GENERATIONS / VANCOUVER 12-16 AUGUST
SIGGRAPH2018

OBJECT-AWARE GUIDANCE FOR AUTONOMOUS SCENE RECONSTRUCTION

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

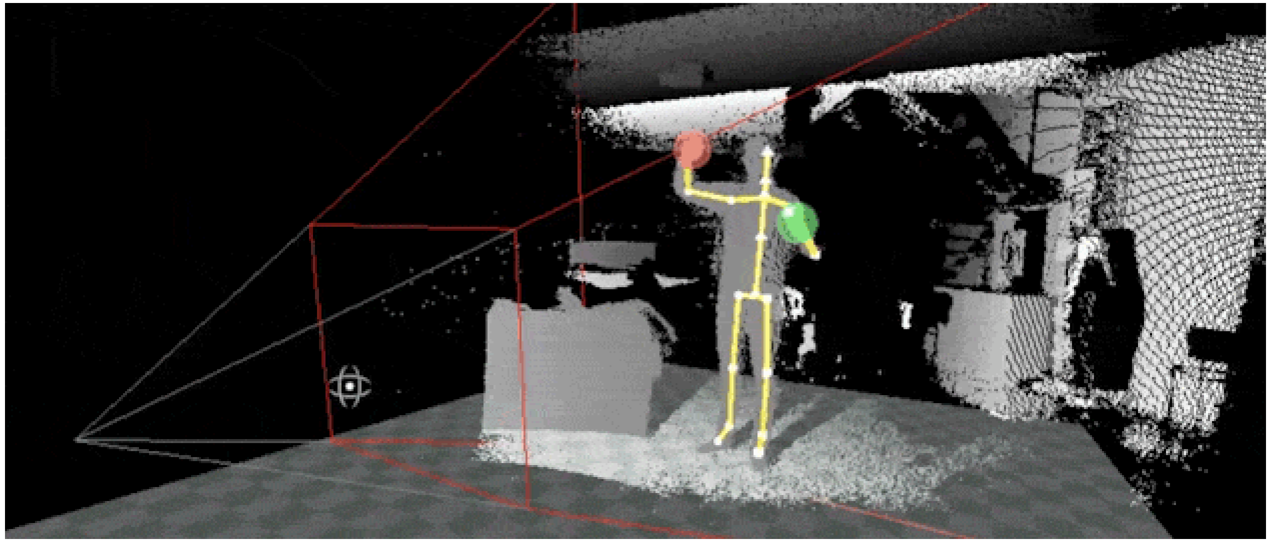
Juzhan Xu, Bin Chen, Hui Huang, Kai Xu

University of Science and Technology of China
Shenzhen University
National University of Defense Technology

Photography & Recording Encouraged

Background

- Commodity RGB-D sensors



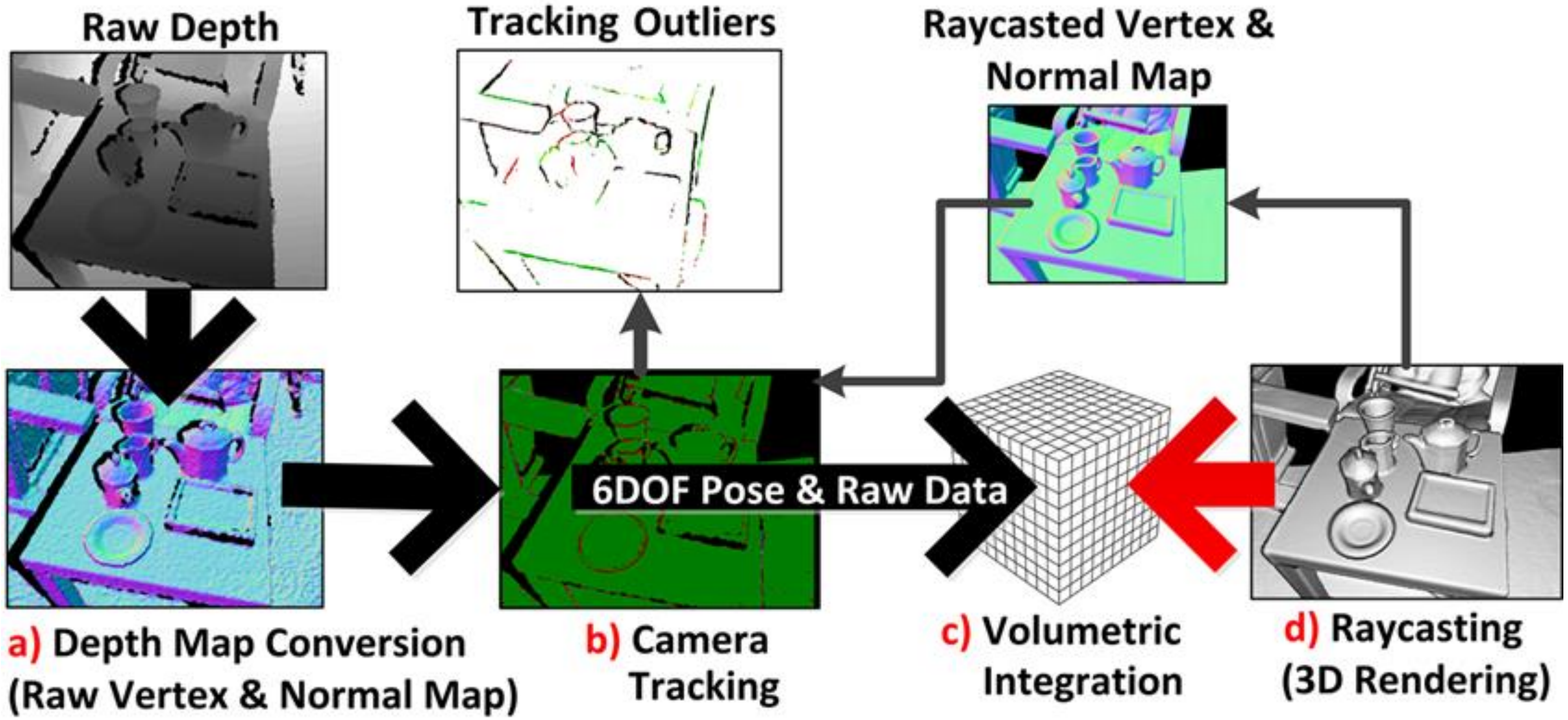
Microsoft Kinect

PrimeSense

Intel RealSense

Background

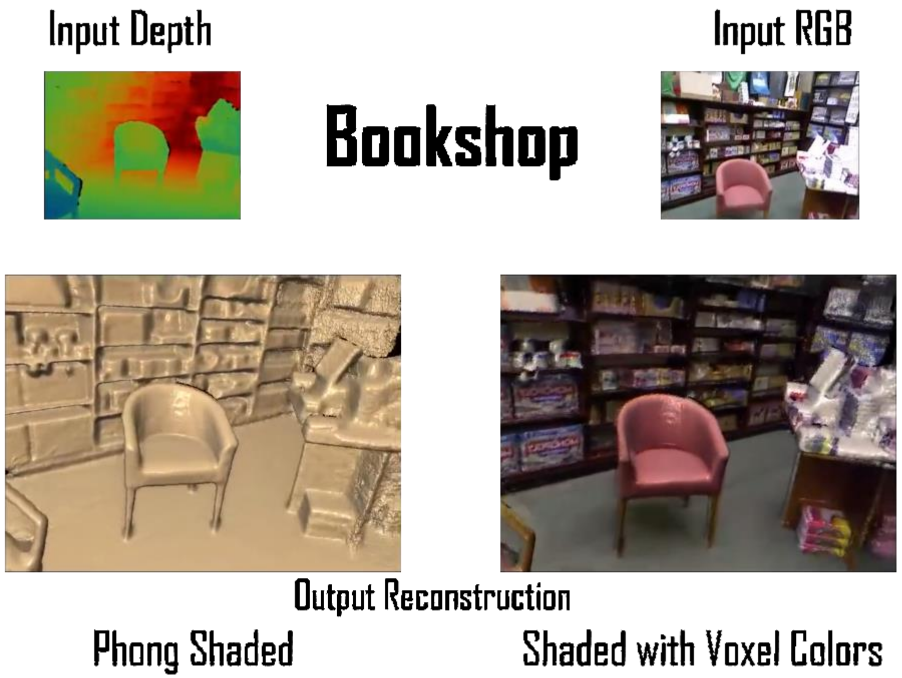
- RGB-D sensor allows real-time reconstruction



KinectFusion
[Izadi et al. 2011]

Background

- Other real-time reconstruction methods

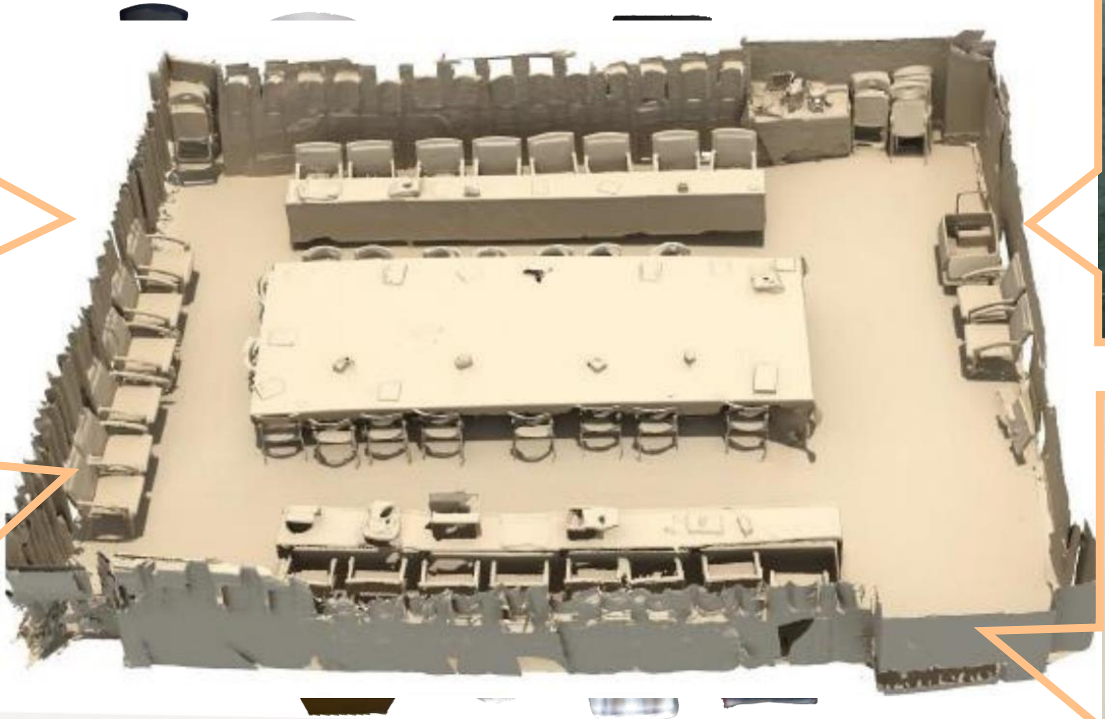


Voxel Hashing
[Nießner et al. 2013]

ElasticFusion
[Whelan et al. 2015]

Background

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



Background

GENERATIONS / VANCOUVER 12-16 AUGUST
SIGGRAPH2018

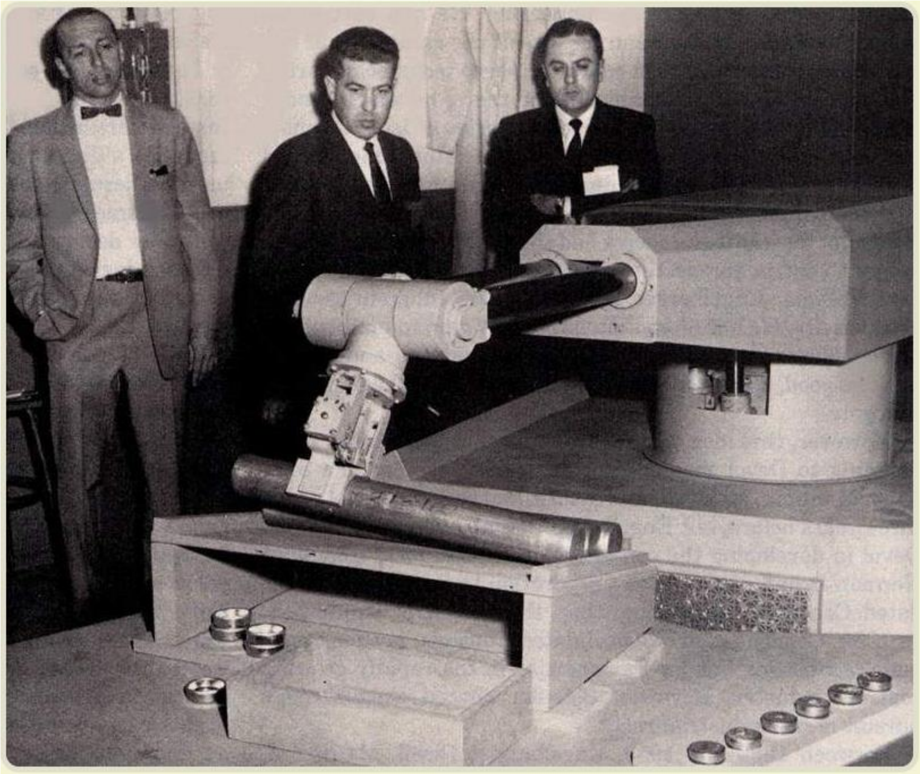
- Human scanning is a laborious task [Kim et al. 2013]

Time consuming

Inaccurate scanning

Background

- Modern robots are more and more reliable and controllable.



Unimation, 1958

Fetch, 2015

Motivation: Autoscanning with Robots

Never feel tired

Automatic

Stable and precise

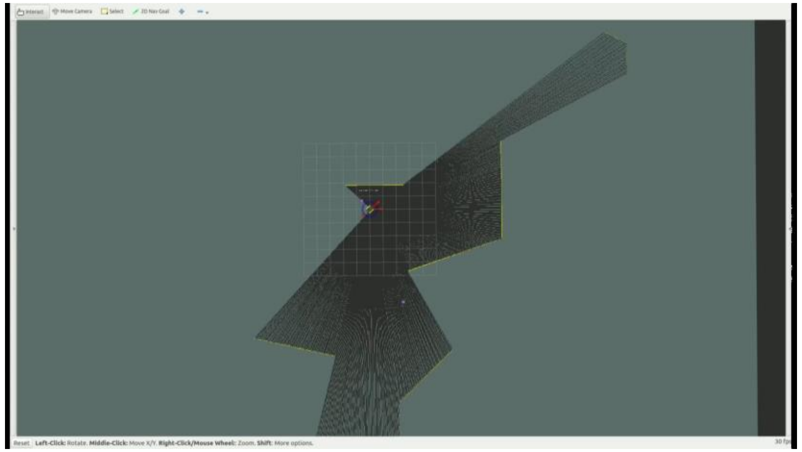
Existing Works: Single Objects

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

Existing Works: Unknown Scenes

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

First Pass | **Second Pass**



frontier-based exploration
[Yamauchi et al. 1997]

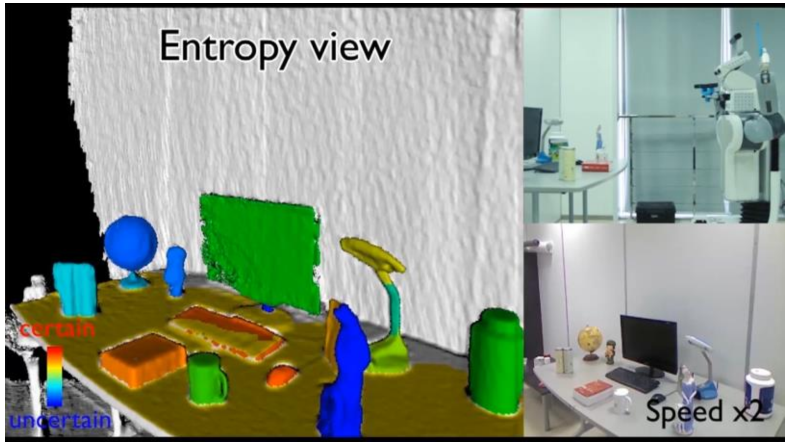
exploration & reconstruction
[Xu et al. 2017]

segmentation & recognition
[Nan et al. 2012]

Existing Works: Unknown Scenes

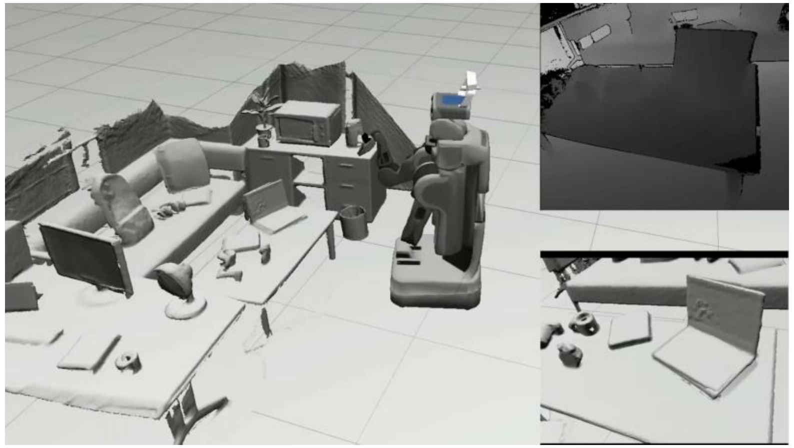
- Two pass scene reconstruction and understanding.
- Can only use **low-level** information in first exploration pass.

First Pass



reconstruction & segmentation
[Xu et al. 2015]

Second Pass



object recognition
[Xu et al. 2016]

The Main Challenge

- How to automatically achieve scene reconstruction and understanding in one pass?

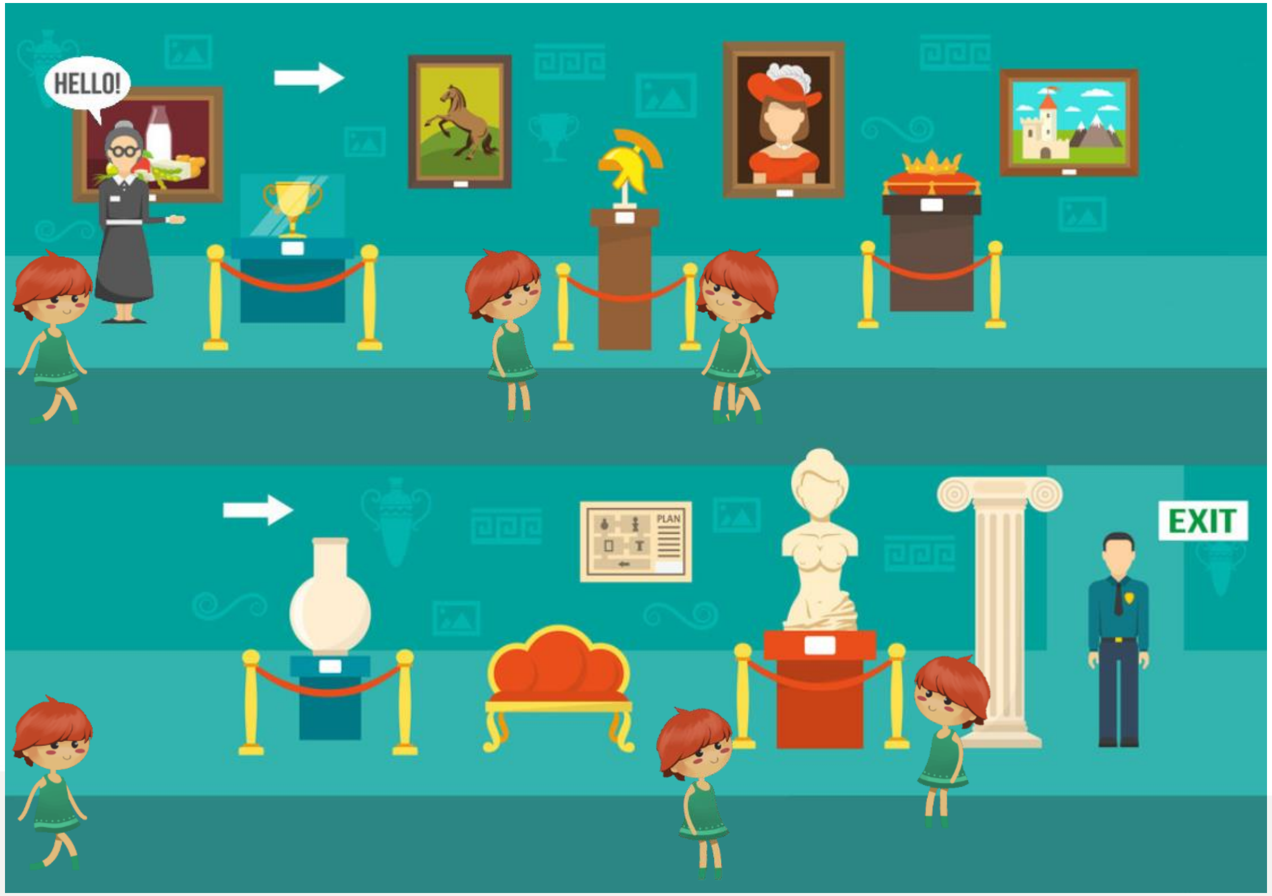
Reconstruction

Understanding

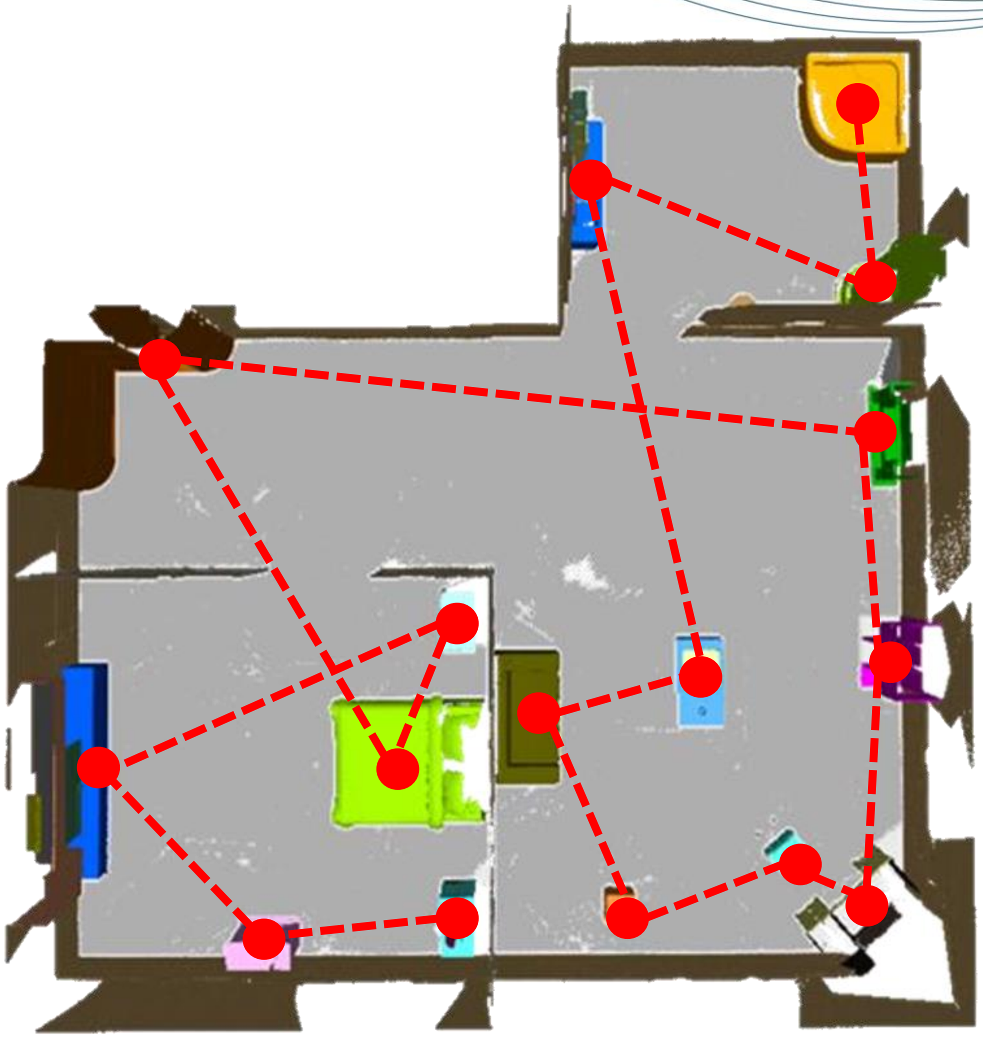
One pass?

Motivation

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



Our Solution



- **Key idea**: using recognized **objects** as a **guidance** map

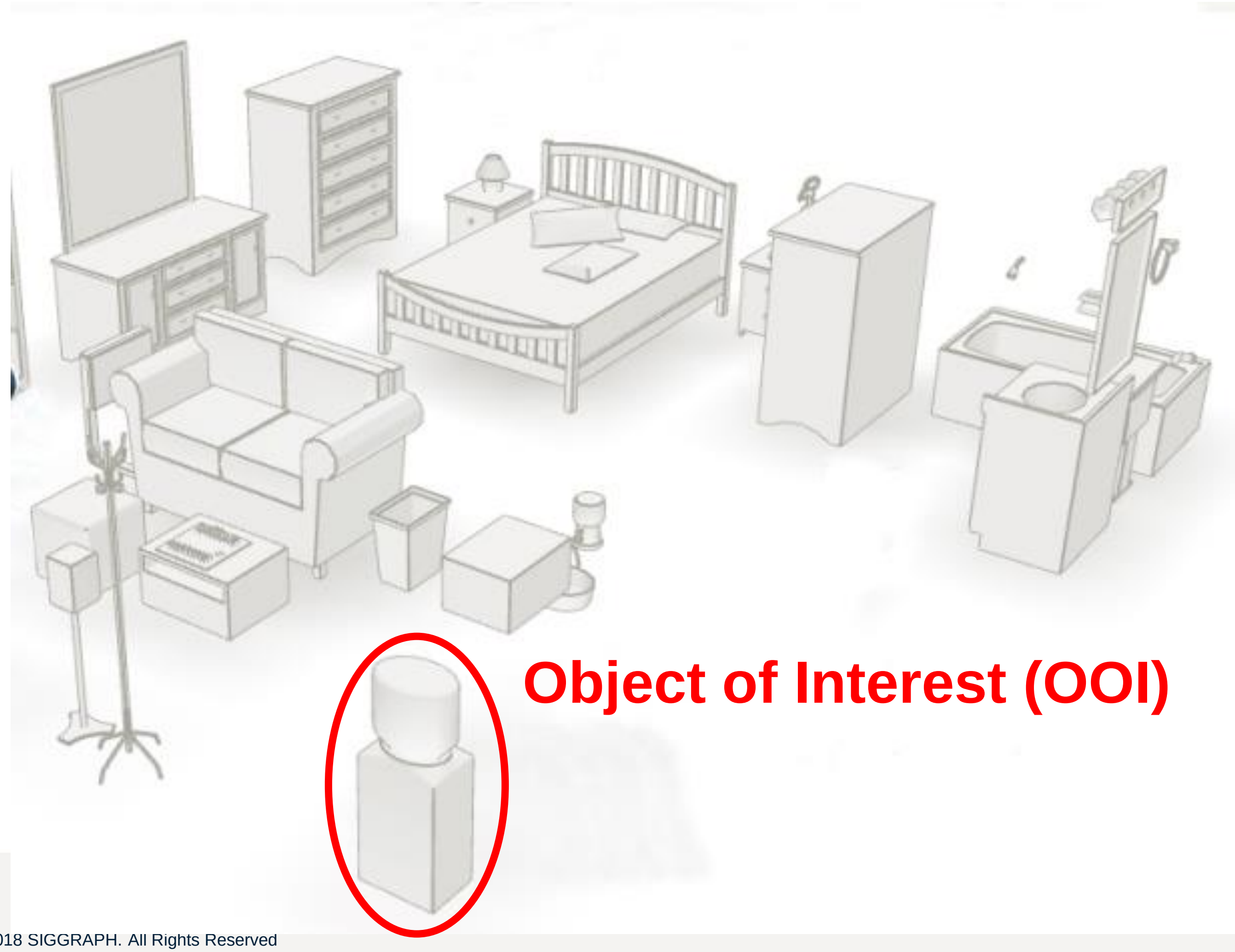
We need to

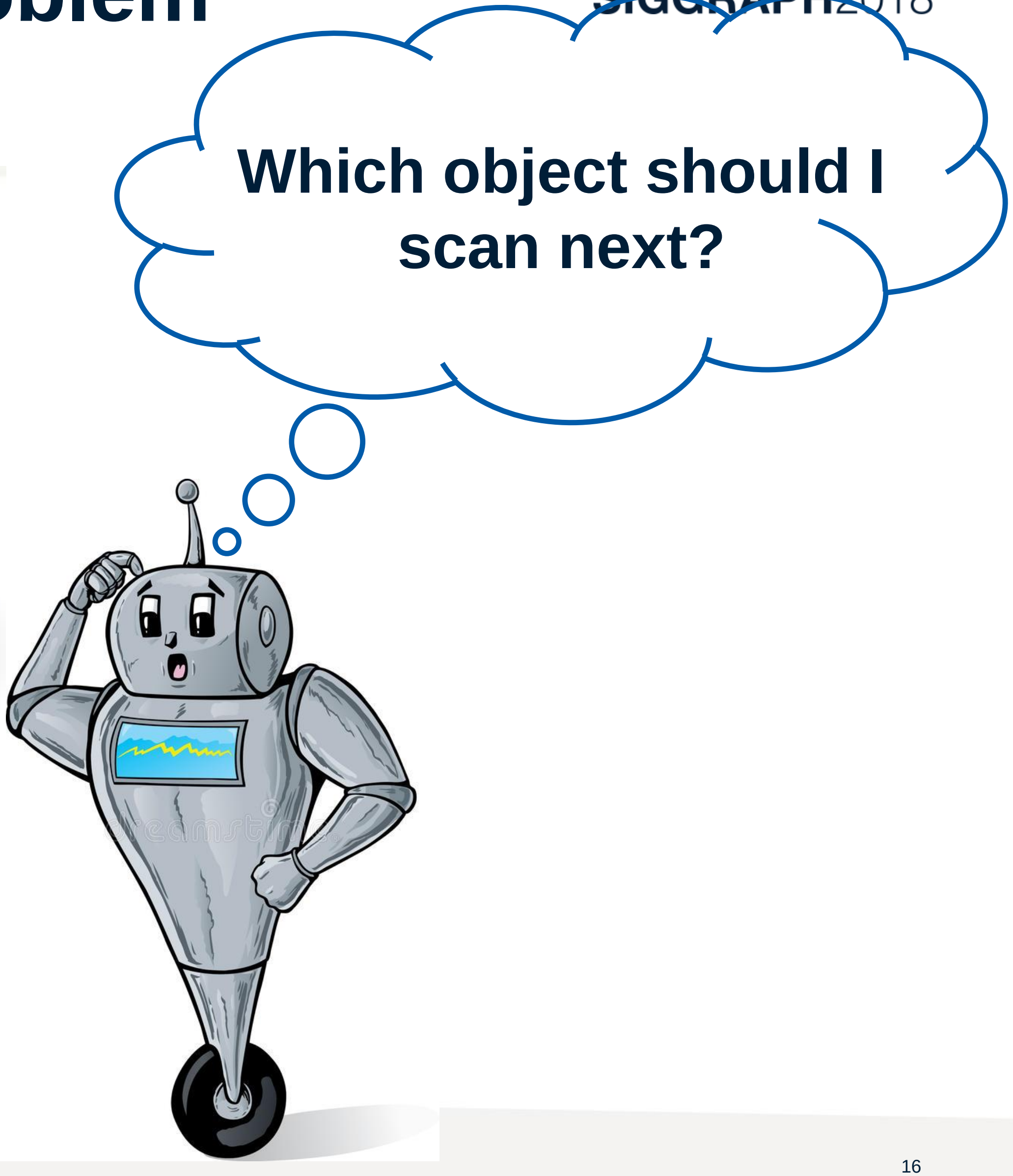
One navigation

Automatic

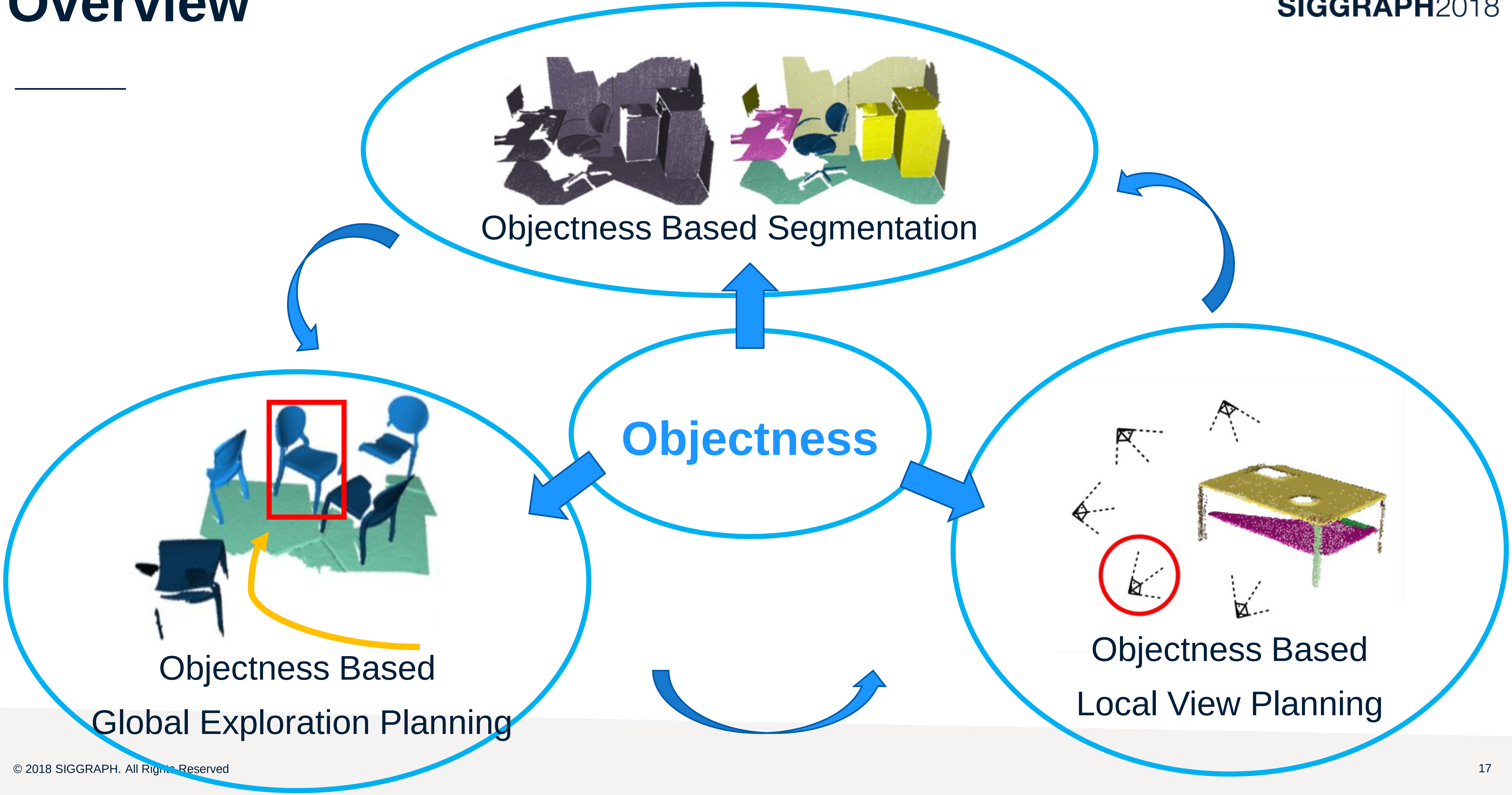
Scene Understanding

Phase 1: The Next Best Object Problem





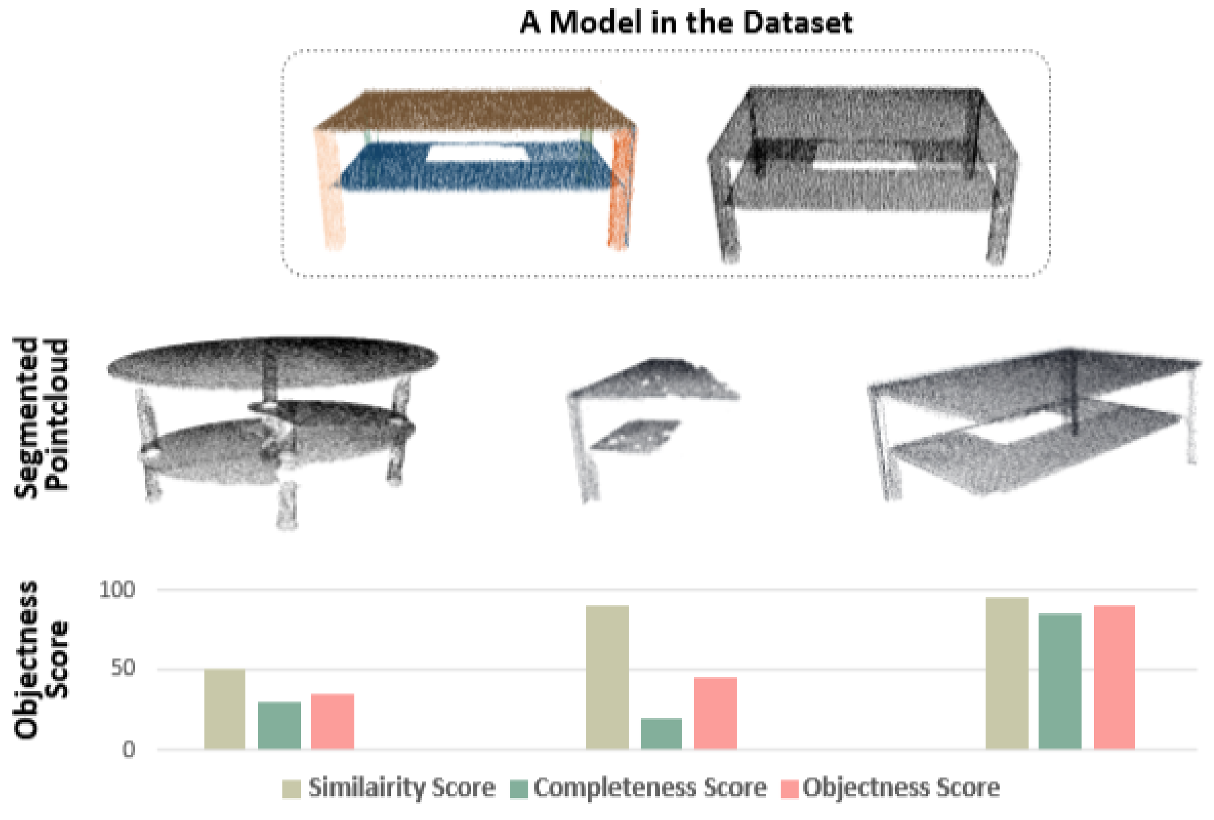
Overview


Objectness Based Segmentation
Objectness
Objectness Based
Global Exploration Planning
Objectness Based
Local View Planning

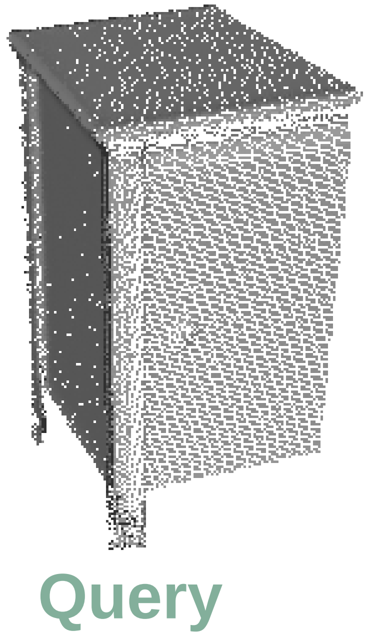
Model-Driven Objectness

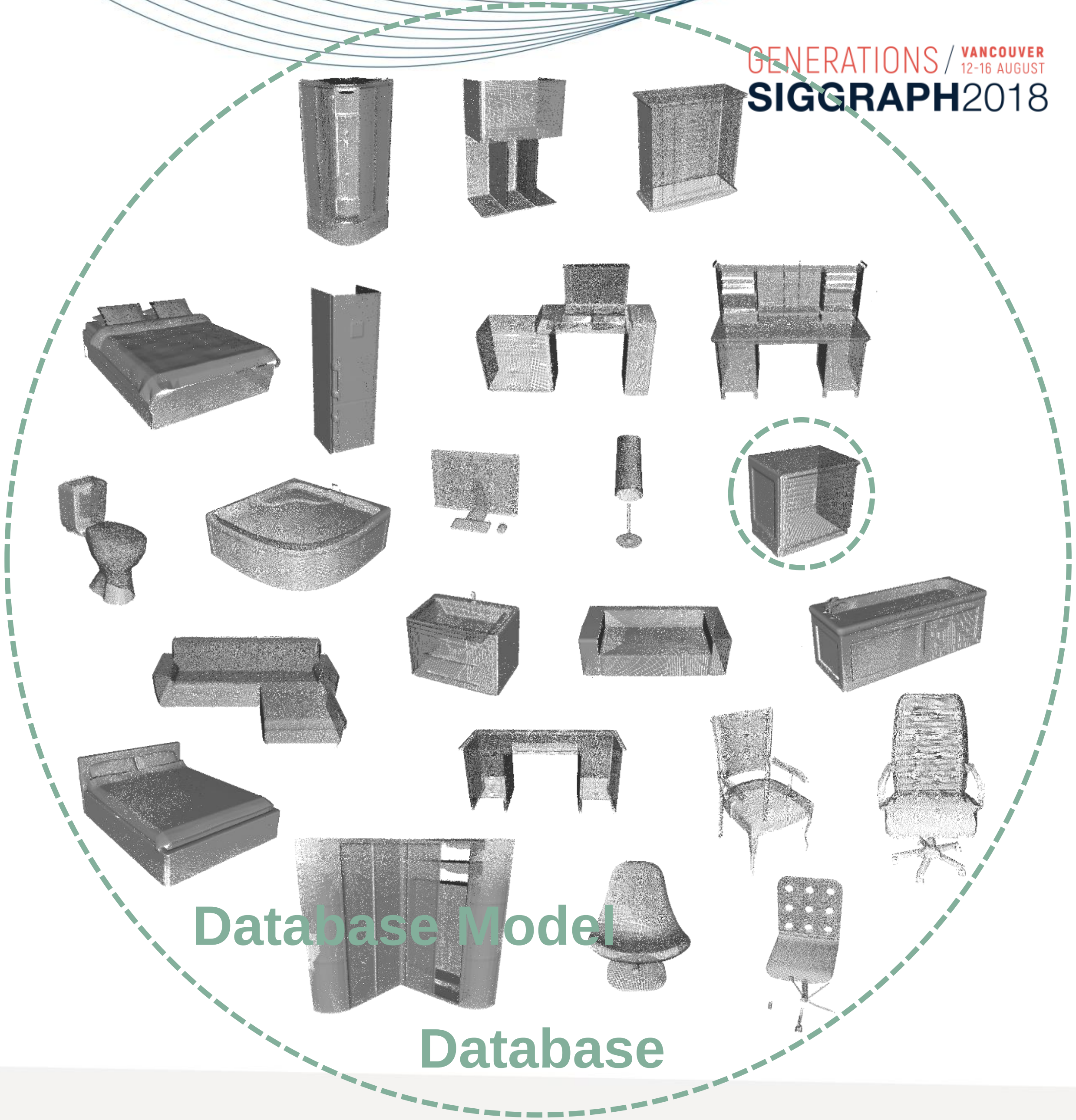
GENERATIONS / VANCOUVER 12-16 AUGUST
SIGGRAPH2018

- Objectness should measure both similarity and completeness

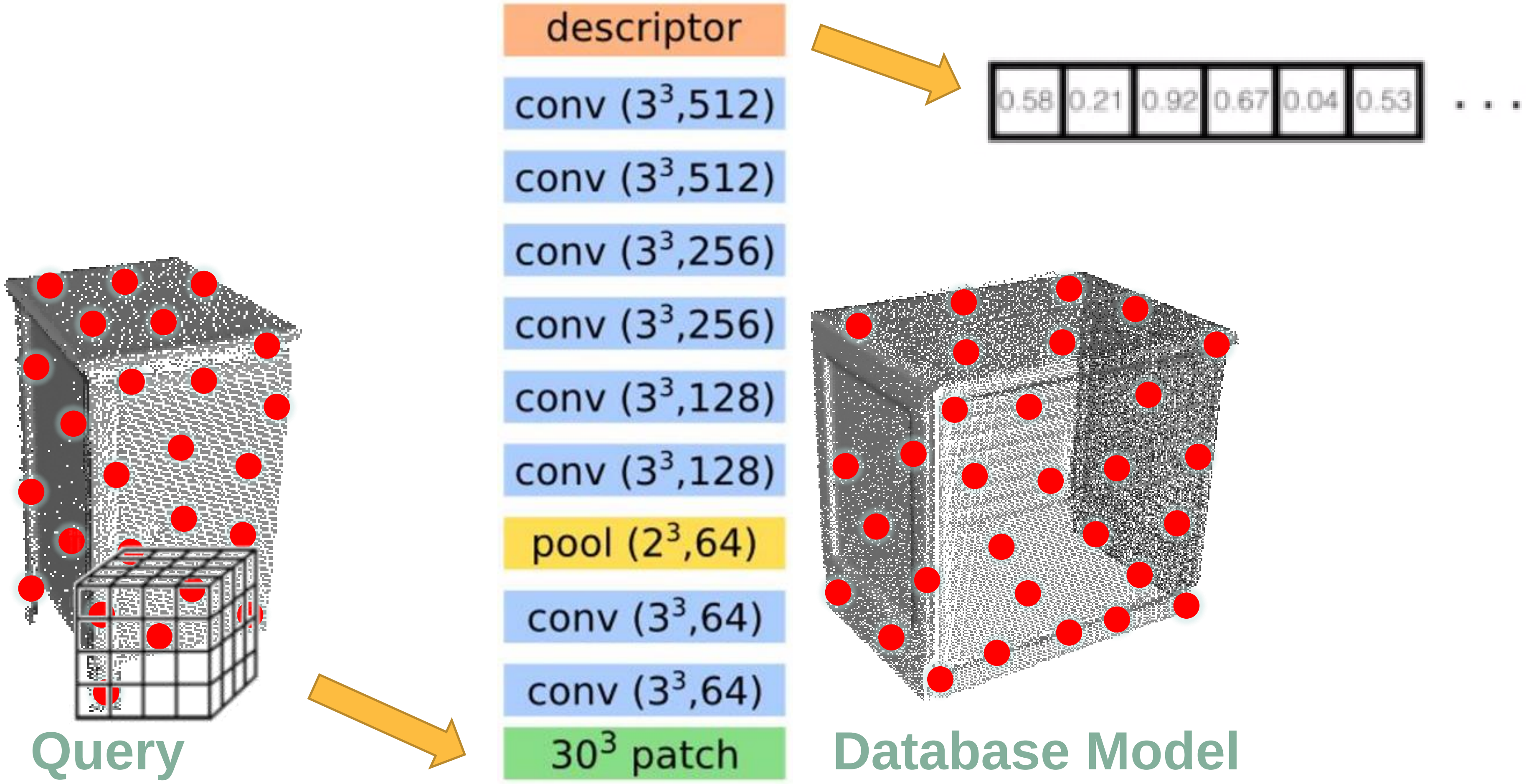


Partial Matching



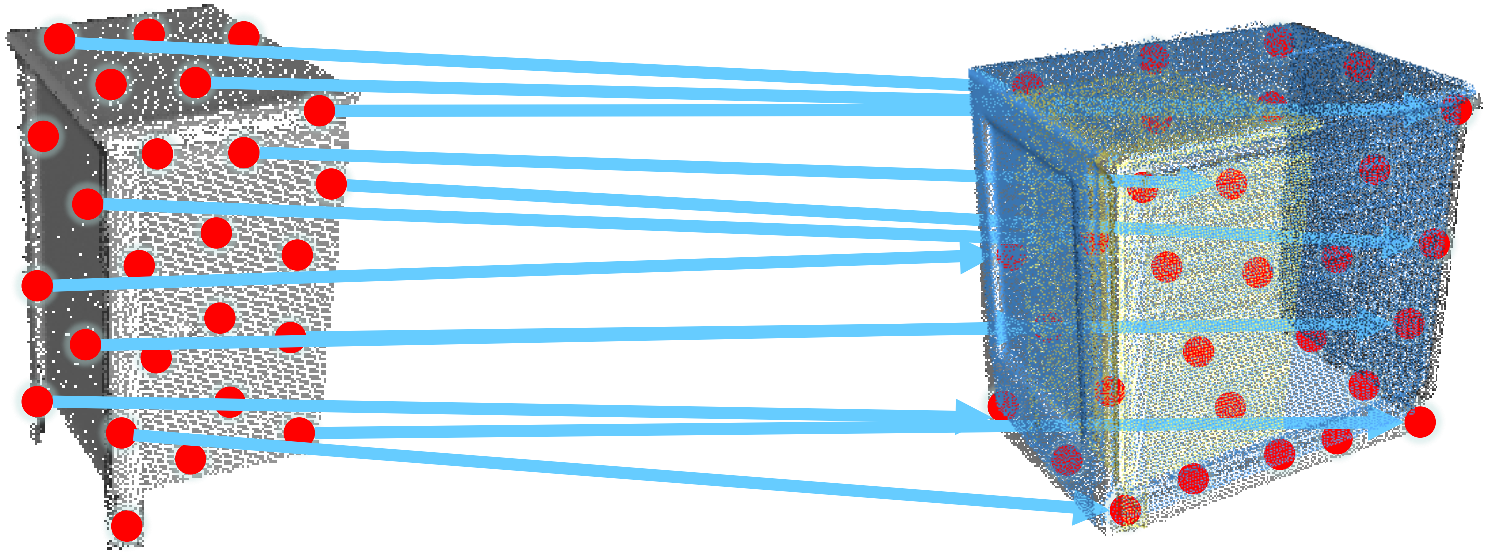


Partial Matching



3DMatch [Zeng et al. 2016]

Partial Matching

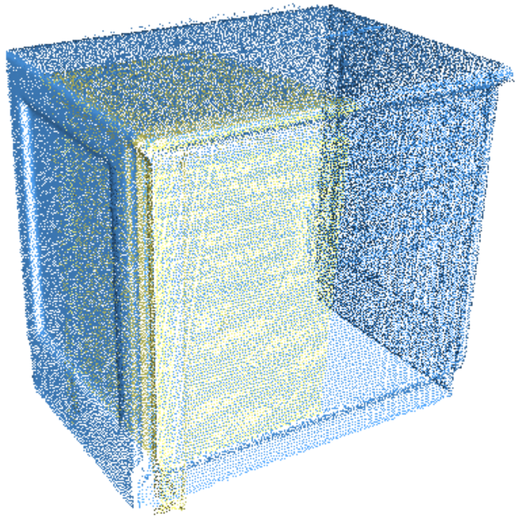


Model-Driven Objectness

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

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

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



Objectness **Similarity** **Completeness**

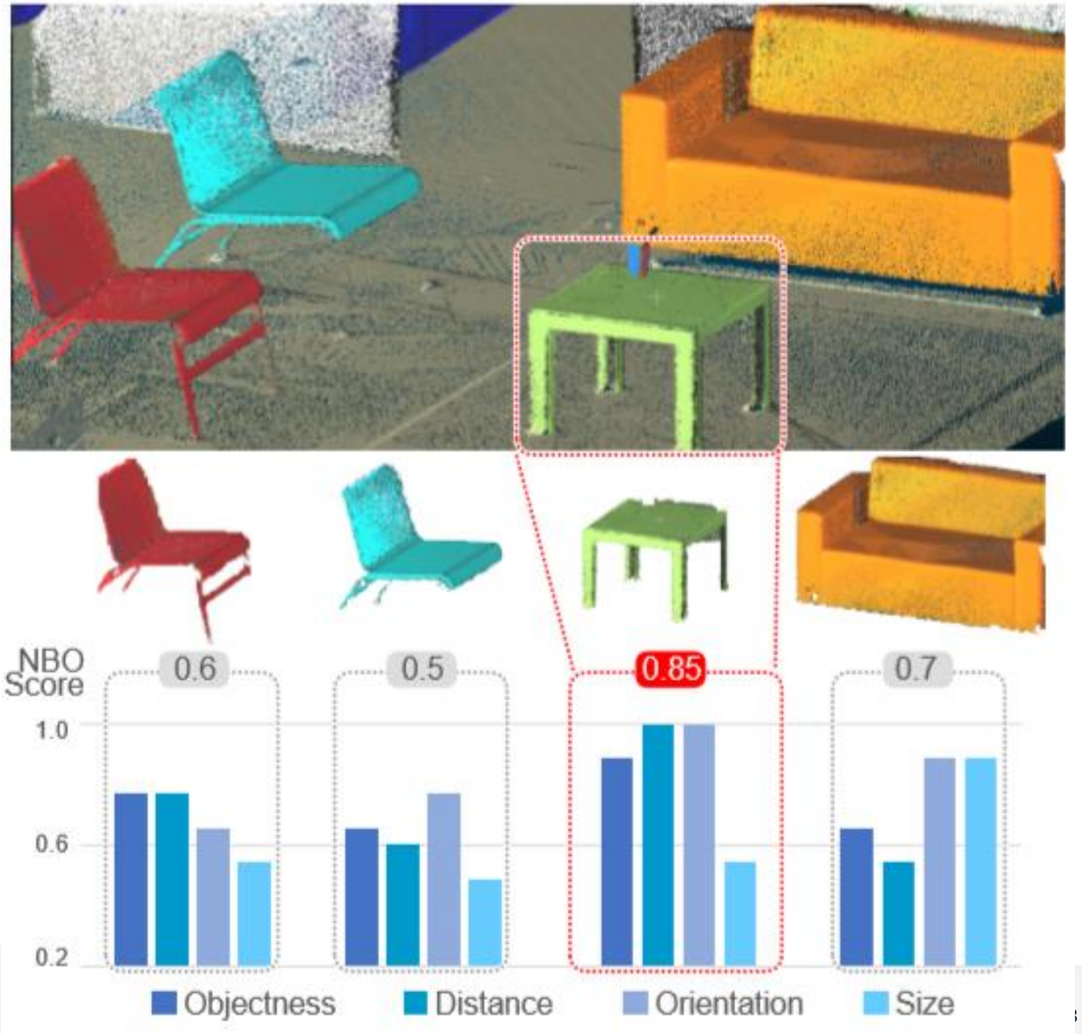
Next Best Object

Objectness

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

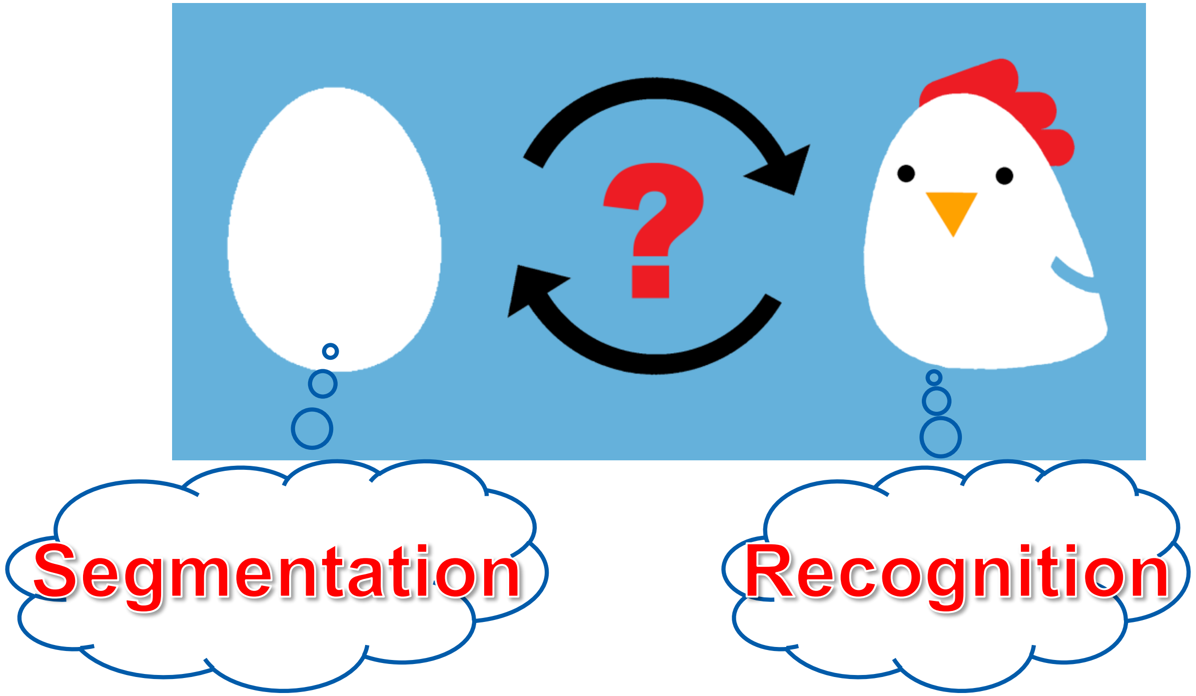
$$S(r) = \underline{w_z S_z(r)} + \underline{w_e S_e(r)} + \underline{w_d S_d(r)}$$

Distance **Orientation** **Size**



Technical Challenge

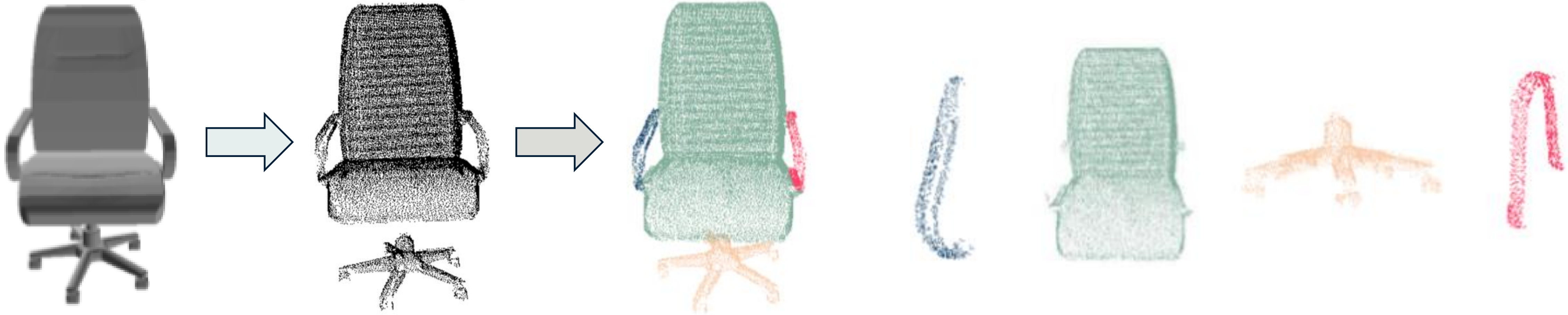
- How to segment and recognize objects during reconstruction?



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

Pre-segmentation

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



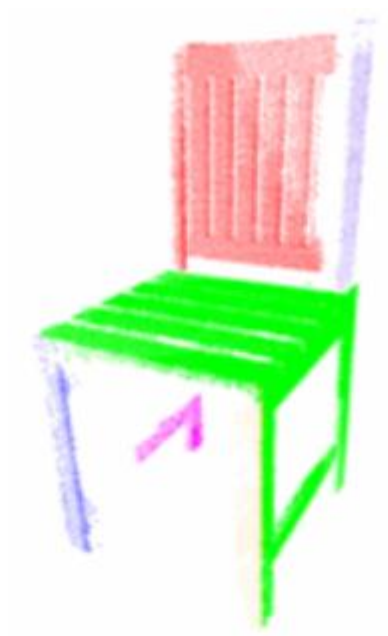
Indoor object

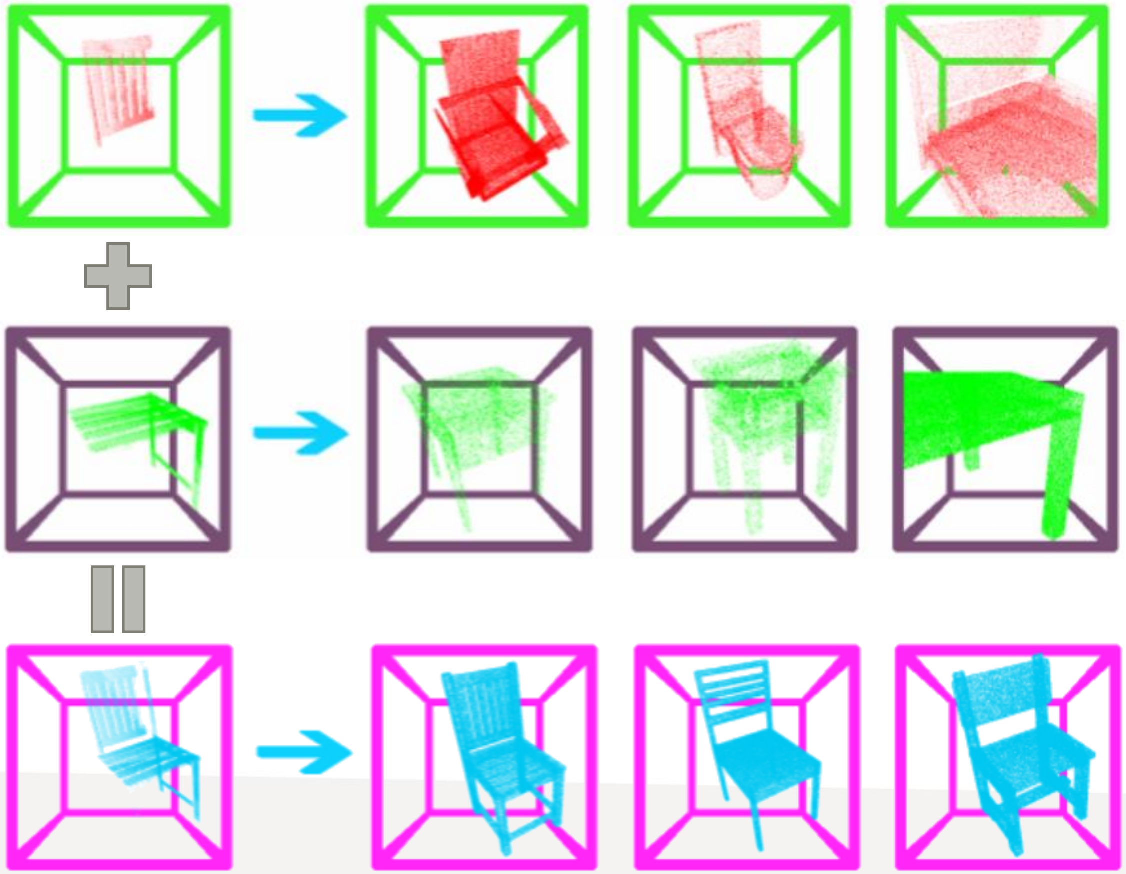
Scanned Model

Pre-segmented Components

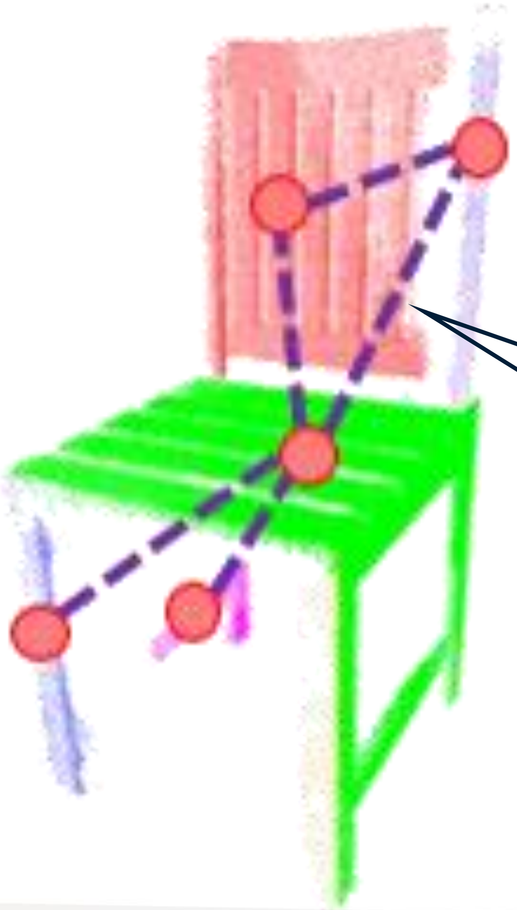
Post-segmentation

- Couples segmentation and recognition in the same optimization





Post-segmentation



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

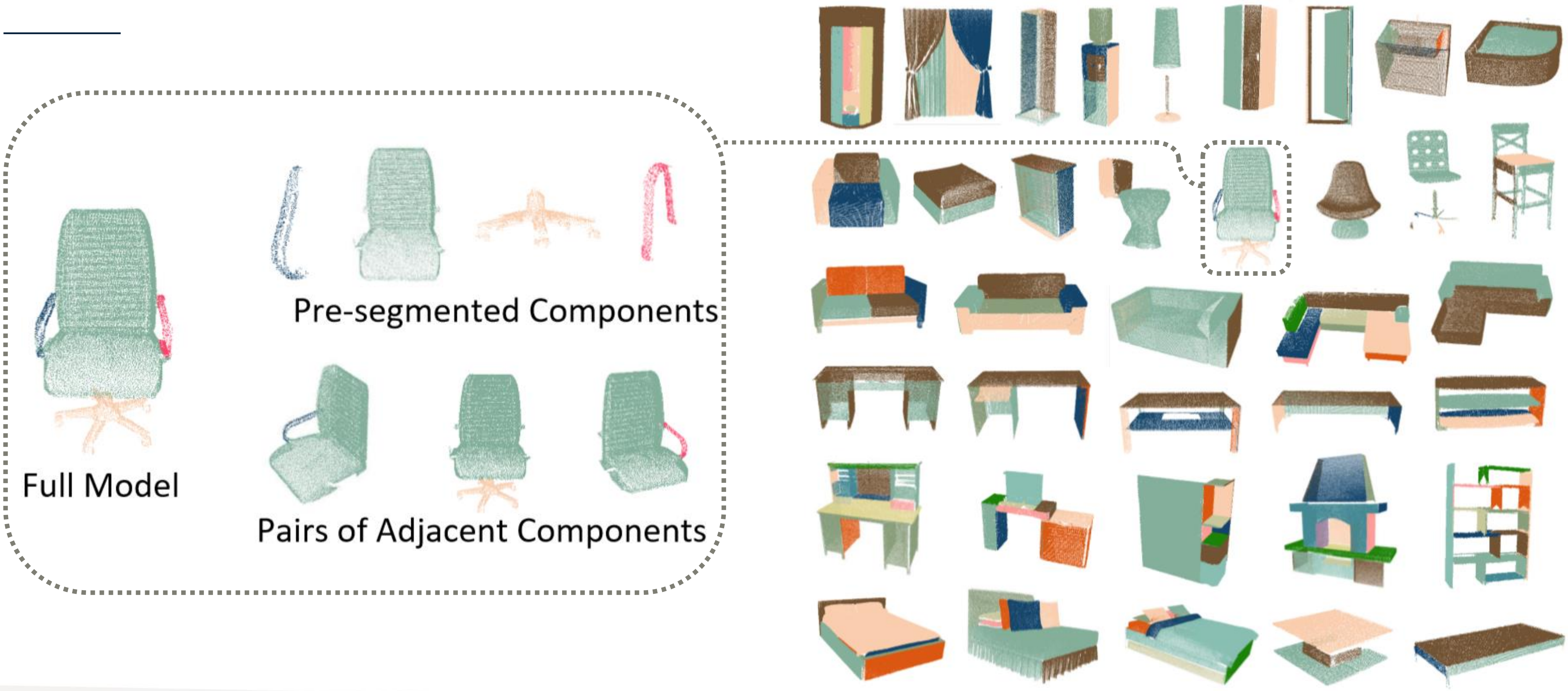
$$E_S(l_c, l_d) = \begin{cases} \max\limits_{m \in M(c \cup d)} O(c \cup d, m), & \text{if } l_c \neq l_d \\ 0, & \text{if } l_c = l_d \end{cases}$$

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

Post-segmentation Results

Database Construction

GENERATIONS / VANCOUVER 12-16 AUGUST
SIGGRAPH2018

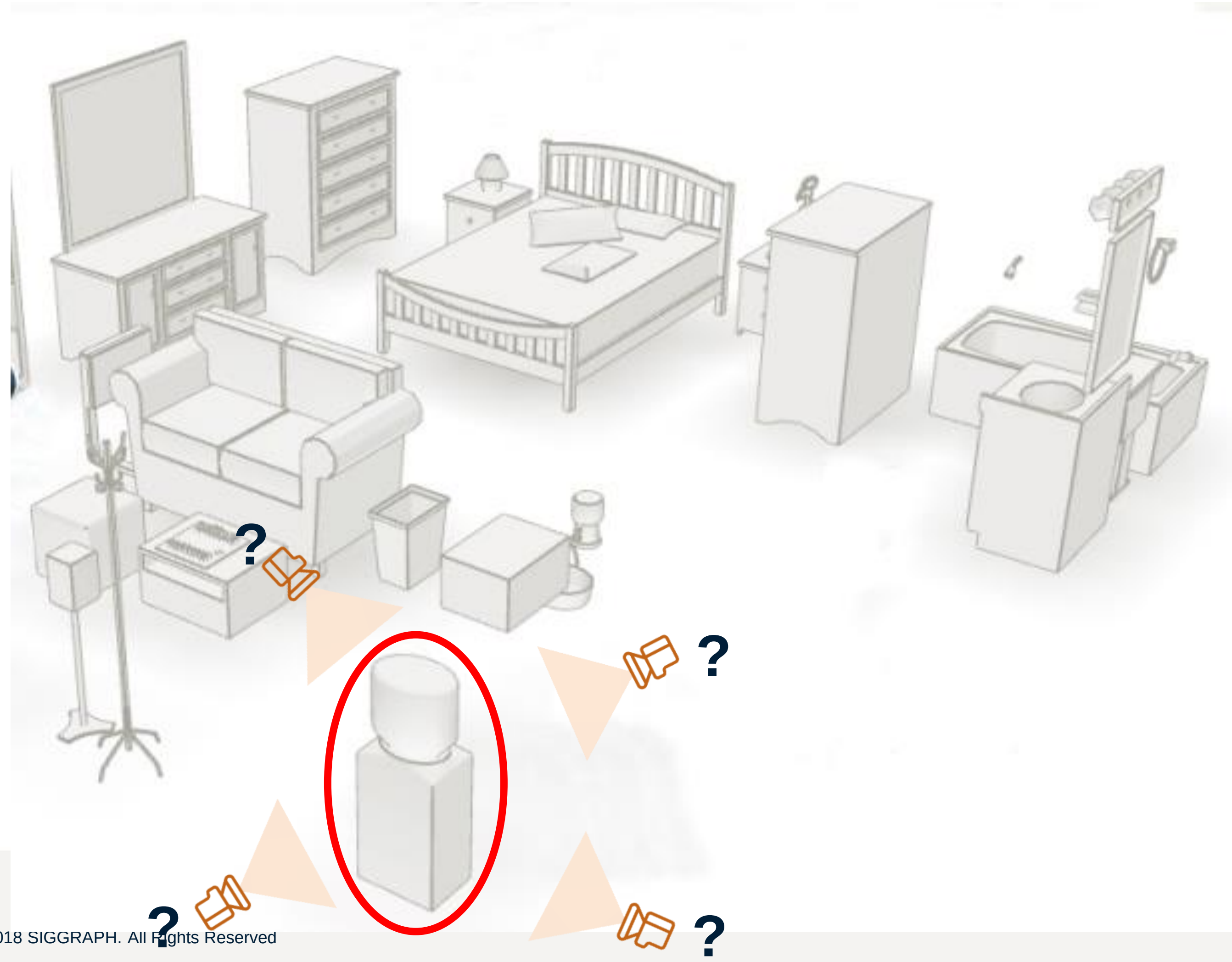


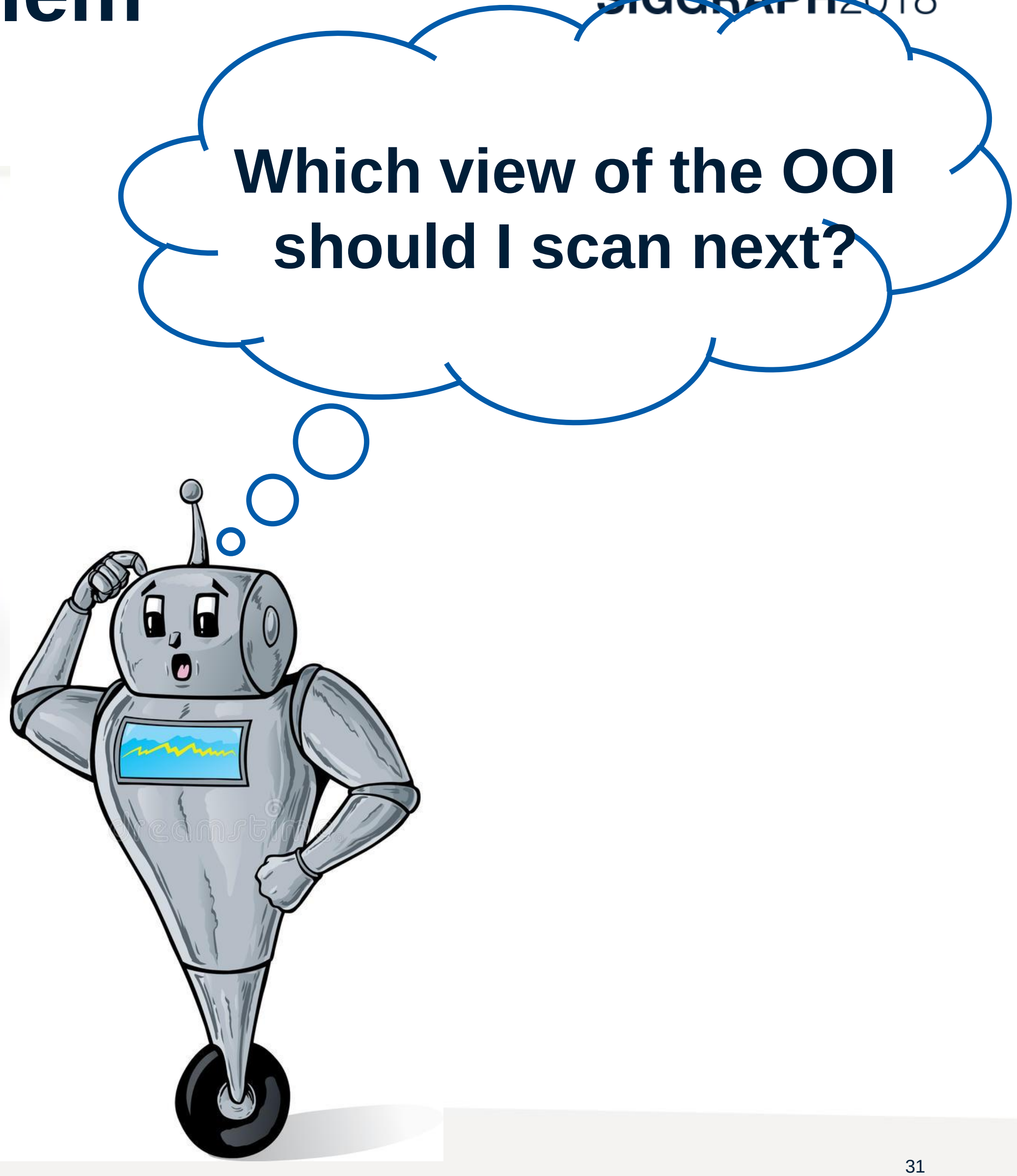
Database Construction

Two advantages:

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

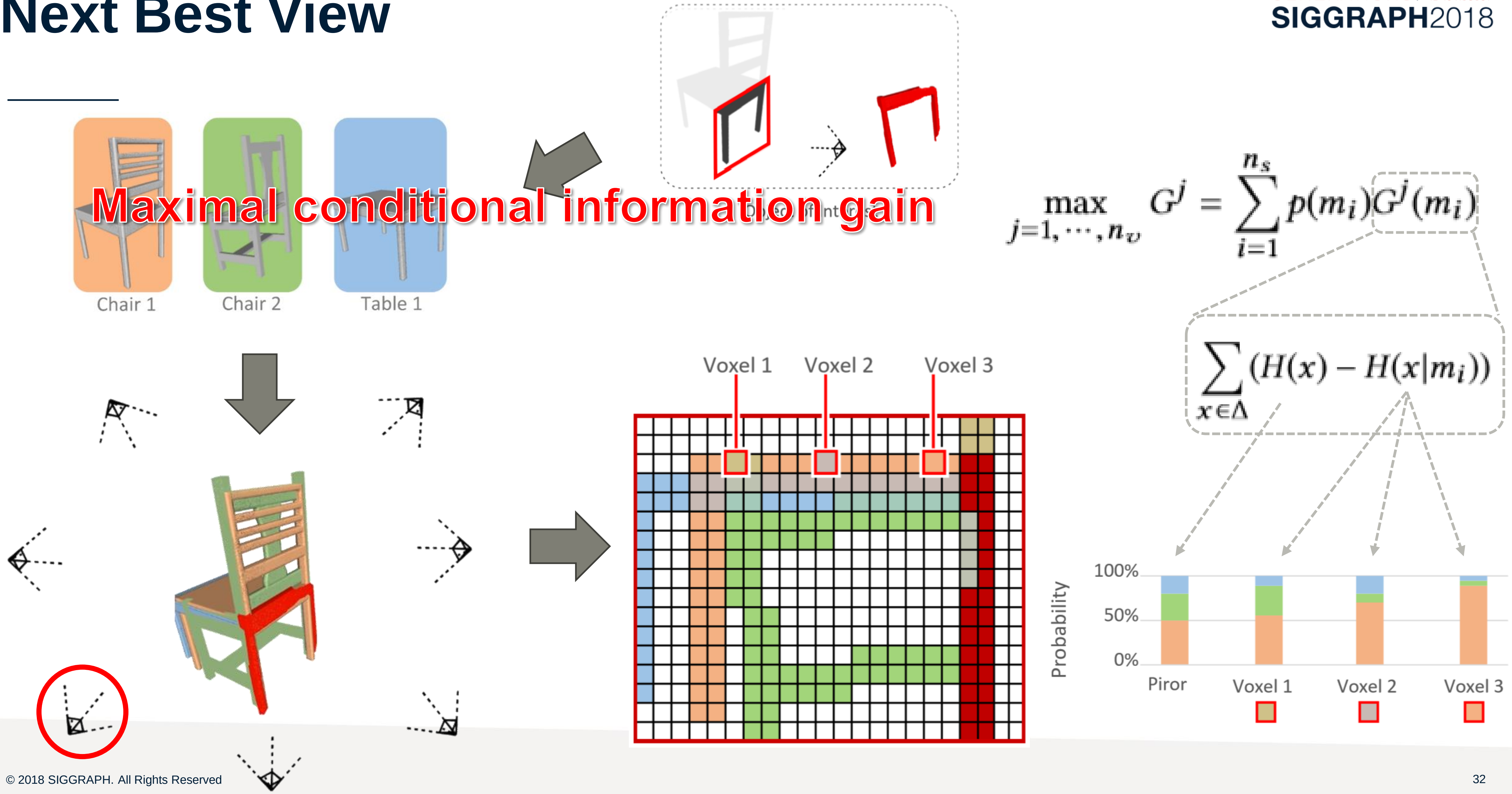
Phase 2: The Next Best View Problem

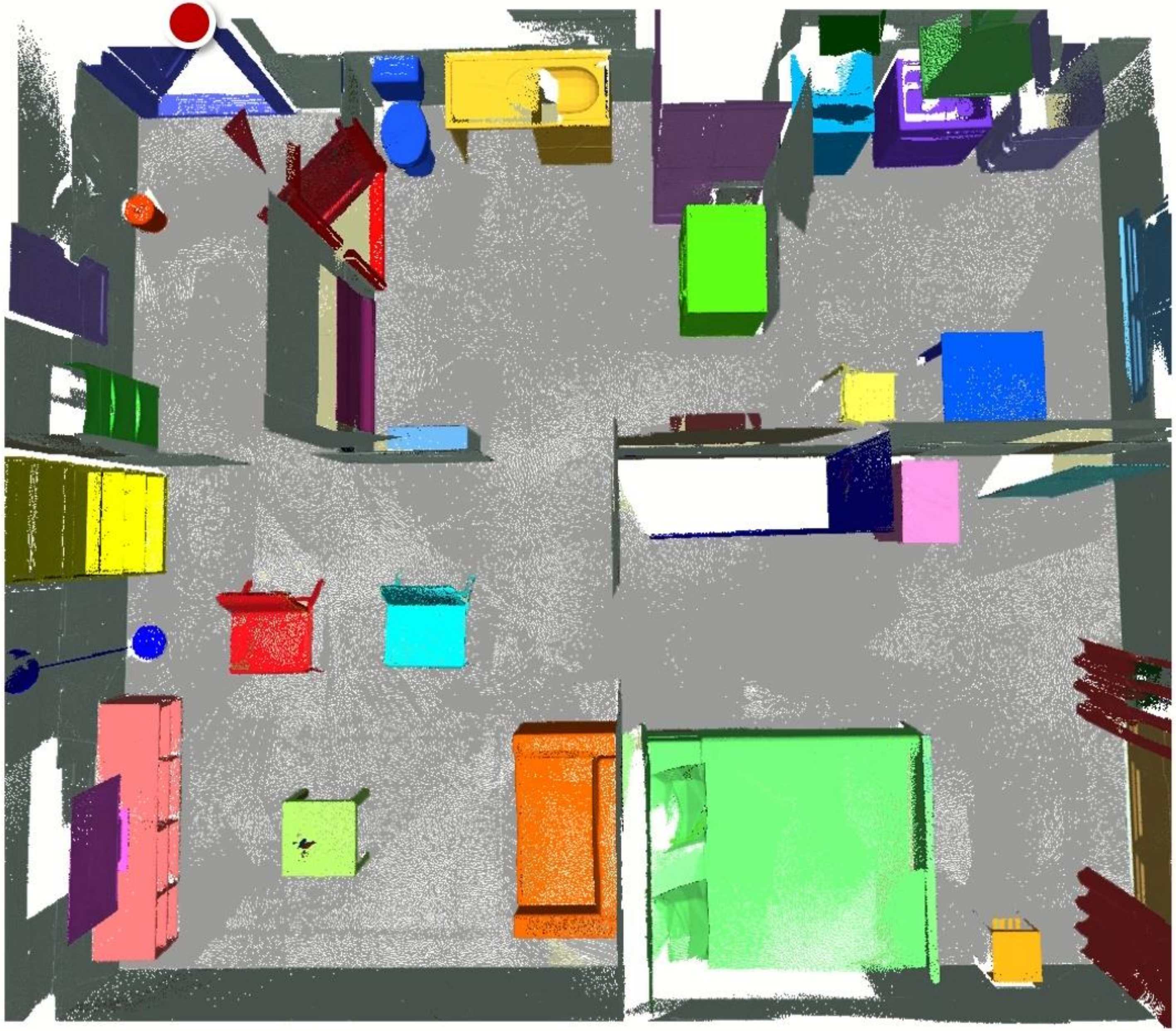




Next Best View

Maximal conditional information gain





Evaluation

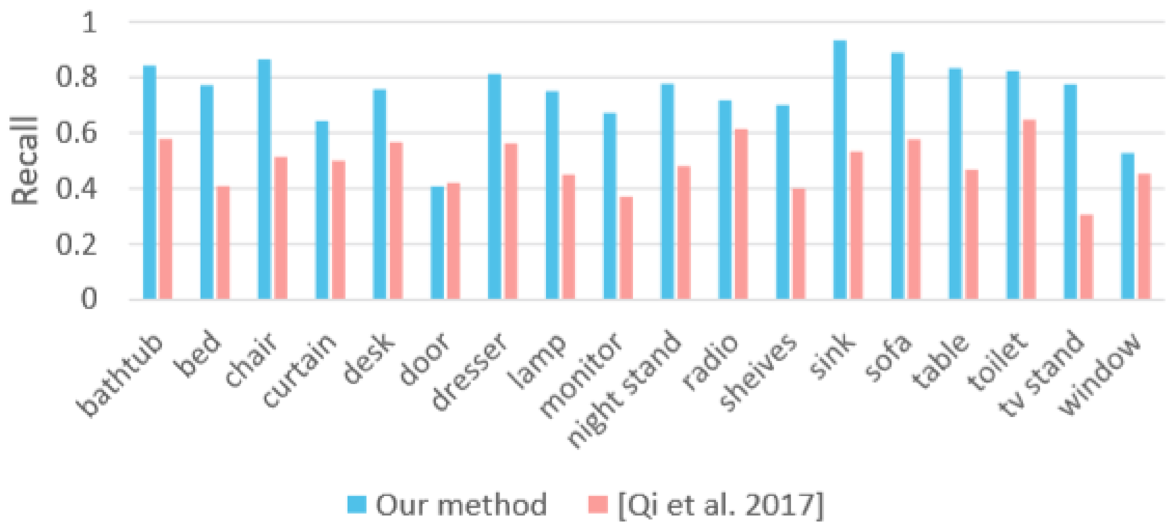
- Virtual scene dataset

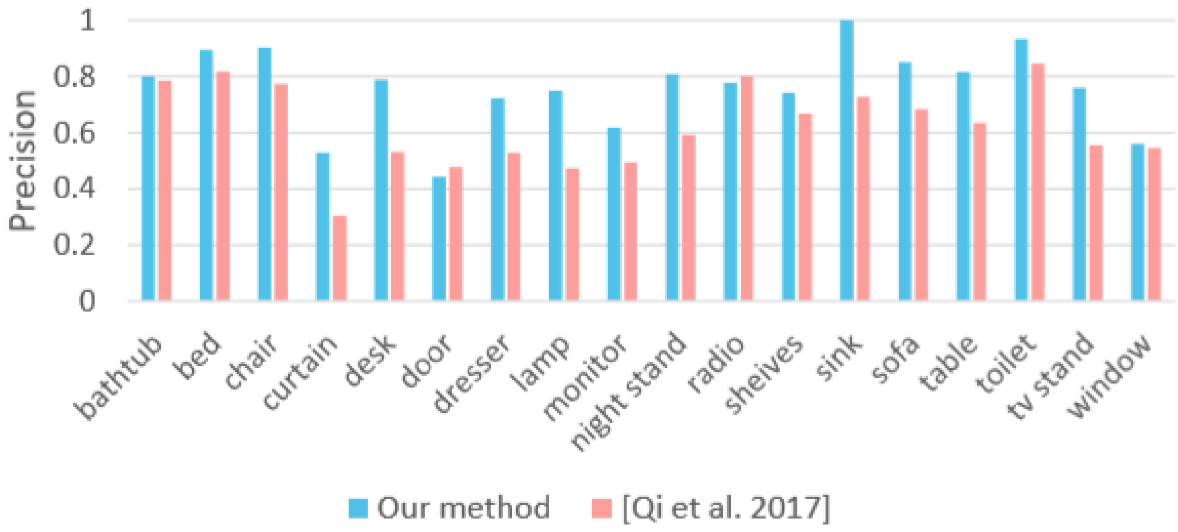
SUNCG (66 scenes)

ScanNet (38 scenes)

Comparison

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

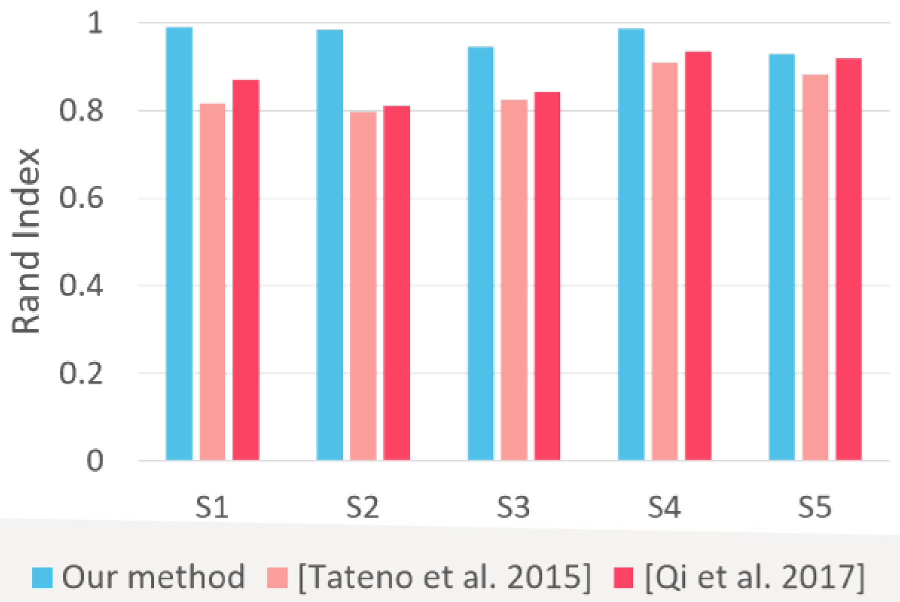


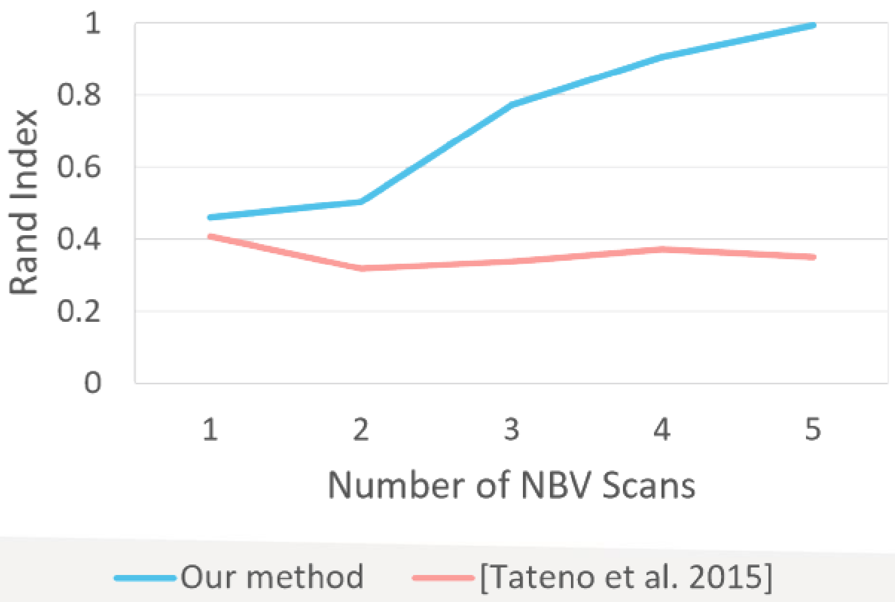


Comparison

- Comparing Rand Index of segmentation

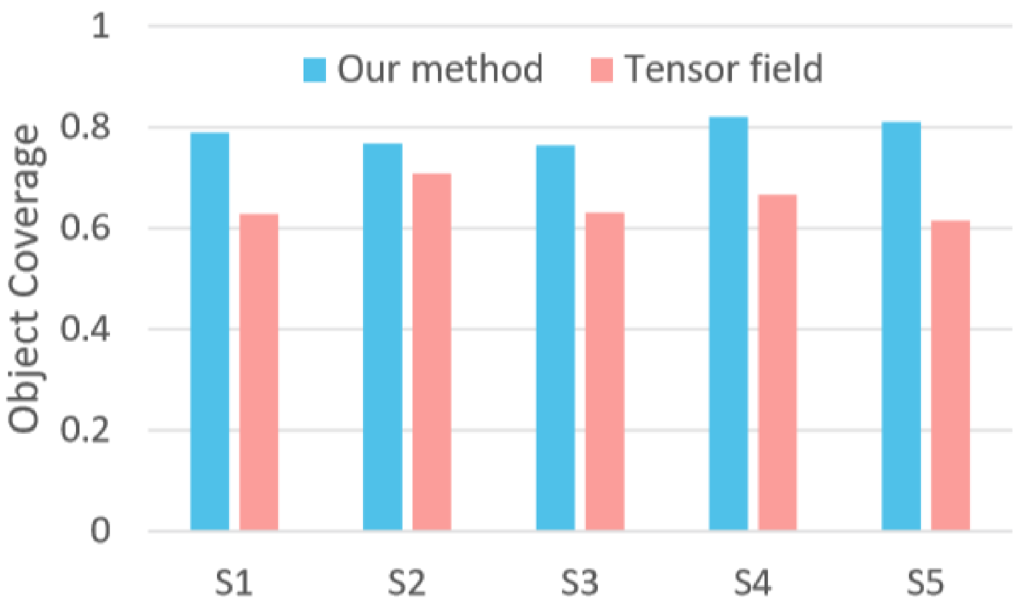
$$RI(S_1, S_2) = \binom{2}{n}^{-1} \sum_{i,j,i<j} [C_{ij}P_{ij} + (1 - C_{ij})(1 - P_{ij})],$$

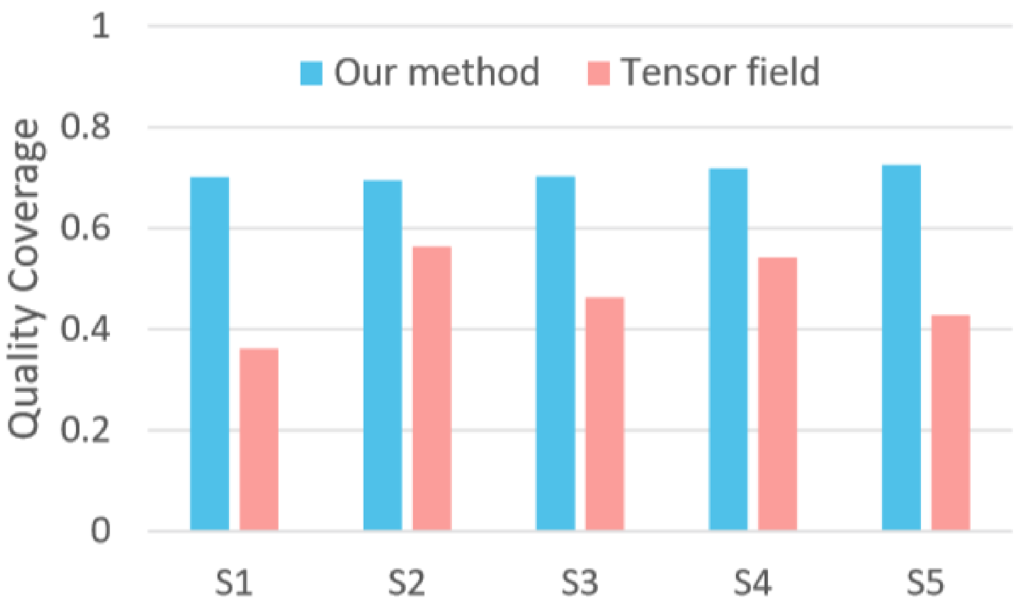




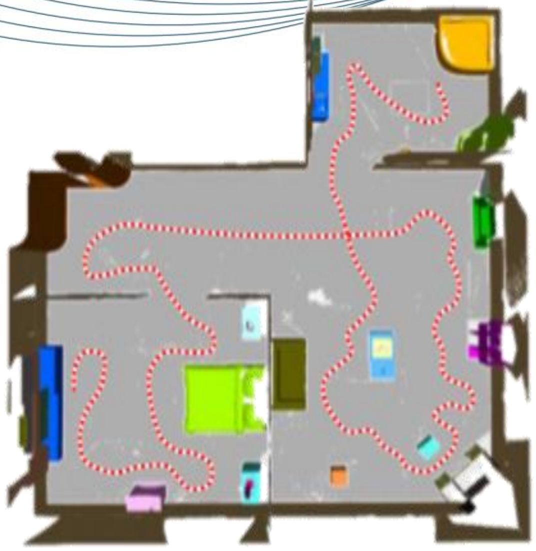
Comparison

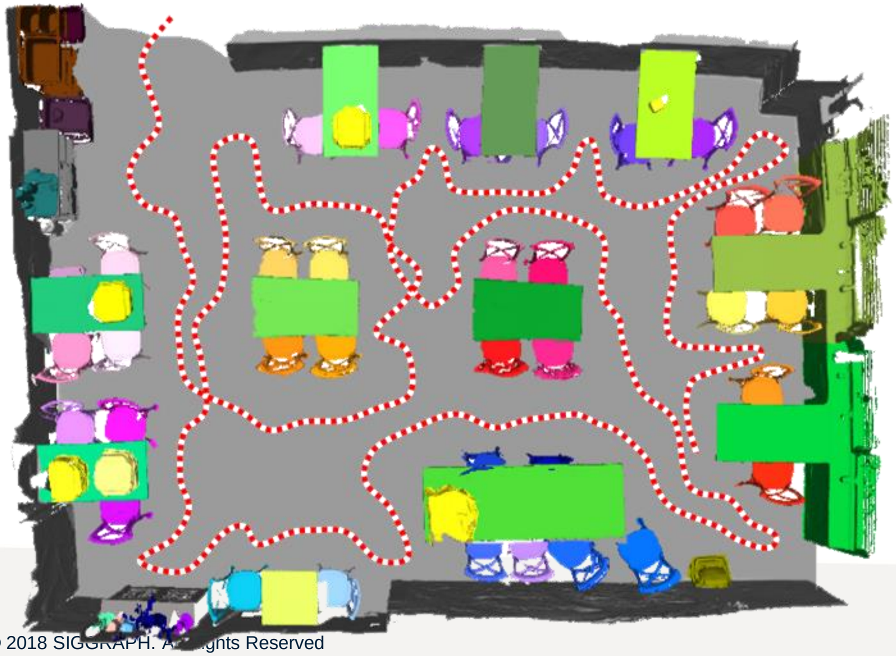
- Comparing object coverage rate and quality against tensor field guided autoscanning [Xu et al. 2017]

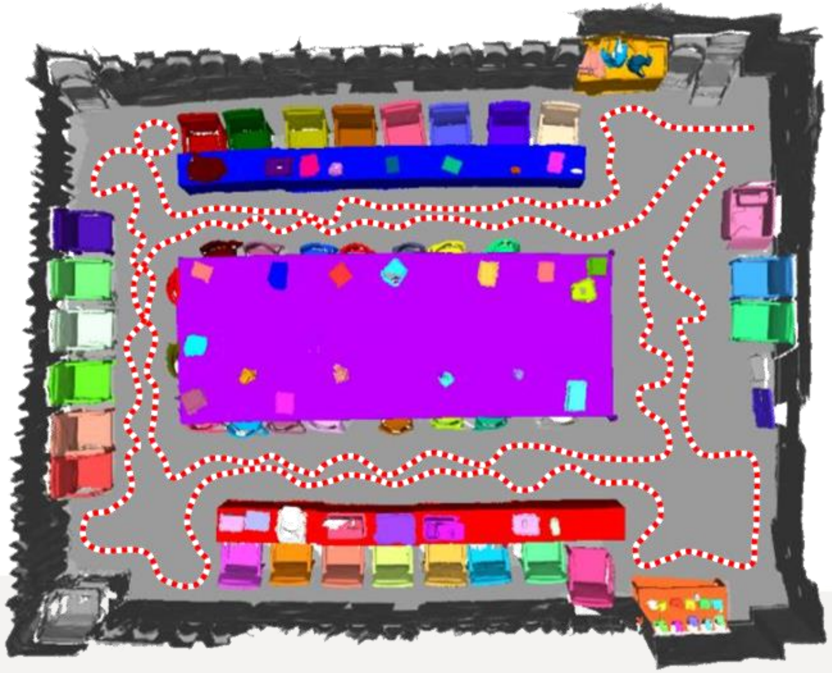




More Results

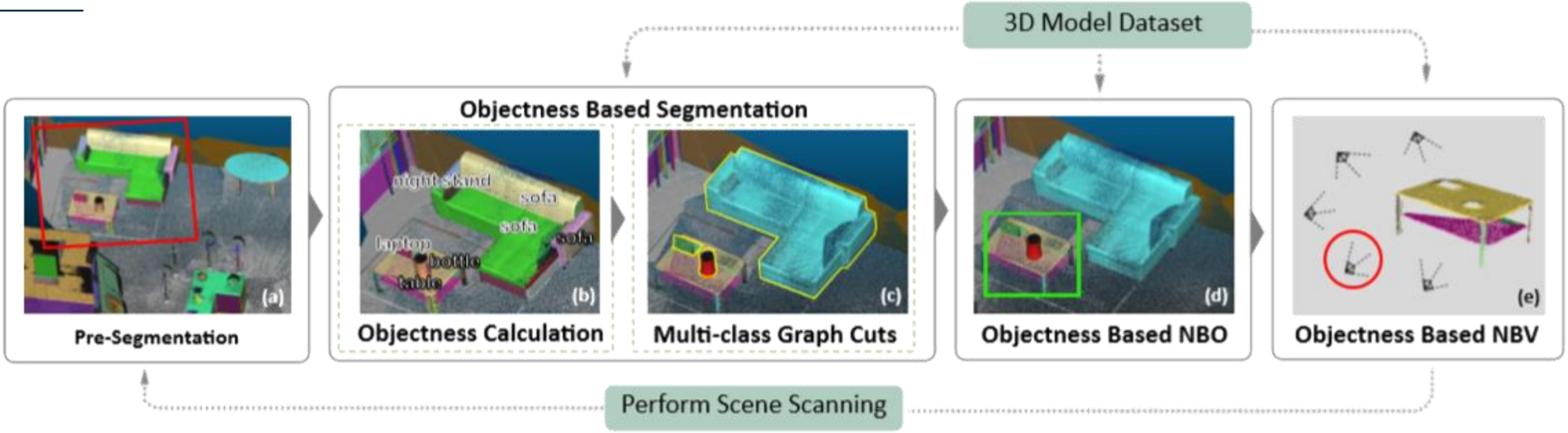






Conclusion

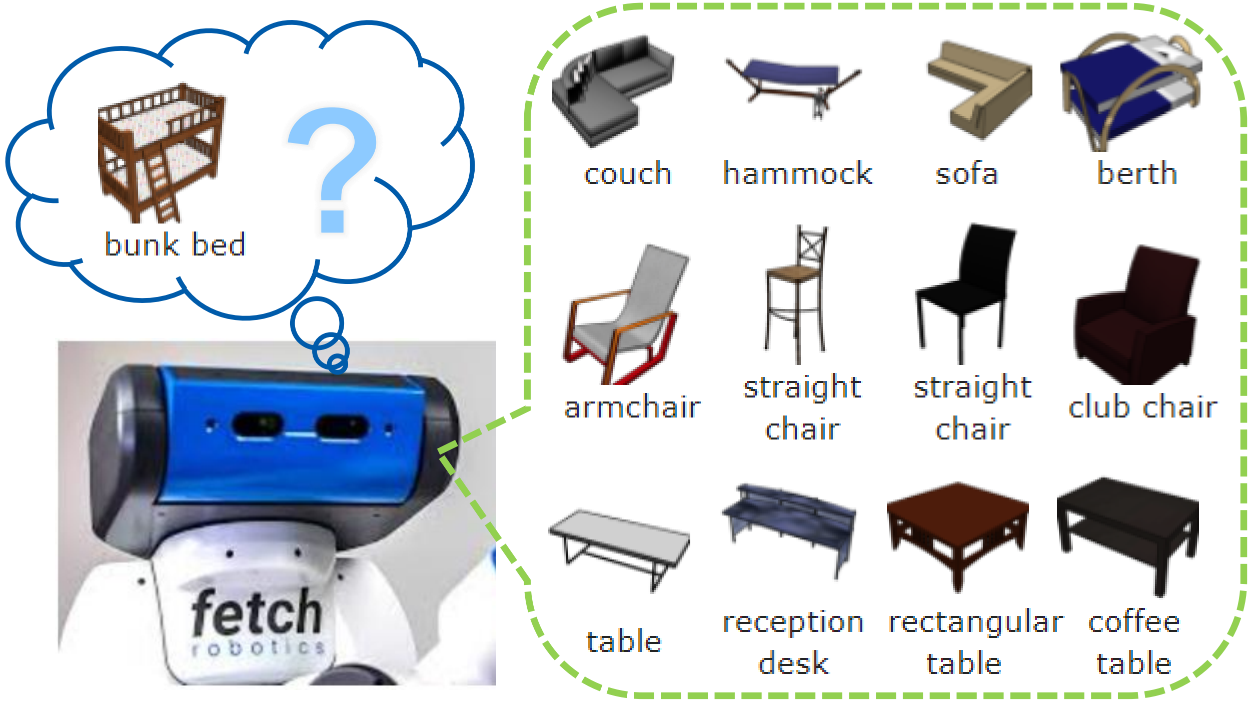
GENERATIONS / VANCOUVER 12-16 AUGUST
SIGGRAPH2018



Key techniques:

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

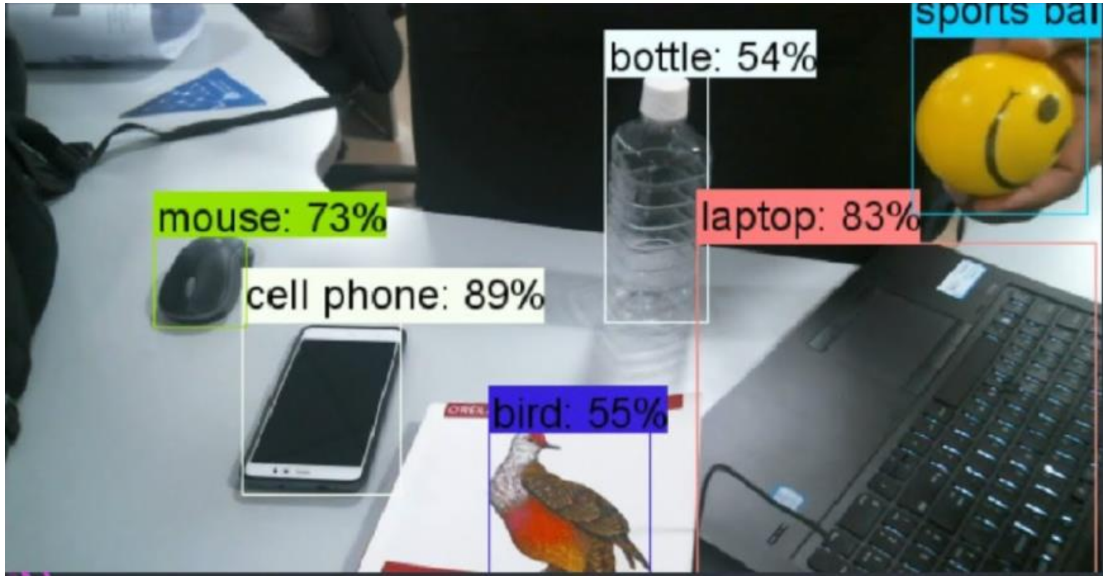
Limitations



No similar models

Cluttered scenes

Future Works



Combine image-based method

Driverless car with LiDAR

Thank you for your attention!

Data and code are available:

http://kevinkaixu.net/projects/nbo.html

Comparison

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

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

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

Depth noise

Time Table

Category	Total	Navigate	Segment	NBO	NBV
Bedroom (V)	47.8	24.1	20.1	2.0	1.6
Living room (V)	57.0	30.4	22.2	2.3	2.1
Kitchen (V)	37.5	16.2	17.6	2.0	1.7
Bathroom (V)	29.5	14.8	12.2	1.3	1.2
Office (V)	40.8	21.3	16.0	1.9	1.6
Meeting room (R)	101.4	62.3	32.4	3.6	3.1
Resting room (R)	78.5	47.9	25.4	2.9	2.3
Office (R)	94.7	56.9	30.3	4.2	3.3

Robot