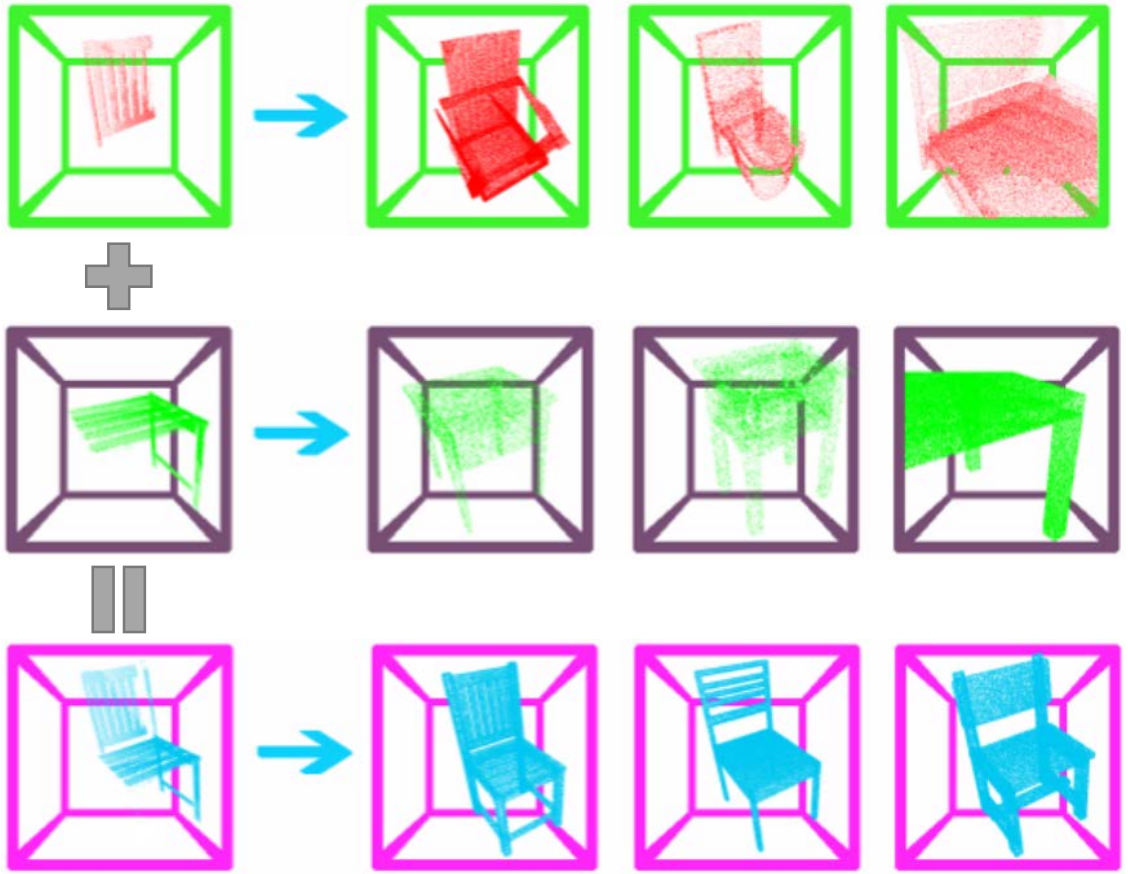
Indoor object Scanned Model



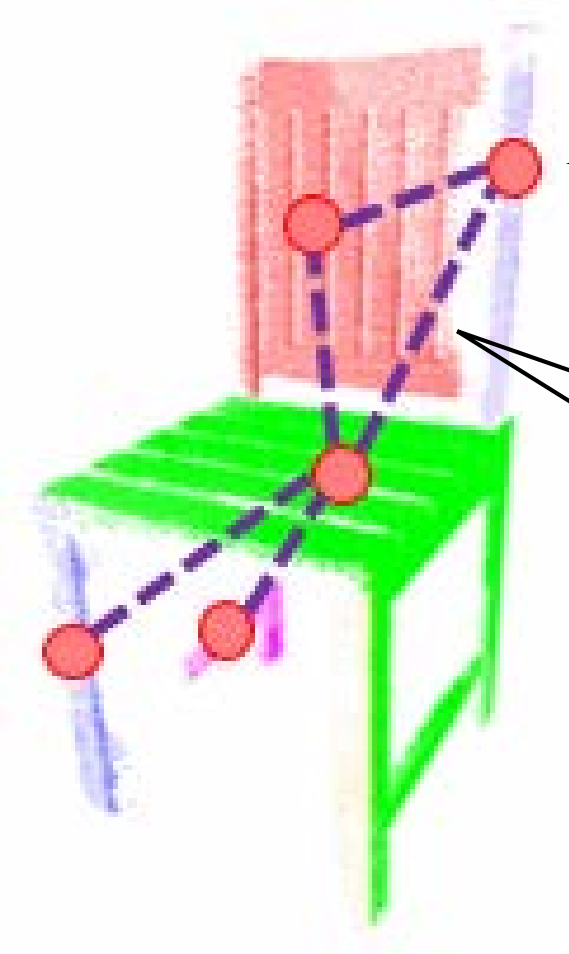
Pre-segmented Components

Post-segmentation

- Couples segmentation and recognition in the same optimization



Post-segmentation



$$E_D(l_c) = \min_{m \in M(c), l_c = L(m)} (1 - O(c, m))$$

$$+$$

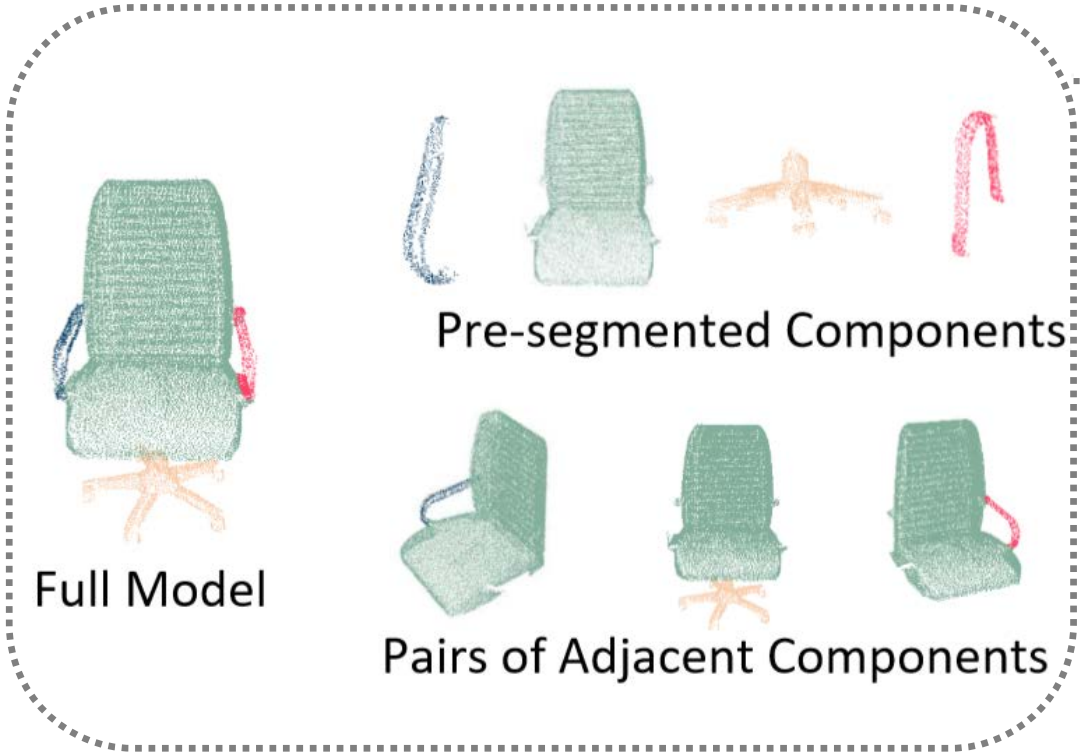
$$E_S(l_c, l_d) = \begin{cases} \max_{m \in M(c \cup d)} O(c \cup d, m), & \text{if } l_c \neq l_d \\ 0, & \text{if } l_c = l_d \end{cases}$$

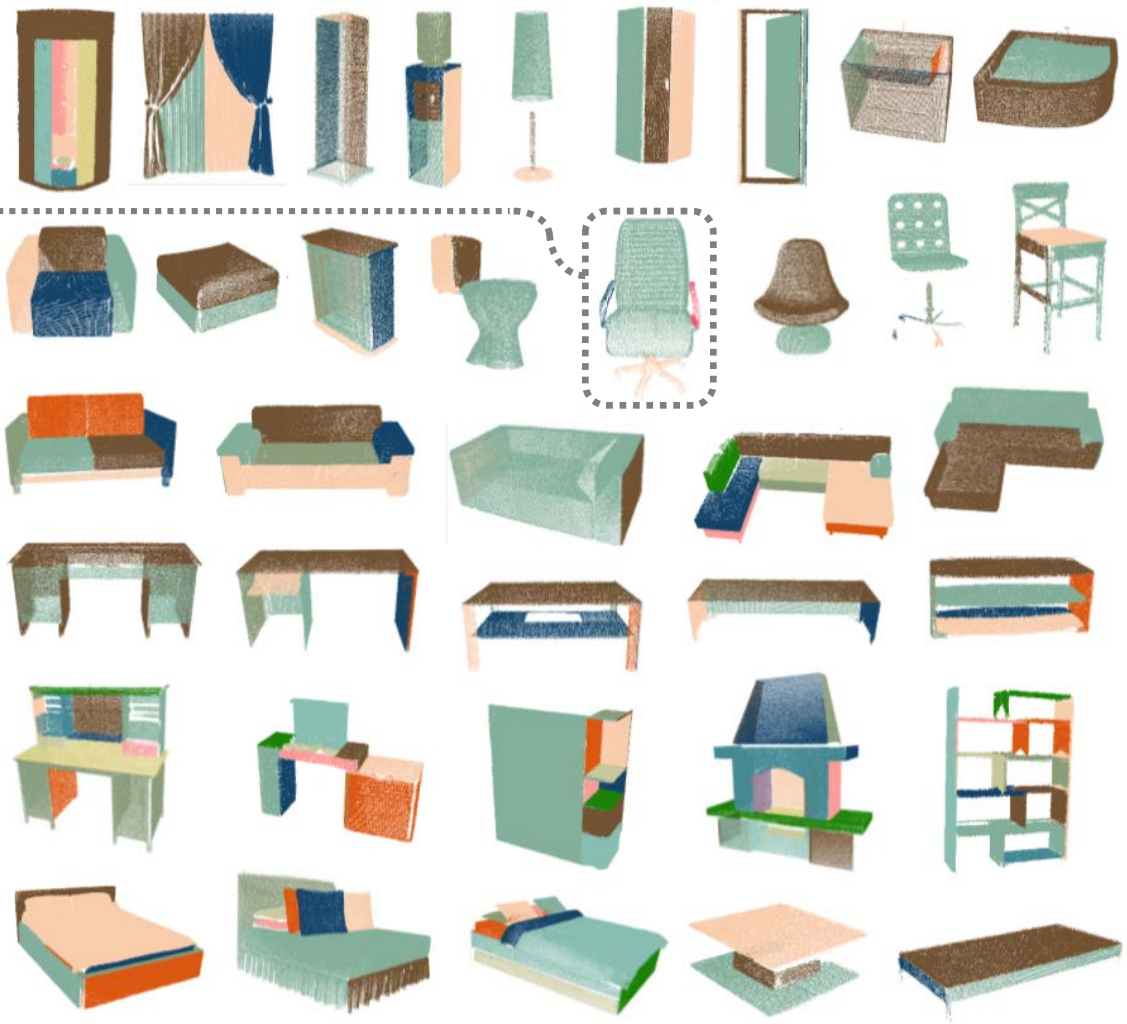
$$=$$

$$\min_{L=\{l_c\}} E(L) = \sum_{c \in \mathcal{V}_c} E_D(l_c) + \sum_{(c,d) \in \mathcal{E}_c} E_S(l_c, l_d)$$

Post-segmentation Results

Dataset Construction



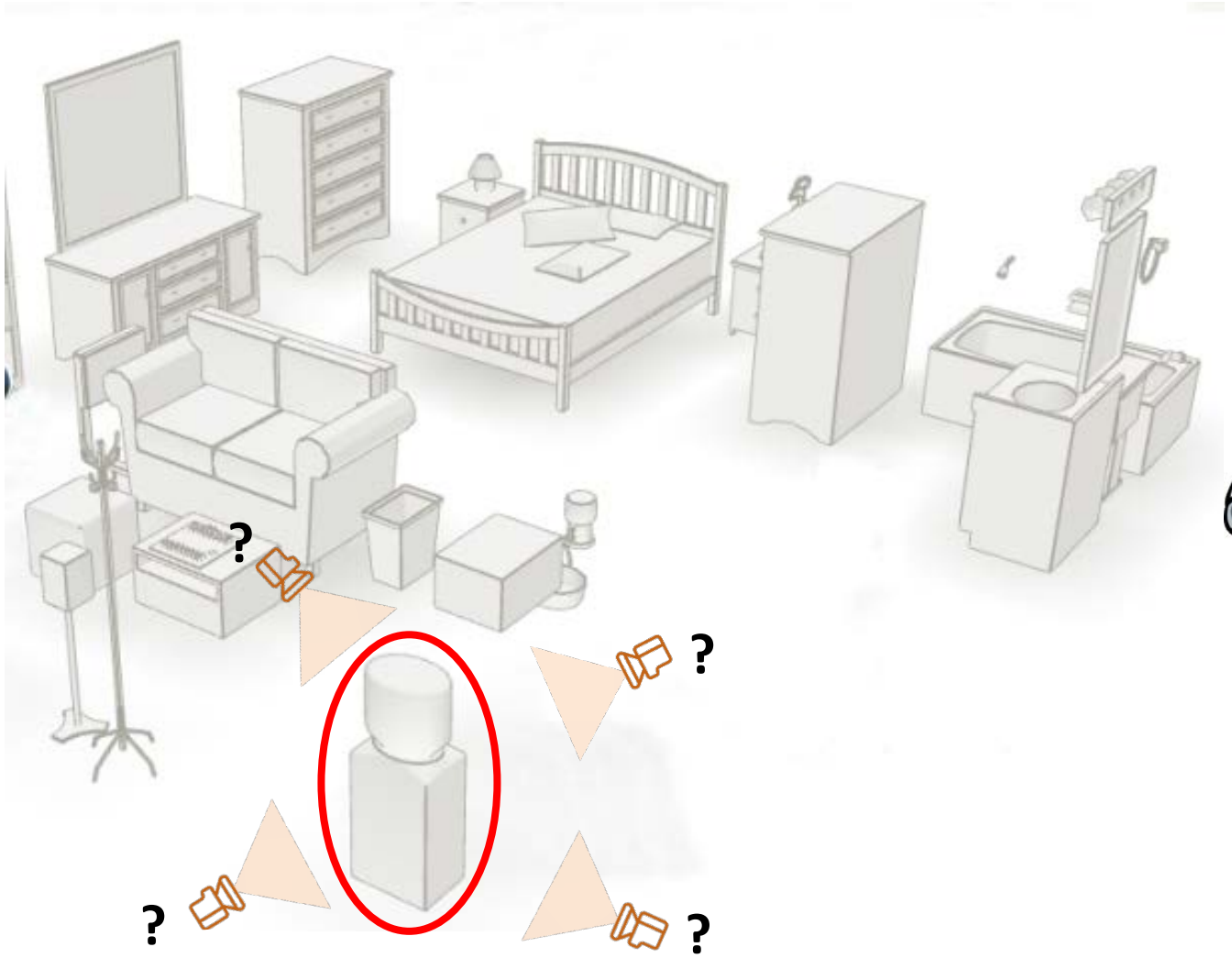


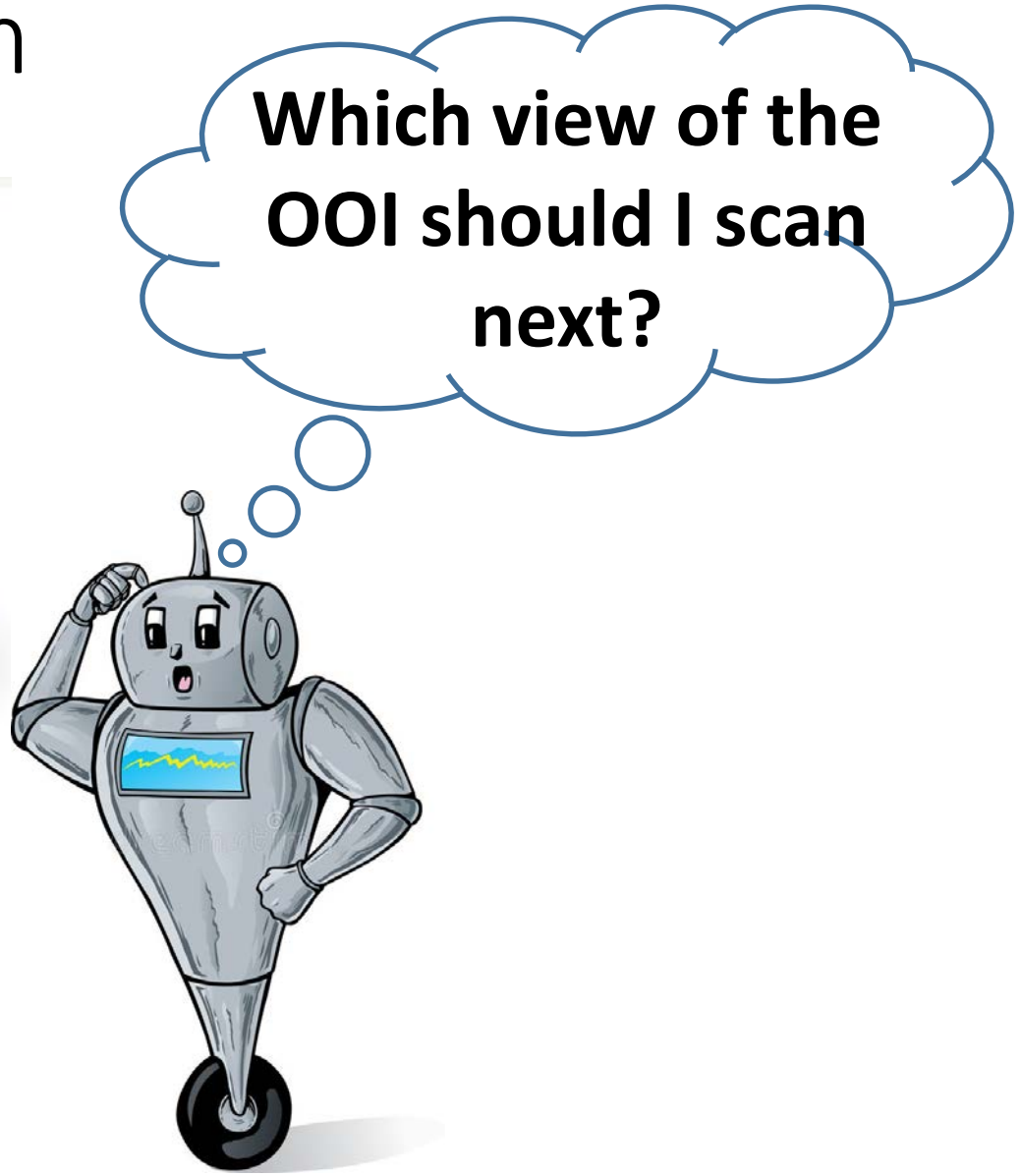
Dataset Construction

Two advantages:

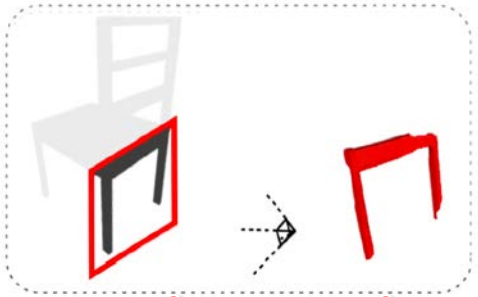
- Decrease the difference between CAD model and scanned model
- Segmented components & component pairs can make retrieval **easier**

The Next Best View Problem





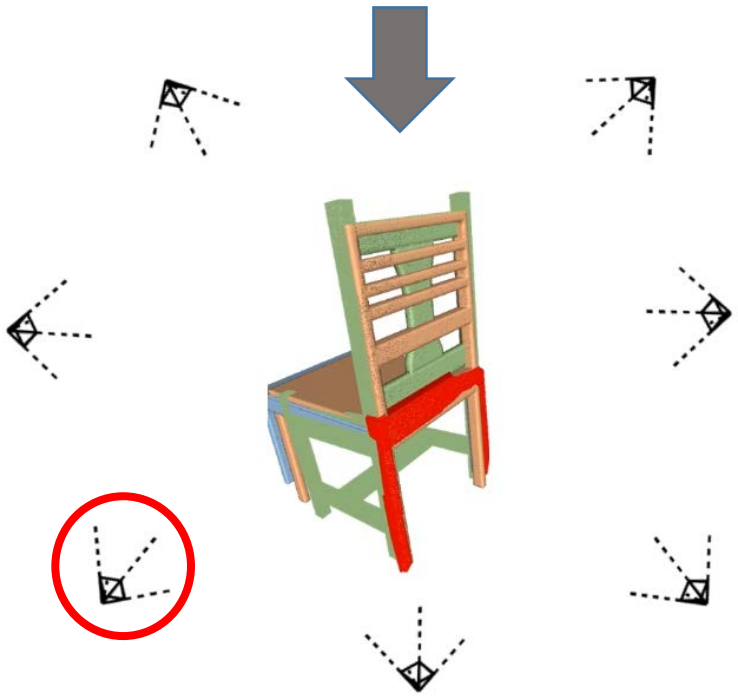
Next Best View

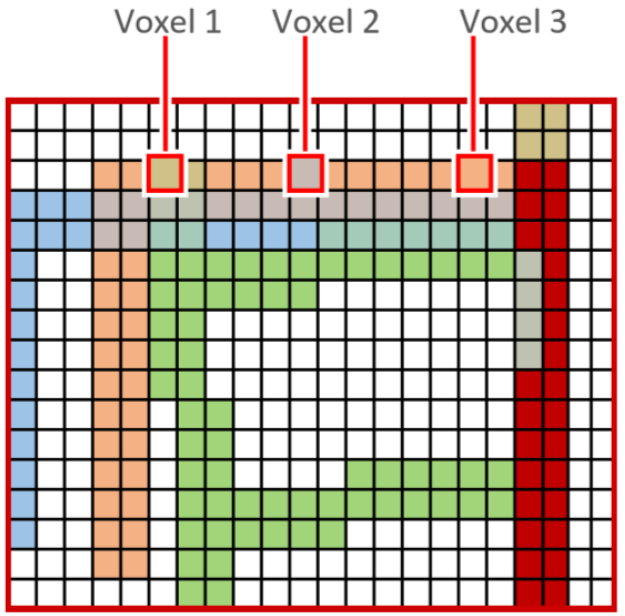


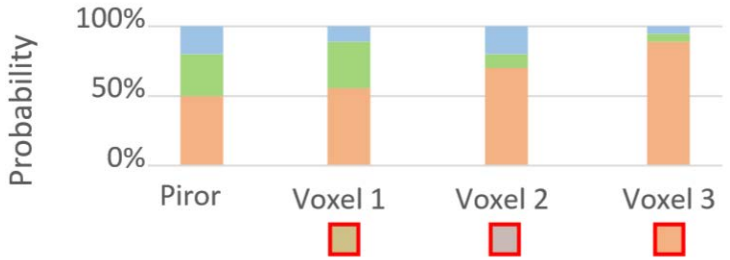
Maximal conditional information gain

$$\max_{j=1,\cdots,n_v} G^j = \sum_{i=1}^{n_s} p(m_i) G^j(m_i)$$

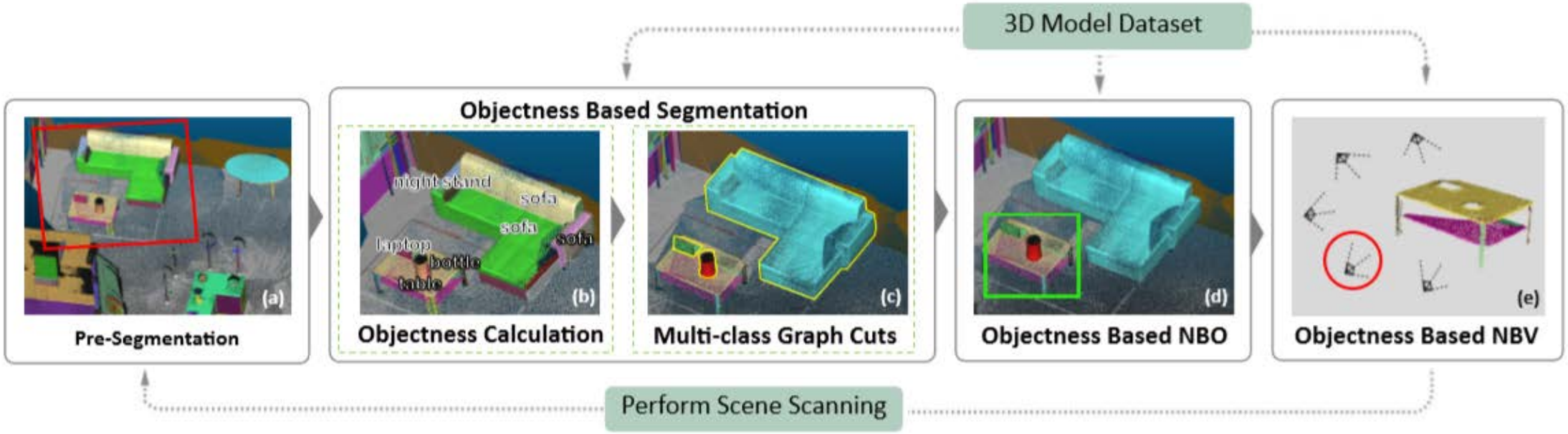
$$\sum_{x \in \Lambda} (H(x) - H(x|m_i))$$





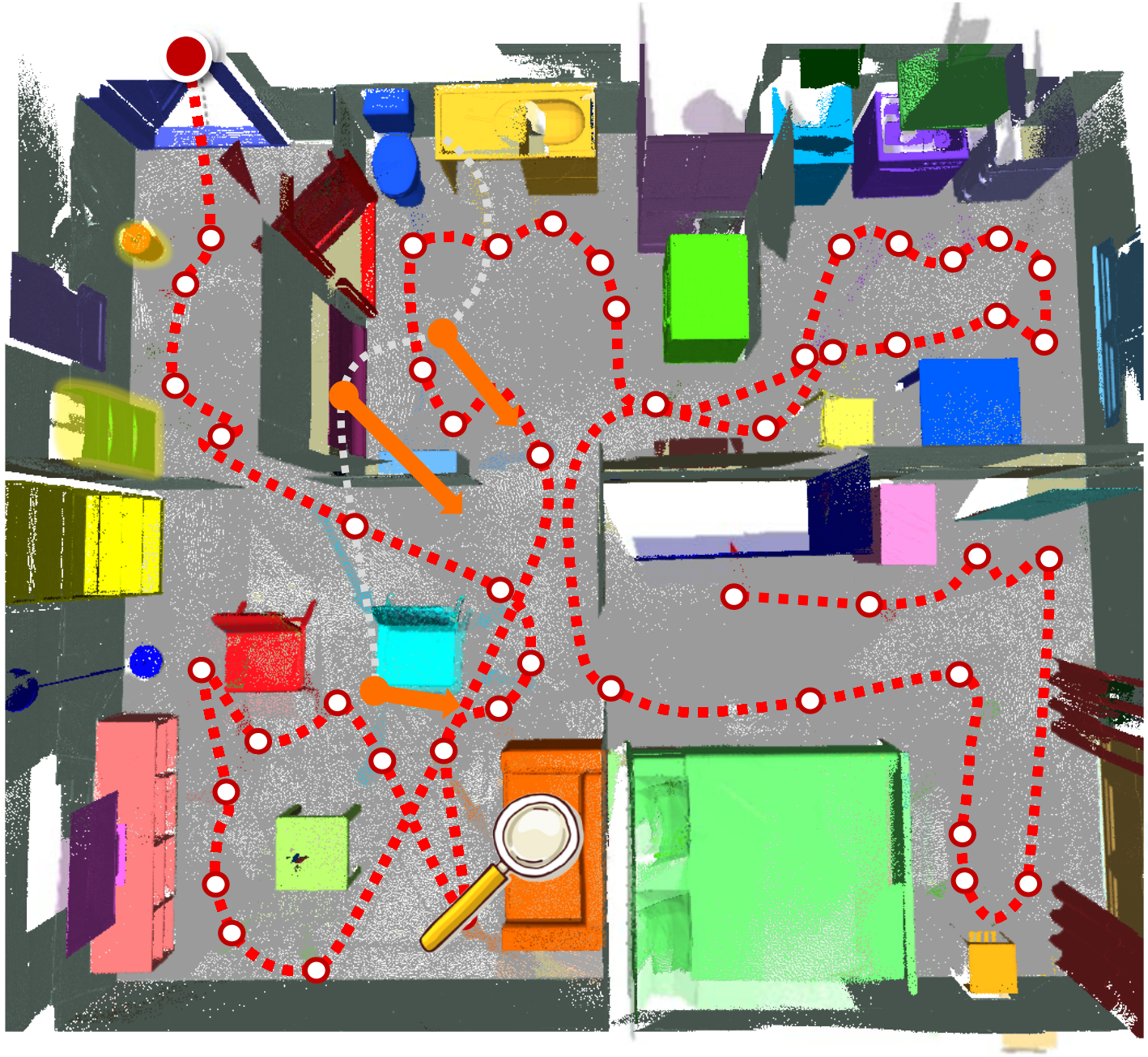


System Pipeline



Key techniques:

- Objectness based segmentation
  - Pre-segmentation
  - Post-segmentation (important)
- Objectness based reconstruction
  - The next best object (NBO)
  - The next best view (NBV)



Evaluation

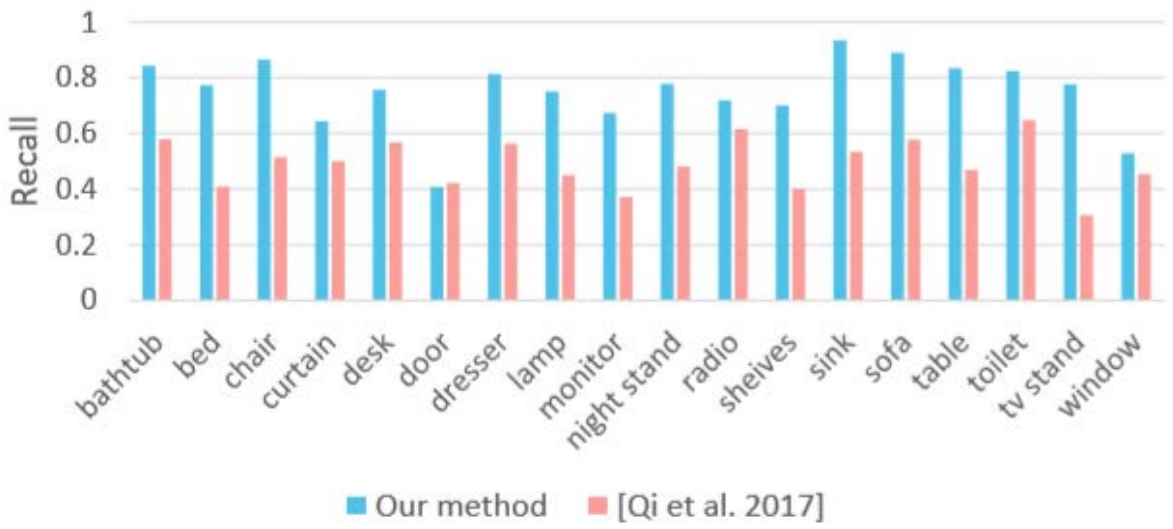
- Virtual scene dataset

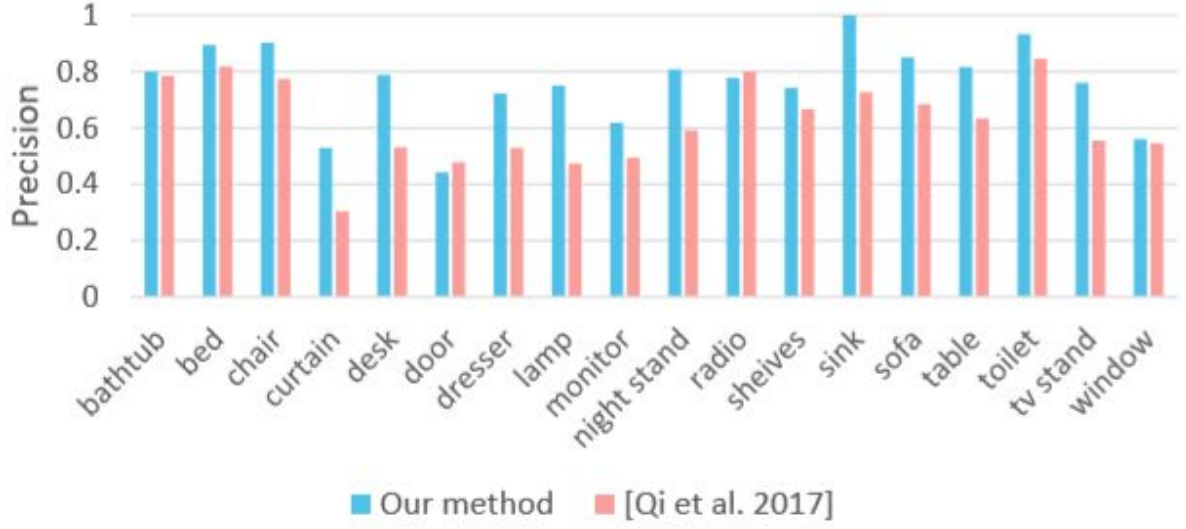
SUNCG (66 scenes)

ScanNet (38 scenes)

Comparison

- Comparing object recognition with PointNet++ [Qi et al. 2017]

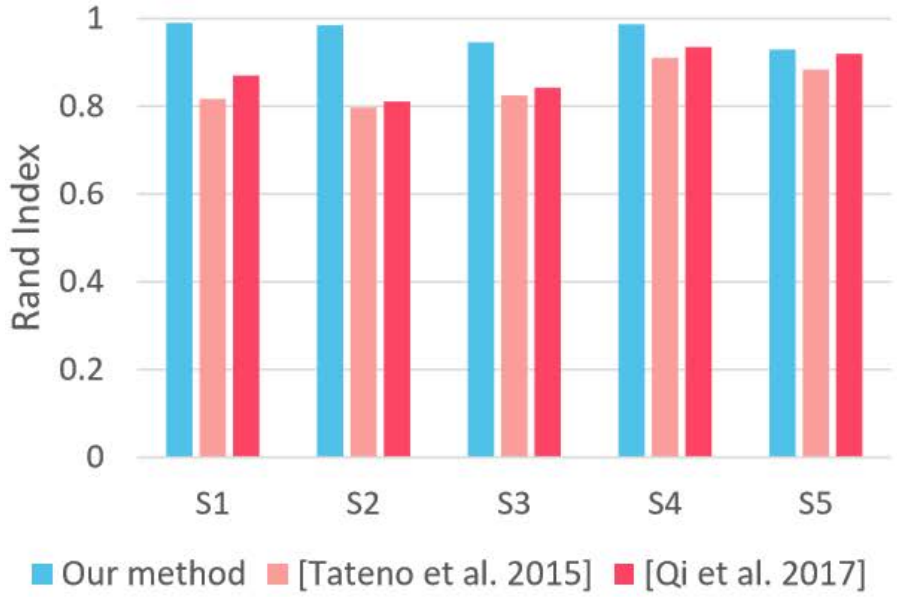


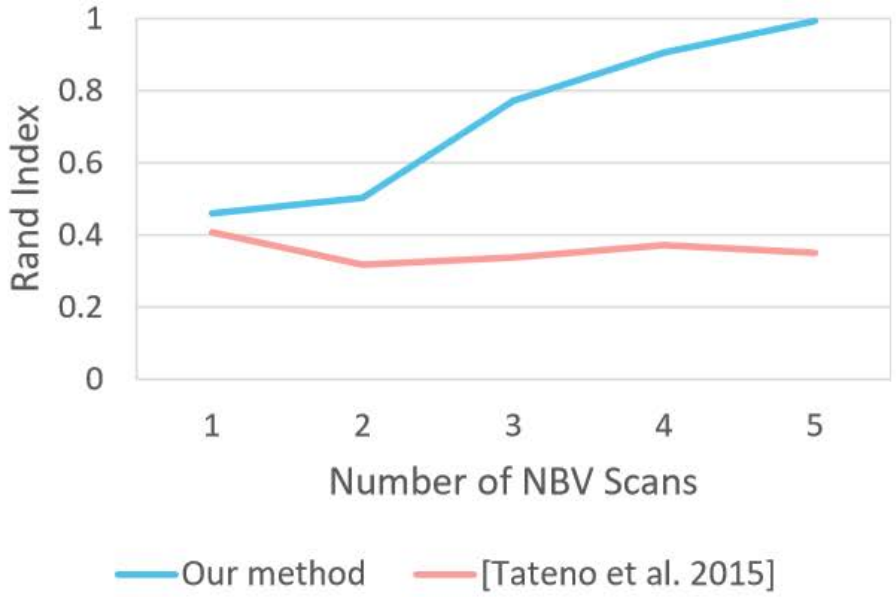


Comparison

- Comparing Rand Index of segmentation

$$RI(S_1, S_2) = \binom{2}{n}^{-1} \sum_{i,j,i<j} [C_{ij}P_{ij} + (1 - C_{ij})(1 - P_{ij})],$$





Comparison

- Comparing object coverage rate and quality against tensor field guided autoscanning [Xu et al. 2017]

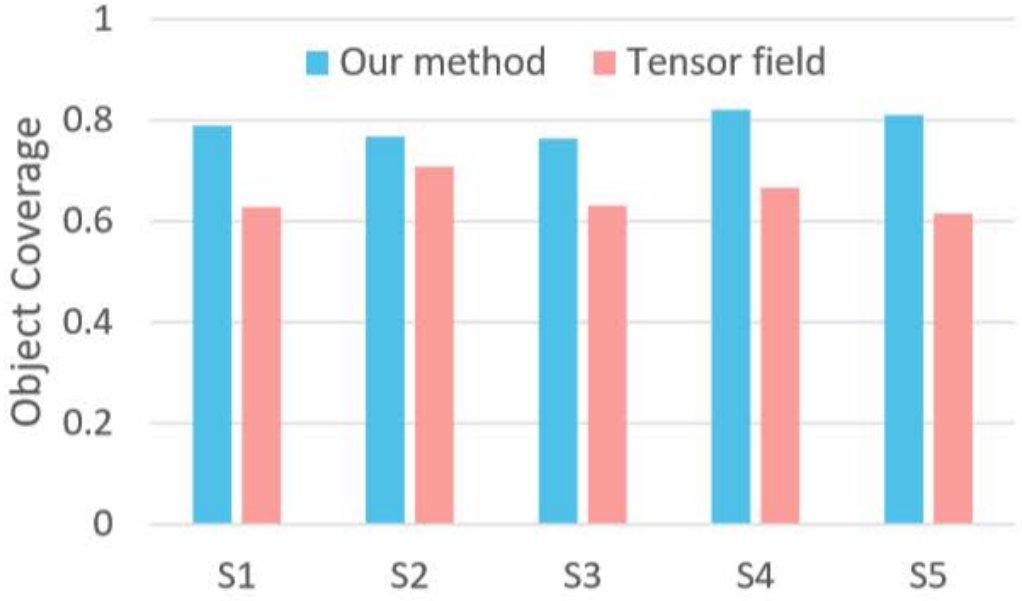
$$R_{\text{cover}} = \frac{1}{|\mathcal{V}_S|} \int_{v \in \mathcal{V}_S} \delta_{\text{detect}}(v) \cdot \delta_{\text{vis}}(v),$$

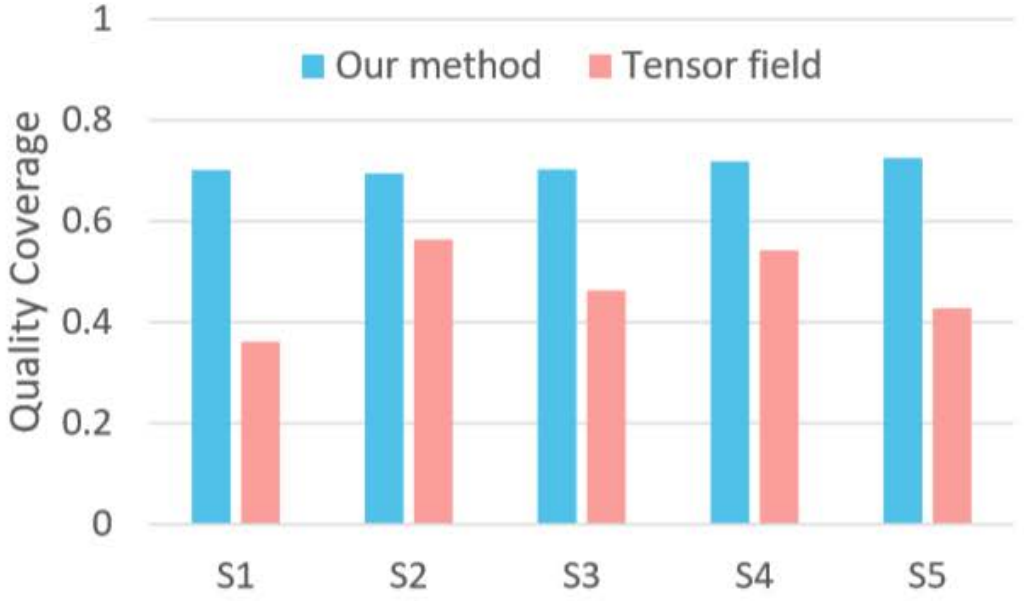
$$Q_{\text{cover}} = \frac{1}{|\mathcal{V}_S|} \int_{v \in \mathcal{V}_S} \delta_{\text{detect}}(v) \cdot \delta_{\text{vis}}(v) \cdot q(v),$$

Depth noise

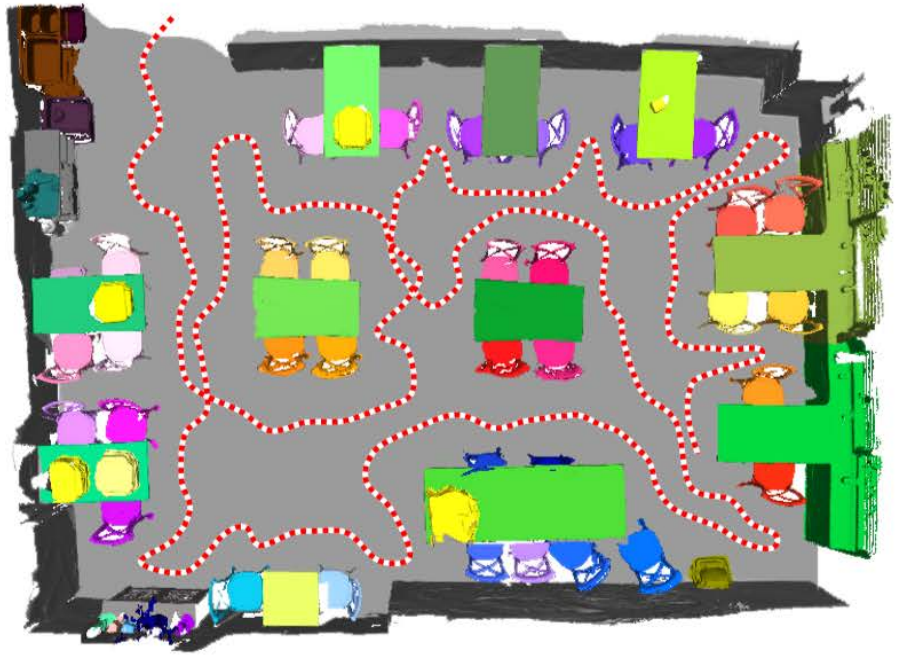
Comparison

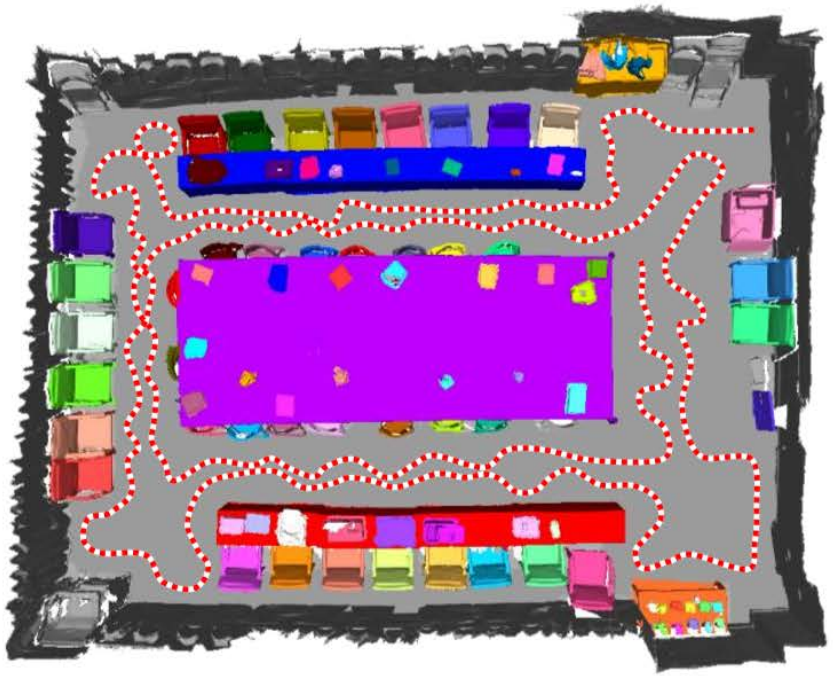
- Comparing object coverage rate and quality against tensor field guided autoscanning [Xu et al. 2017]



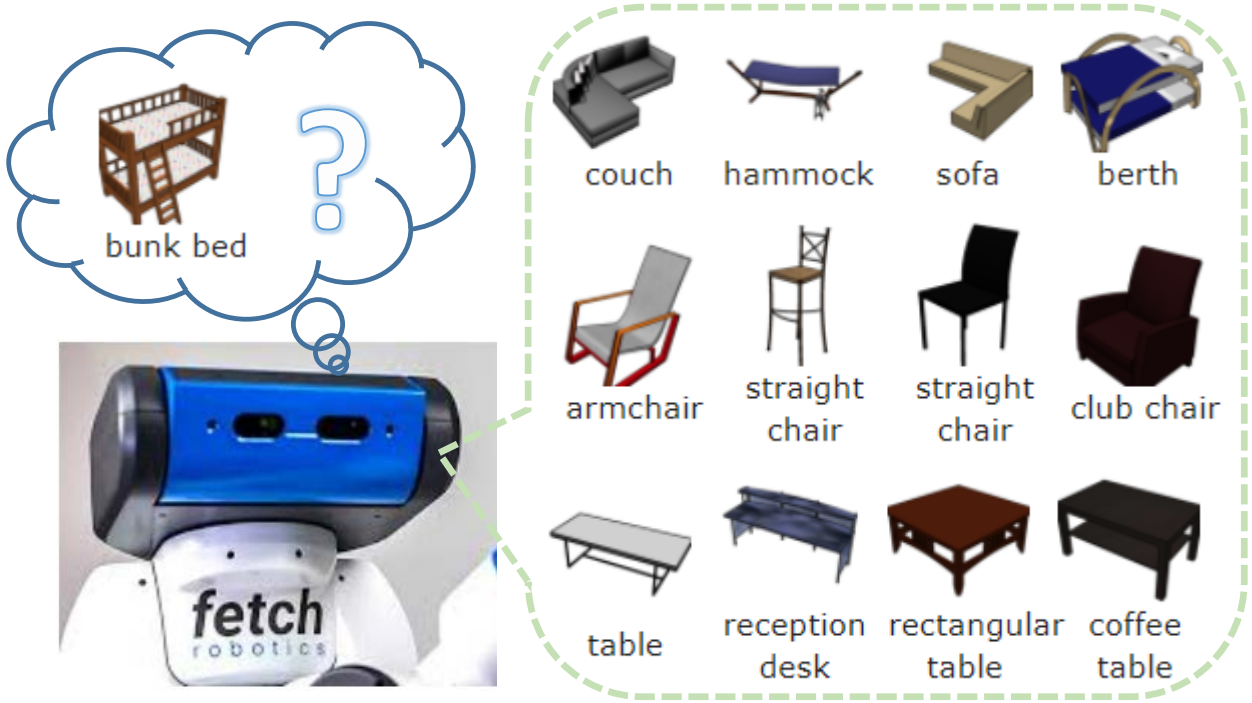


More Results





Limitations



No similar models

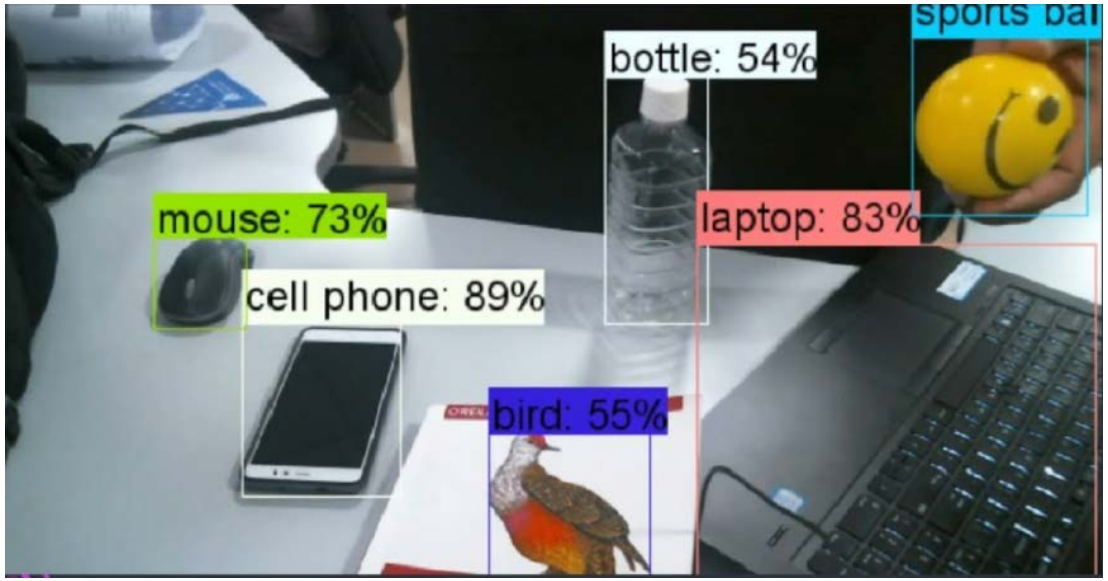
Cluttered scenes

Limitations & Future Works

Single object

Group structure

Future Works



Combine image-based method

Driverless car with LiDAR

Conclusion

- An object-guided approach for autonomous scene exploration, reconstruction, and understanding
  - Model-driven objectness
  - Objectness-based segmentation
  - Objectness-based NBO strategy
  - Objectness-based NBV strategy
- Coupled global exploration and local scanning
- Coupled segmentation and recognition

Thank you!

Q & A