

Volumetric instance-aware semantic mapping and 3D object discovery



Margarita Grinvald, Fadri Furrer,
Tonci Novkovic, Jen Jen Chung,
Cesar Cadena, Roland Siegwart,
Juan Nieto

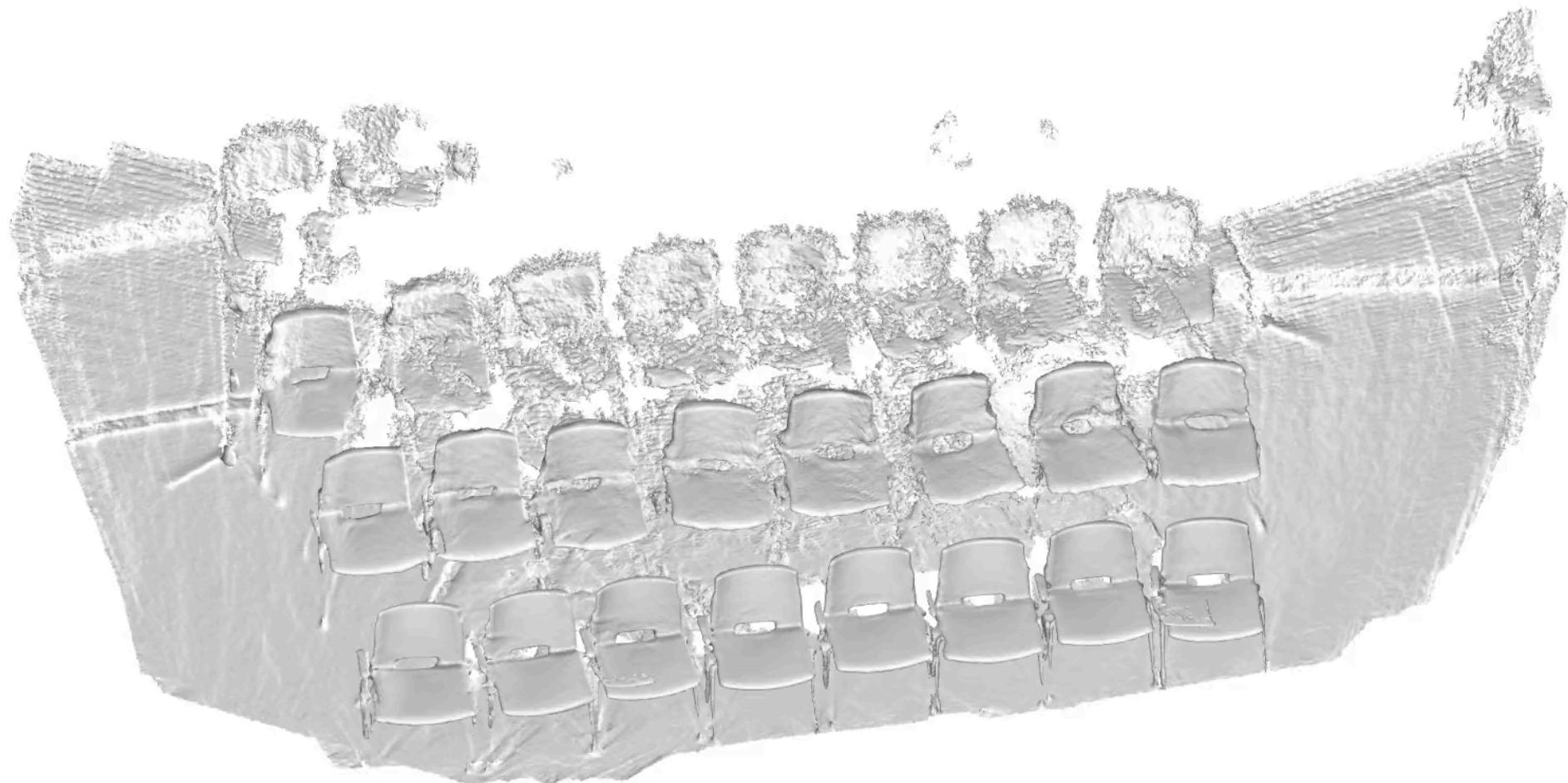
IROS, 5 Nov 2019

ETH zürich

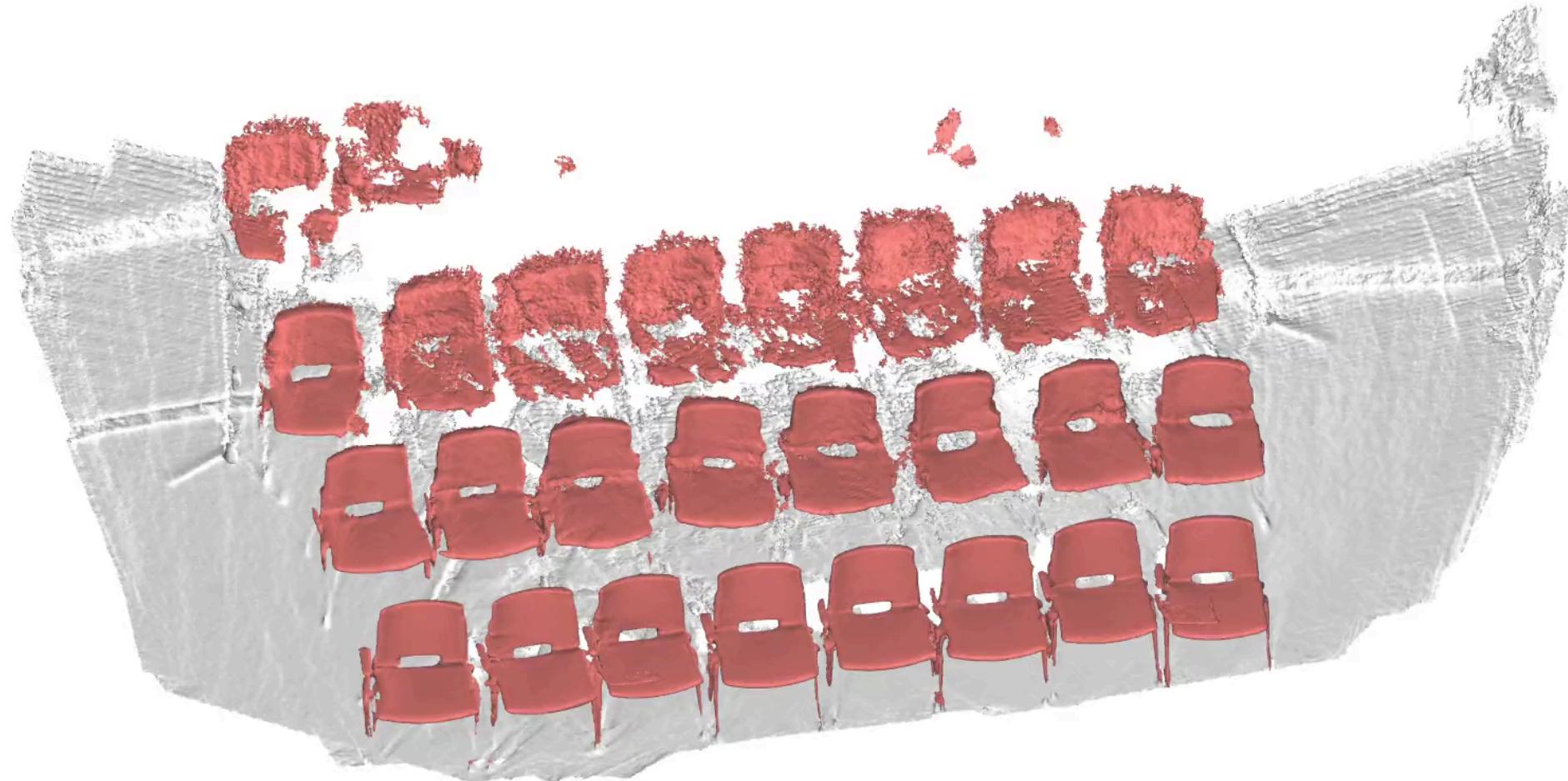




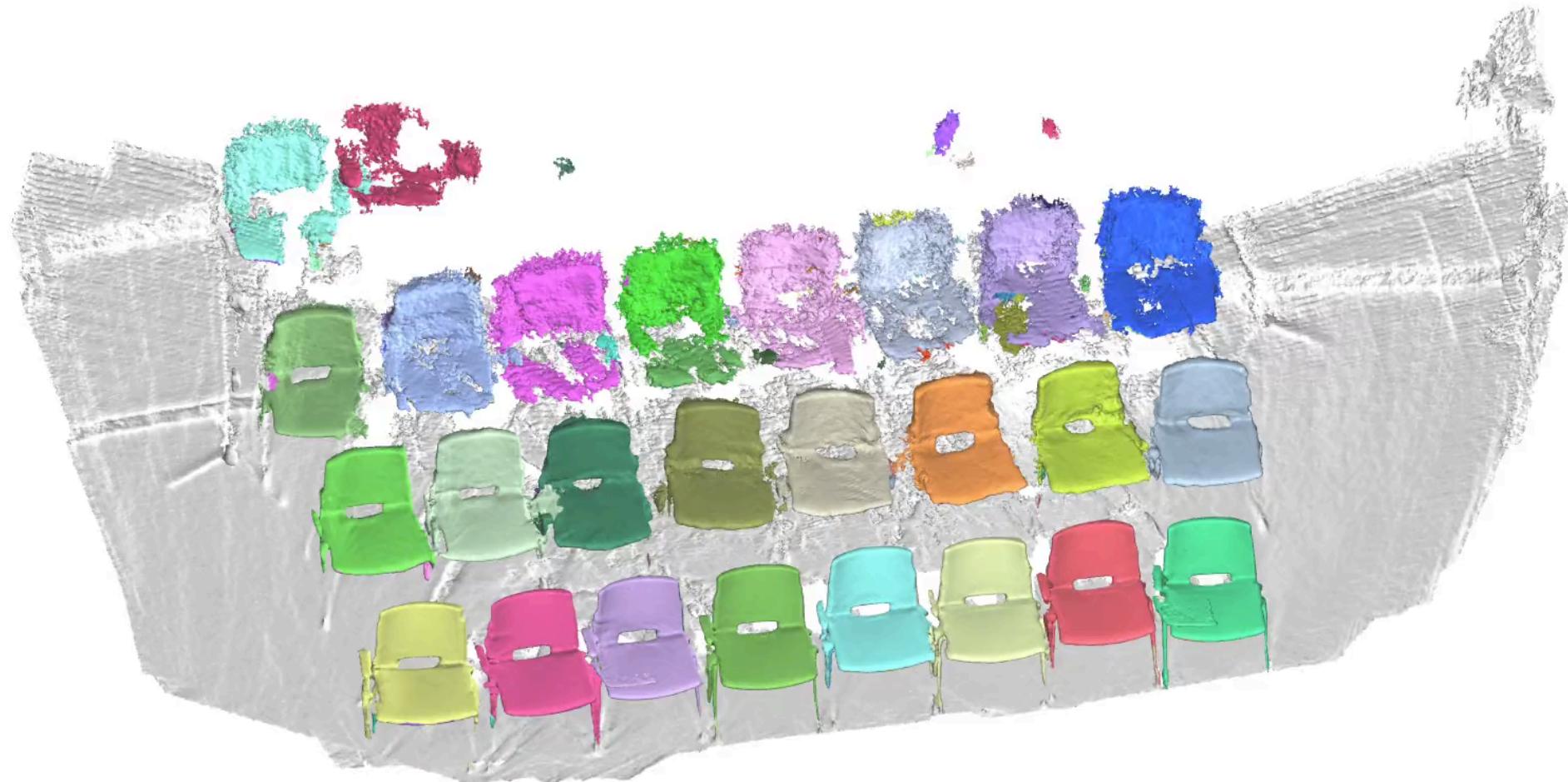
Traditional 3D reconstruction fails to provide any high-level interpretation of the scene



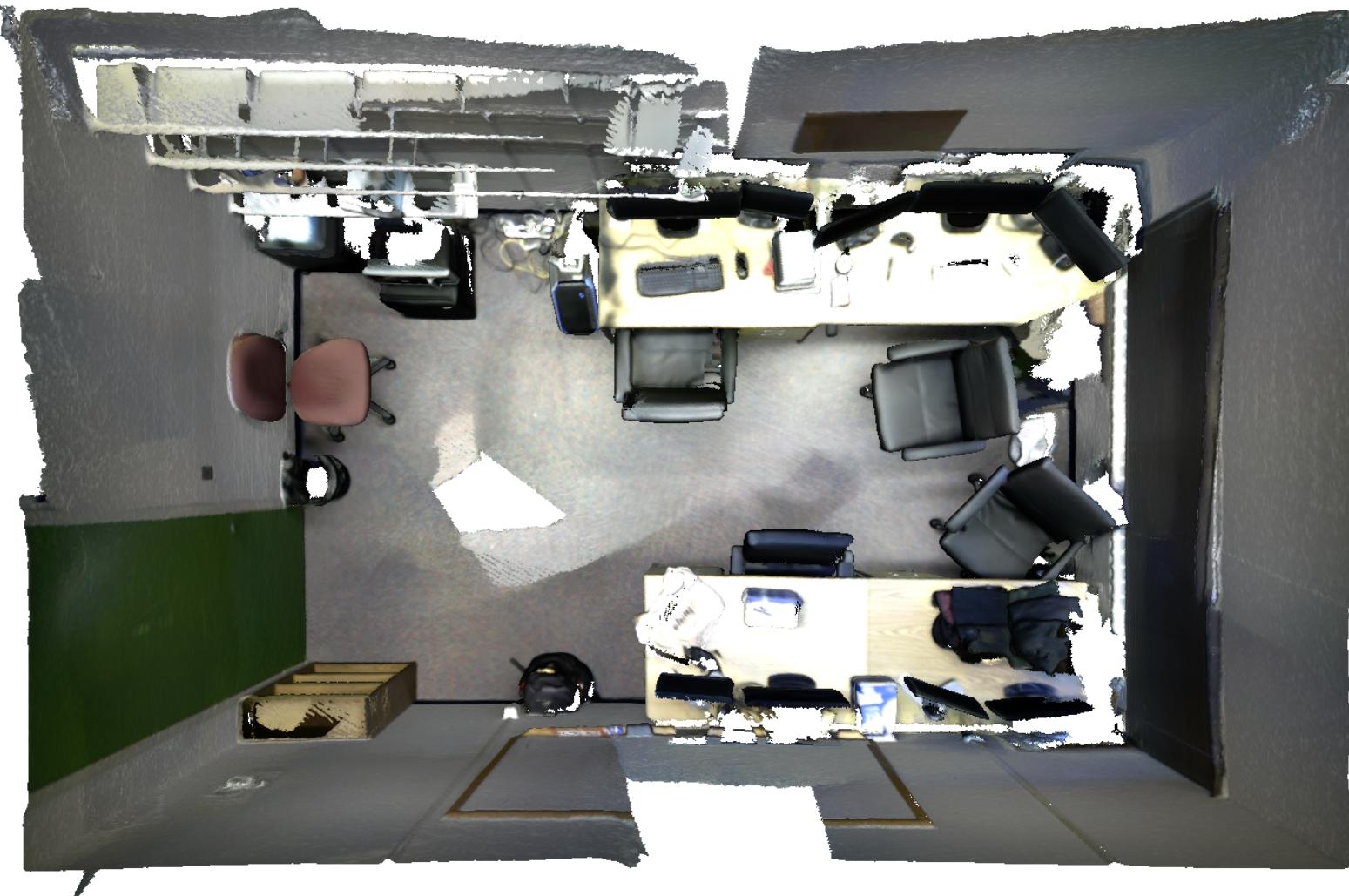
Semantic mapping classifies scene parts by category
but disregards individual object instances



Instance-aware semantic mapping solves detection
and recognition at the level of individual objects



Object-level mapping in the real-world needs to cope with the complexity of an open-set environment



Object-level mapping in the real-world needs to cope with the complexity of an open-set environment



Object-level mapping in the real-world needs to cope with the complexity of an open-set environment



Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup

Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

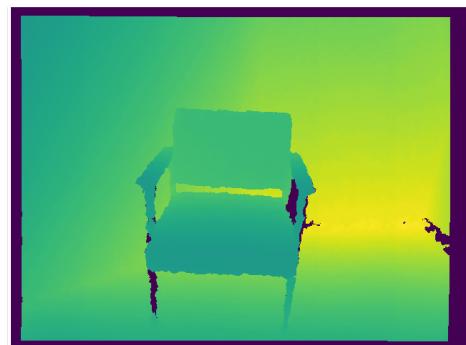
Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup



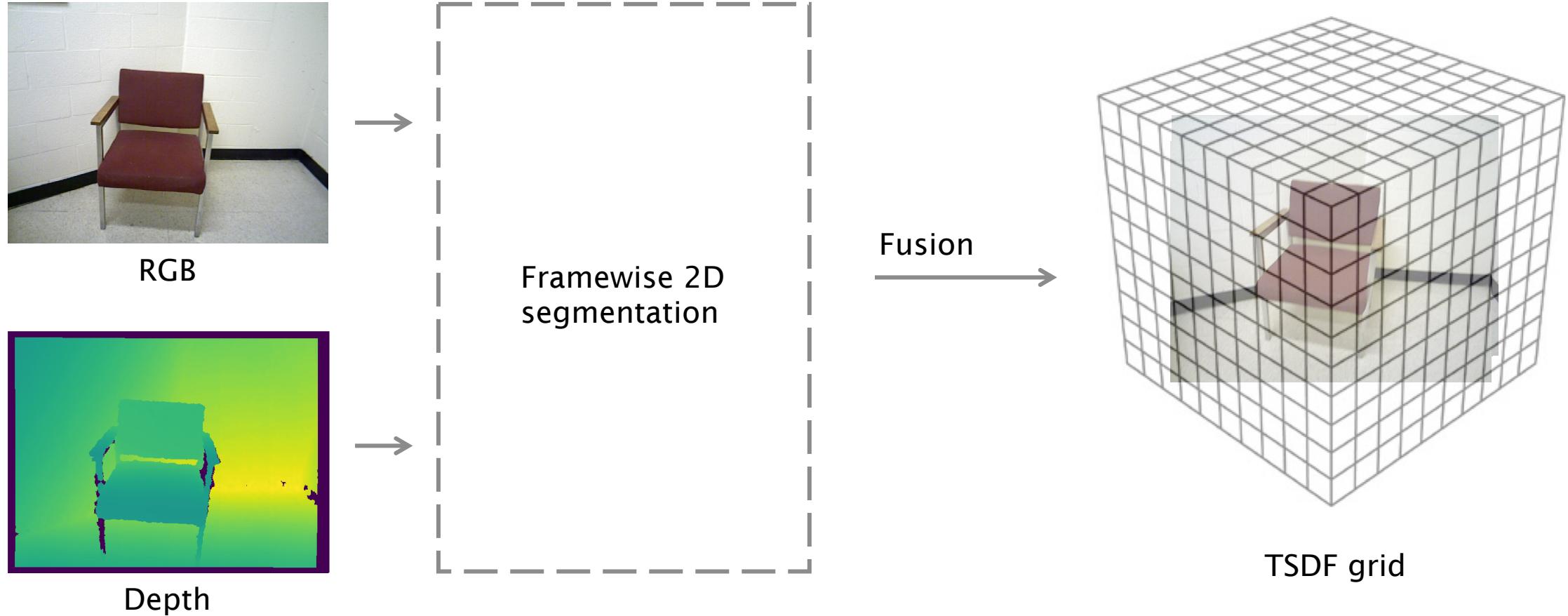


RGB

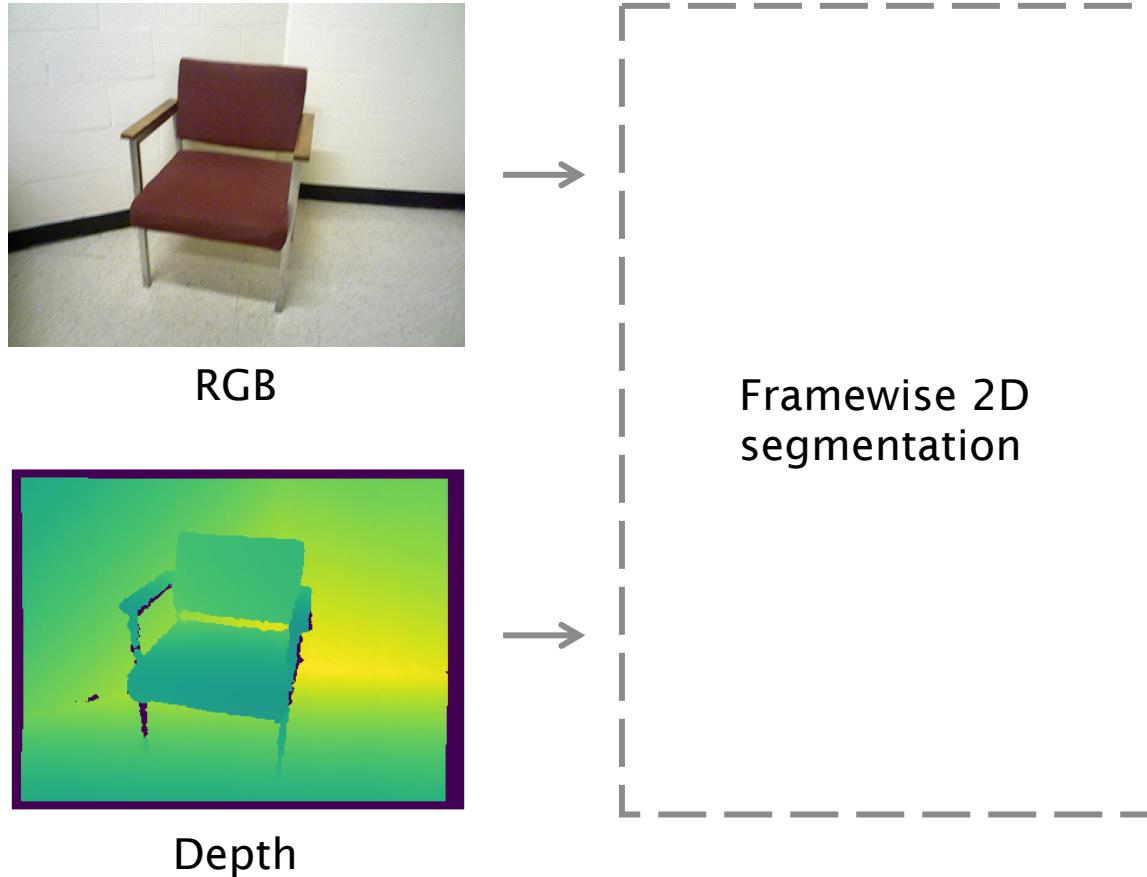


Depth

A dense volumetric object-level map is built online
by incrementally fusing per-frame 2D segmentation



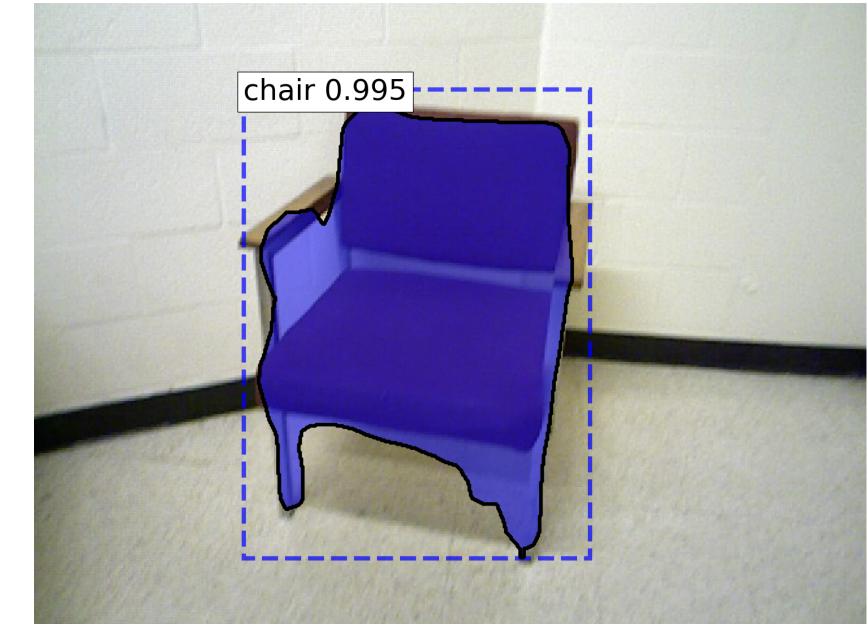
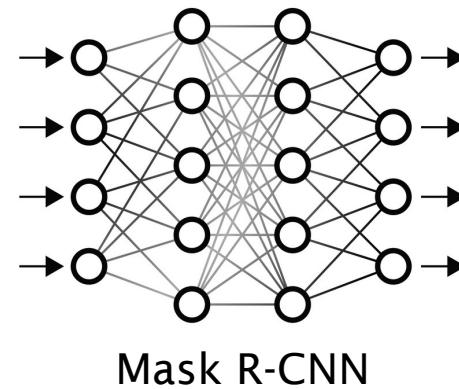
A dense volumetric object-level map is built online
by incrementally fusing per-frame 2D segmentation



A neural network detects recognized objects in the RGB frame and predicts for each a (loose) segmentation mask

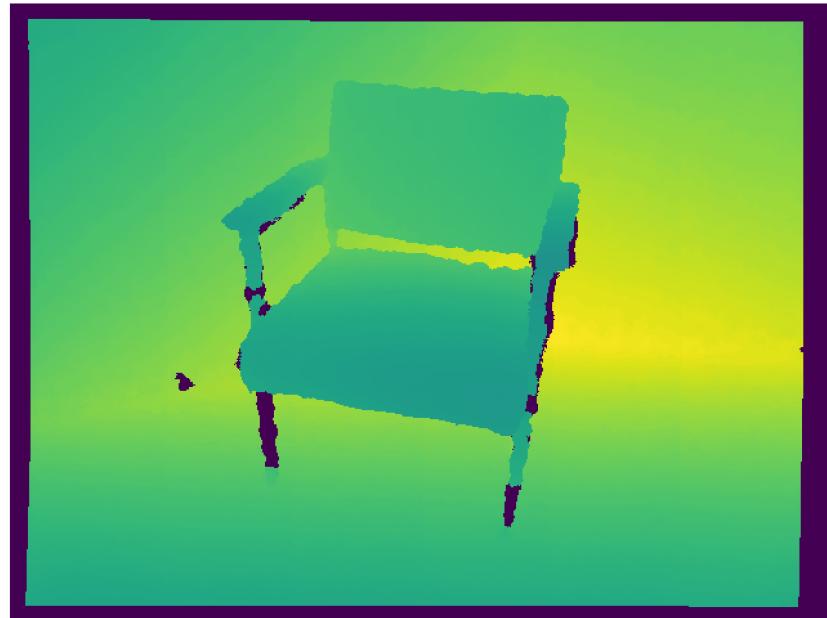


RGB

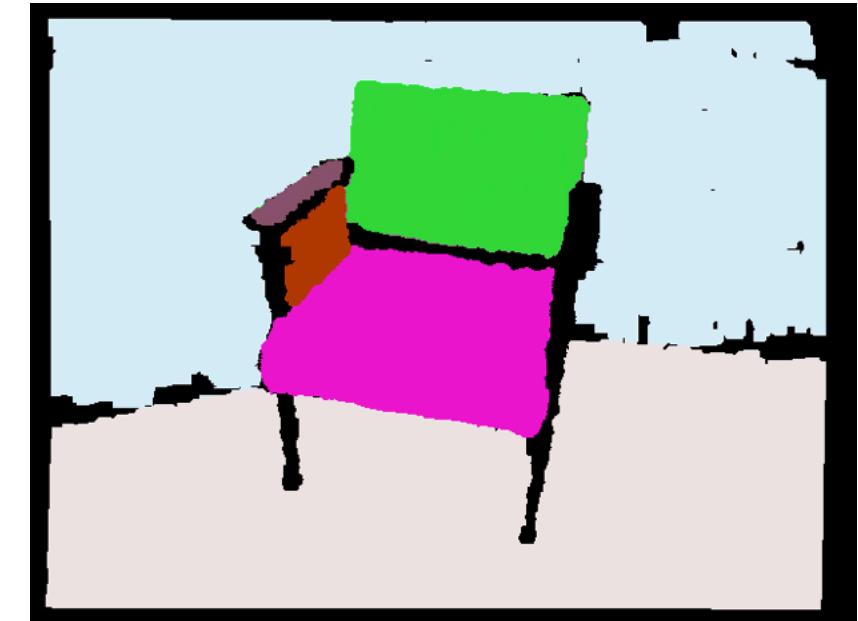
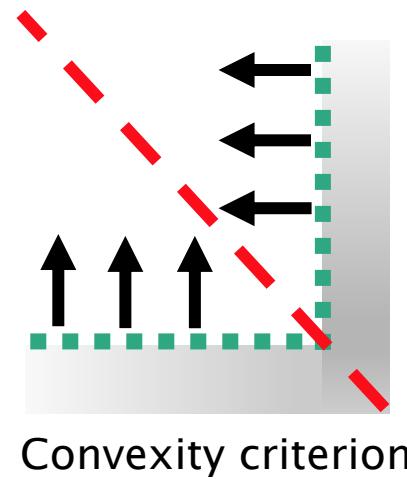


Semantic instance-aware segmentation

An unsupervised geometric method
exhaustively (over)segments the depth frame



Depth

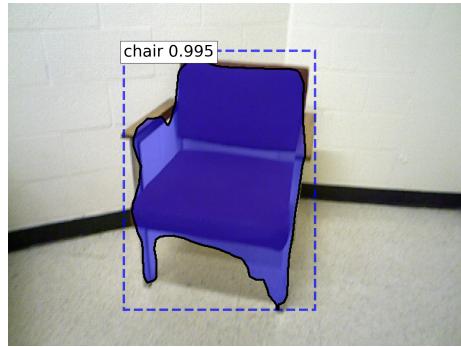


Convexity-based segmentation

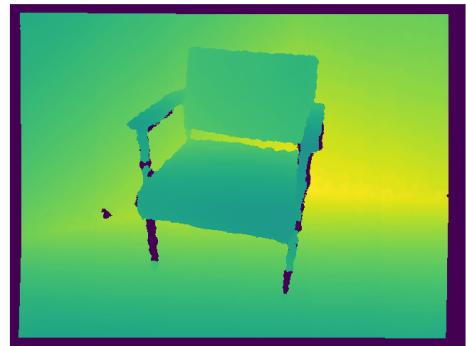
The semantic masks group sets of convex segments as part of the same object instance



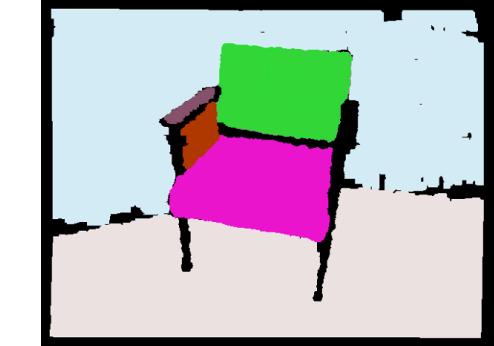
RGB



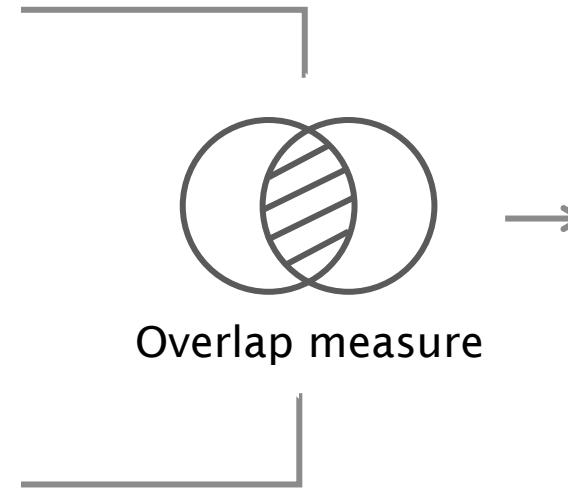
Mask R-CNN



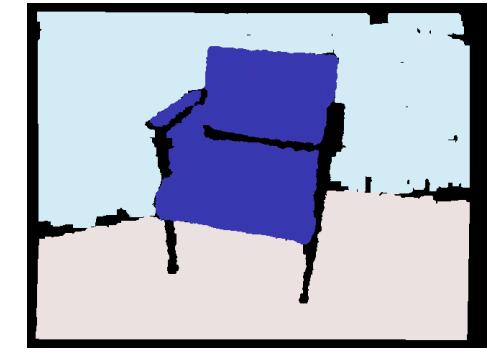
Depth



Geometric segmentation



Overlap measure



Semantically refined geometric segmentation

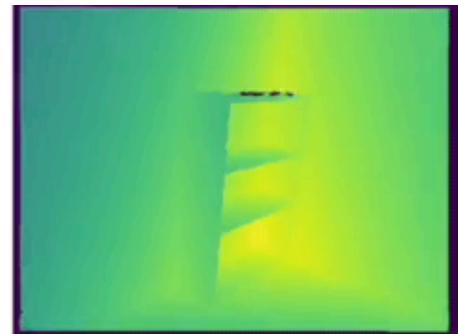
The partial per-frame geometry and segmentation observations are incrementally integrated into a volumetric map



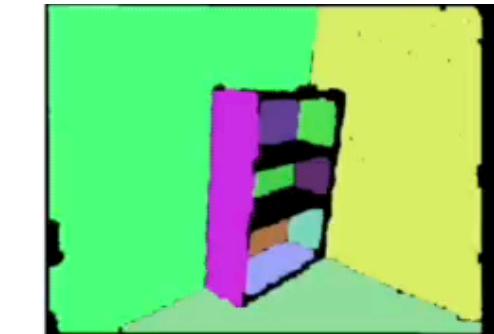
RGB



Mask R-CNN



Depth



Geometric segmentation



*not actual speed

Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup

Volumetric instance-aware semantic mapping and 3D object discovery

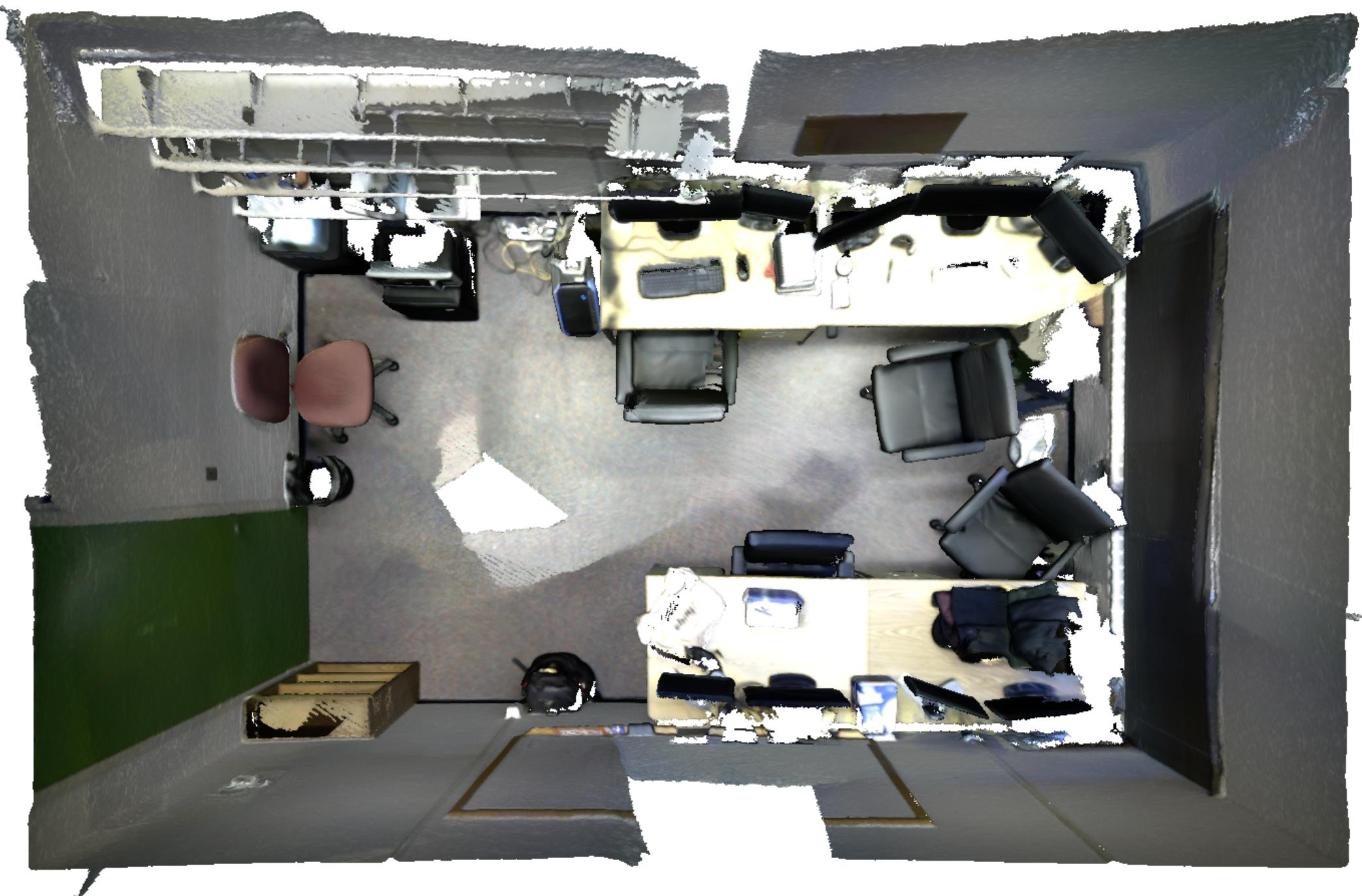


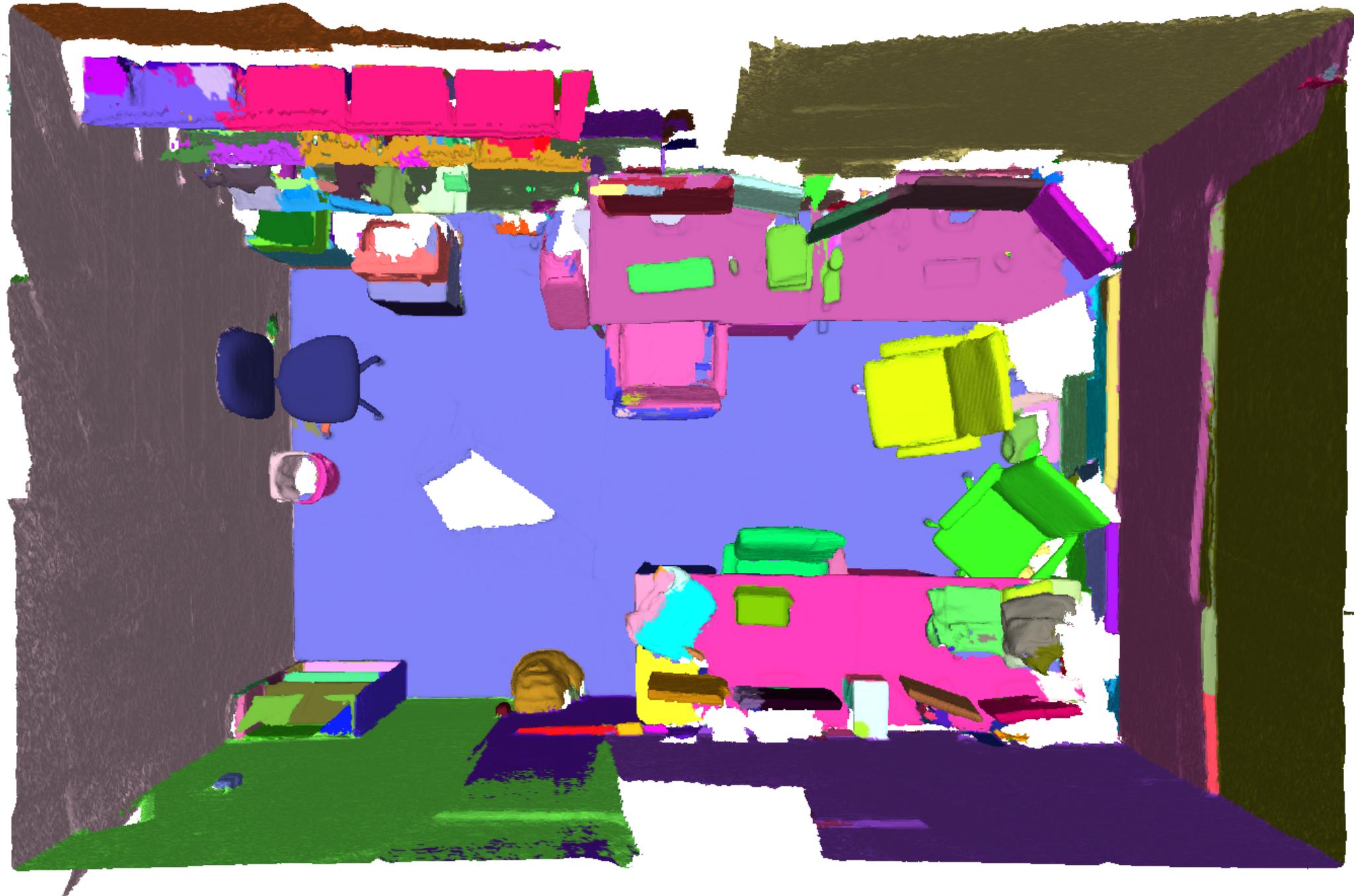
Dense object-level mapping
with a localized RGB-D camera

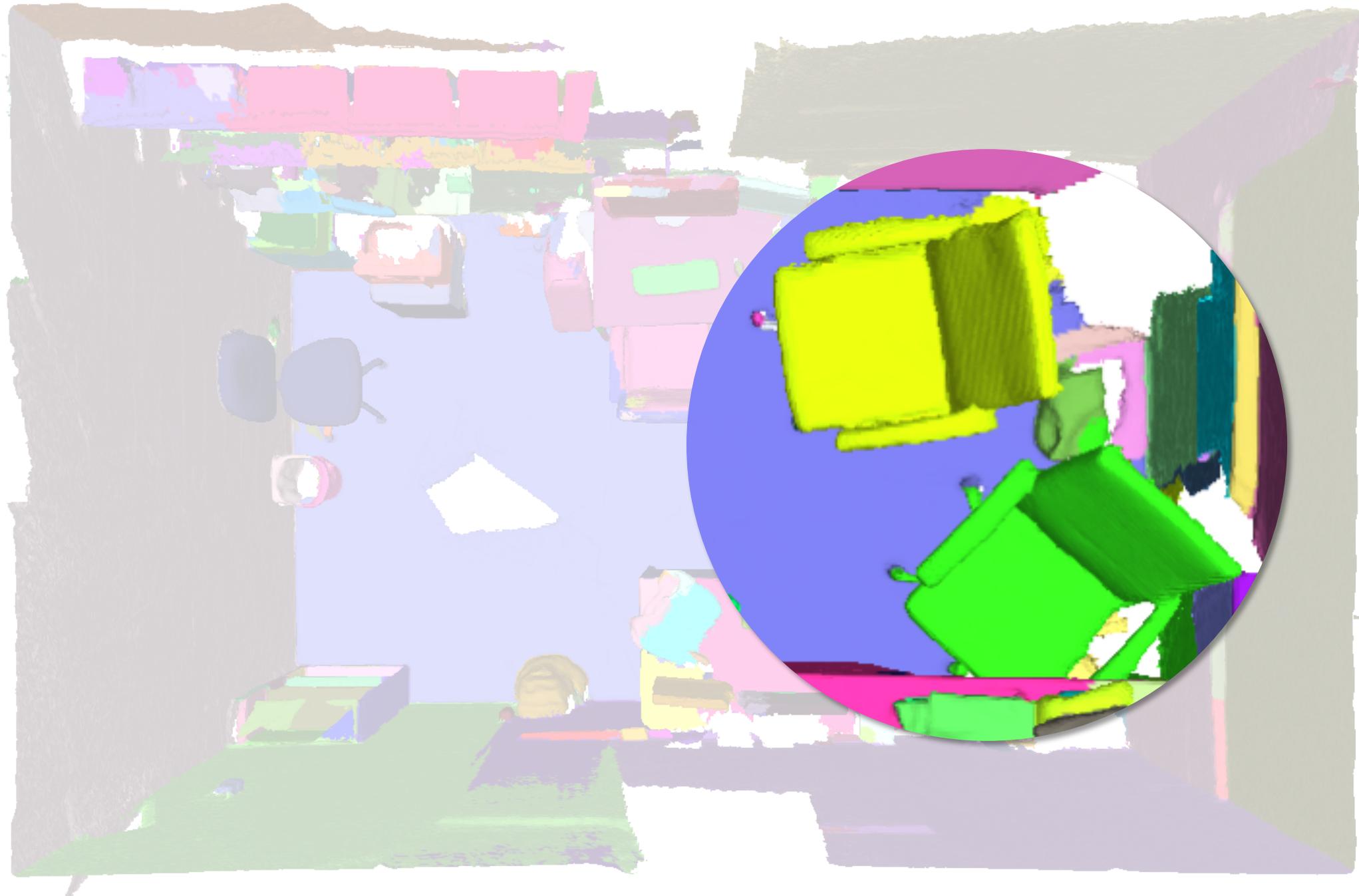
Object detection in an open-set world
by fusing classic and modern computer vision

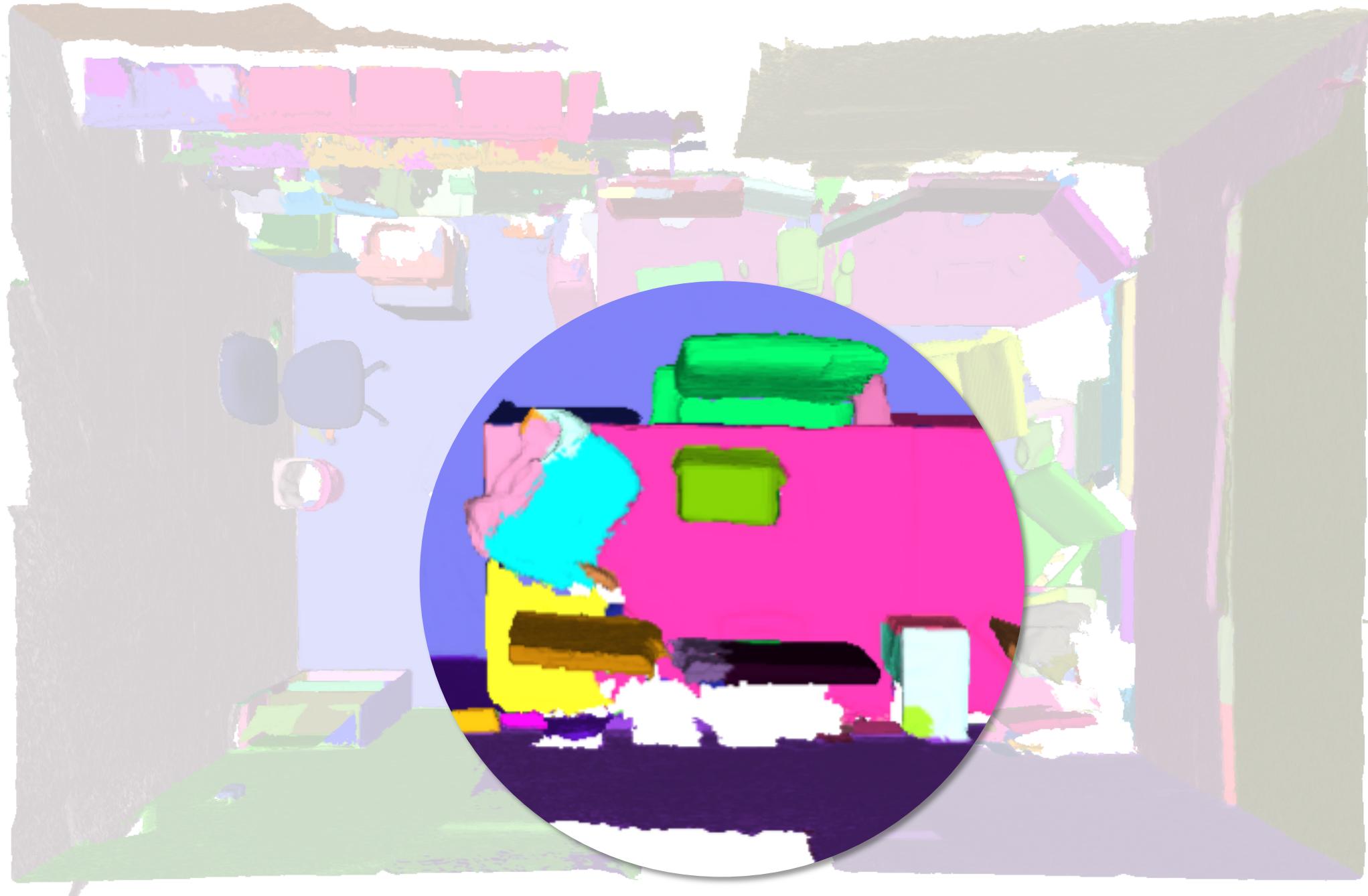
Efficient online framework
well-suited for a real-world robotic setup

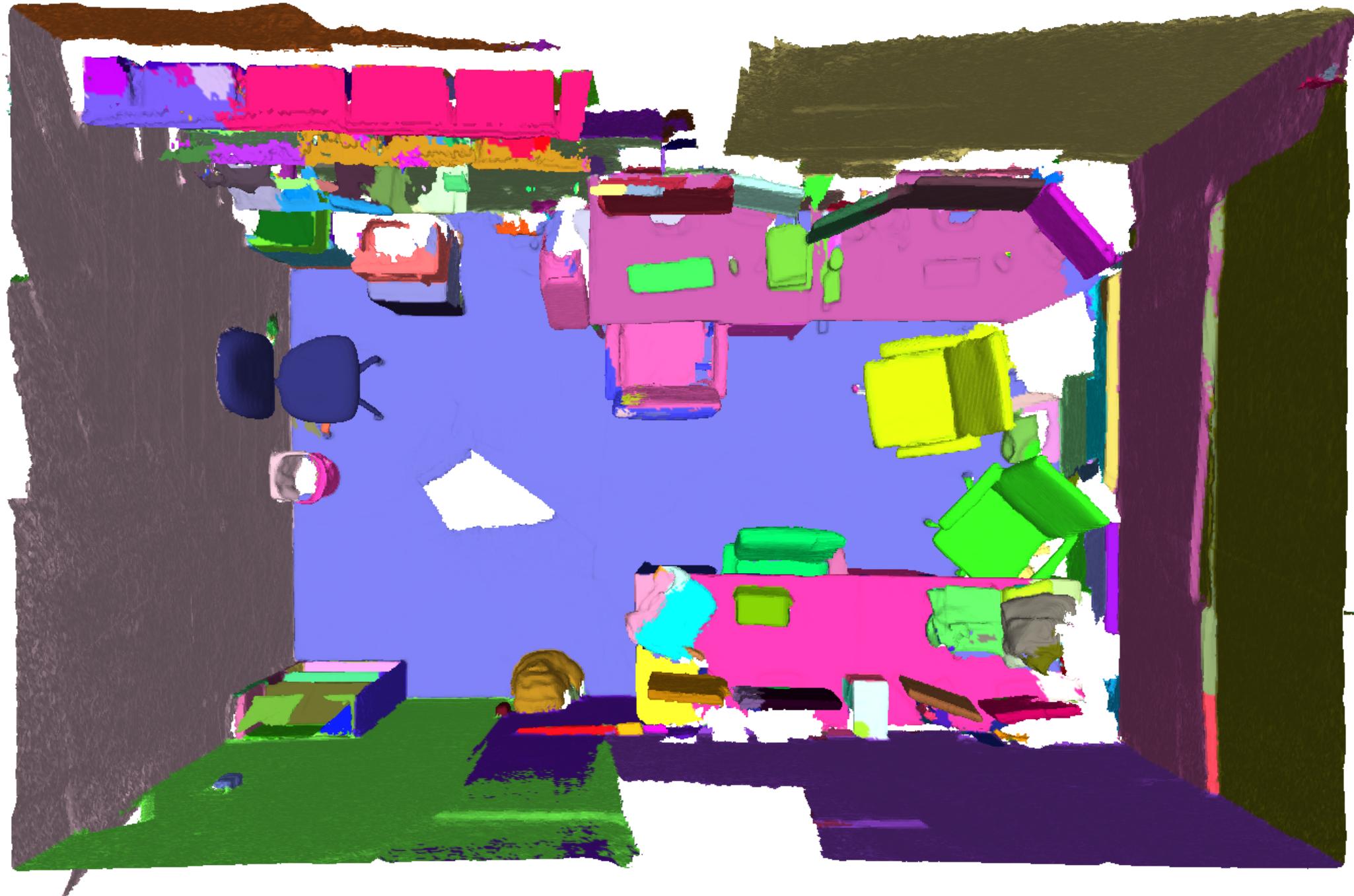
The framework detects both **recognized instances**
and previously **unseen object-like elements**

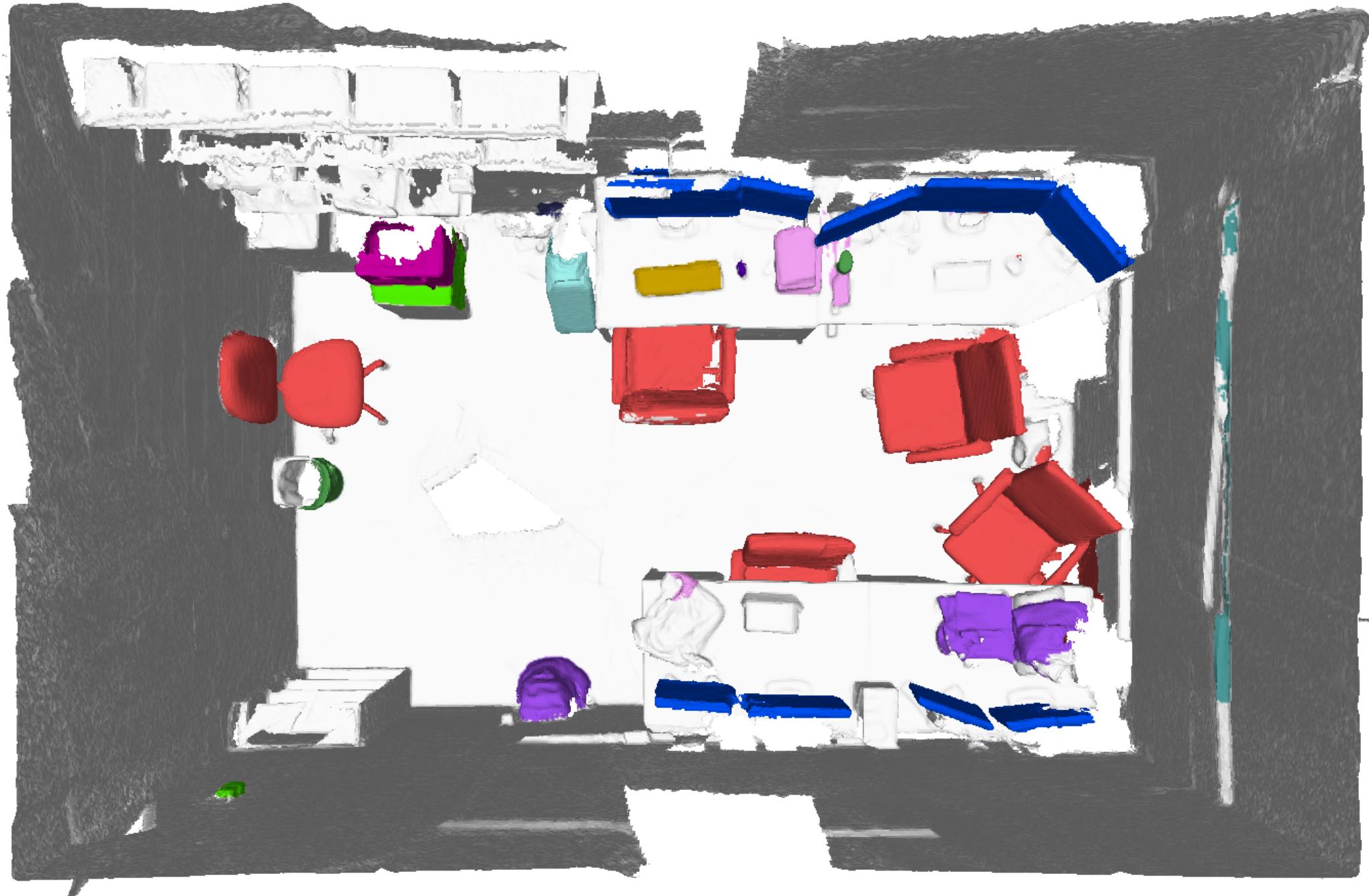


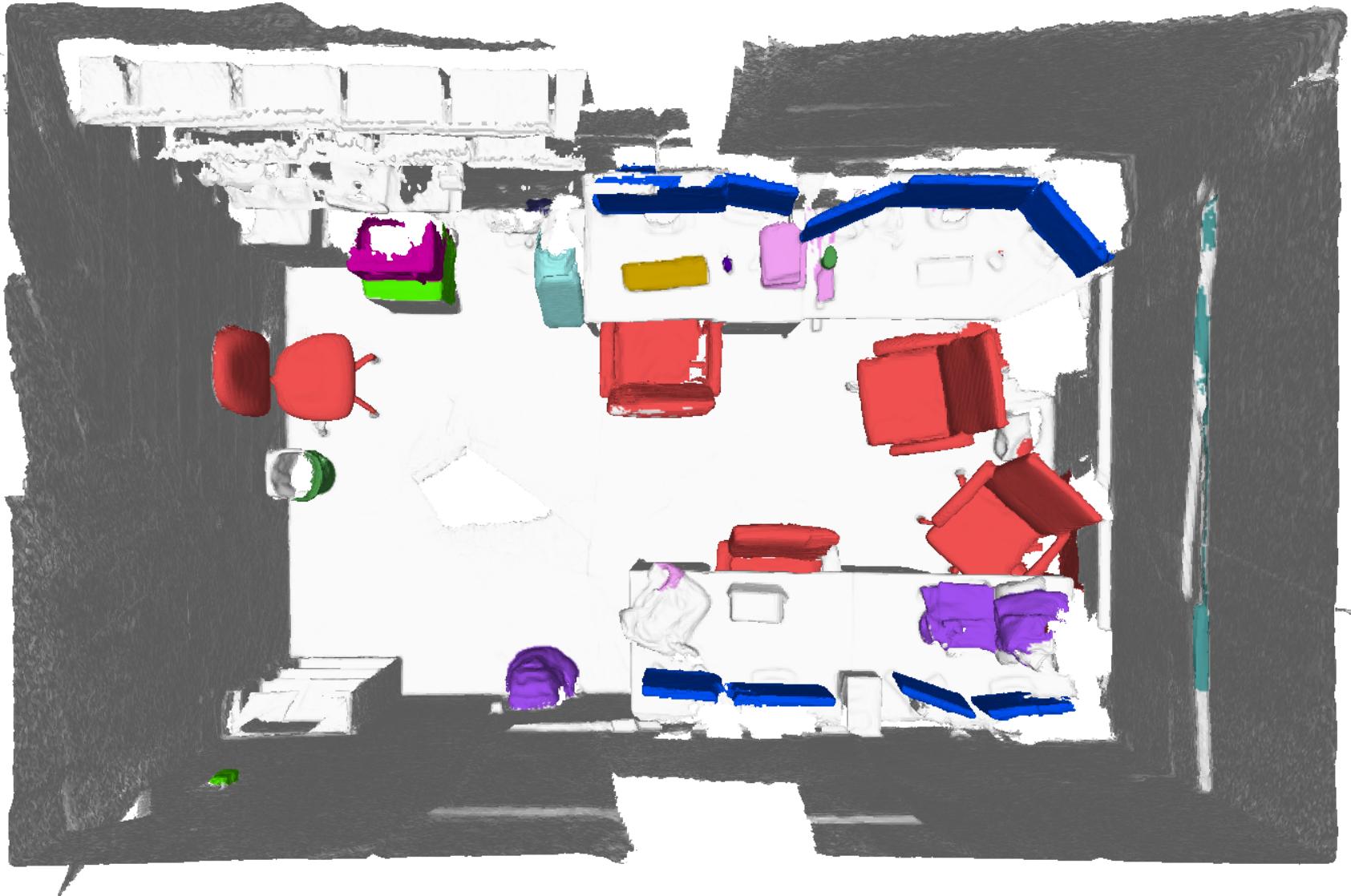












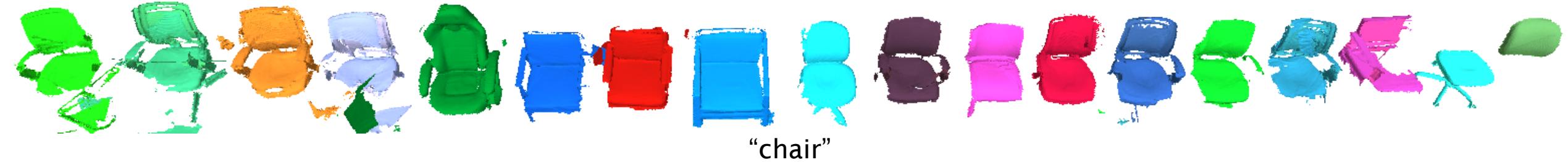
Monitor
Chair
Backpack

Keyboard
Mouse
Cup

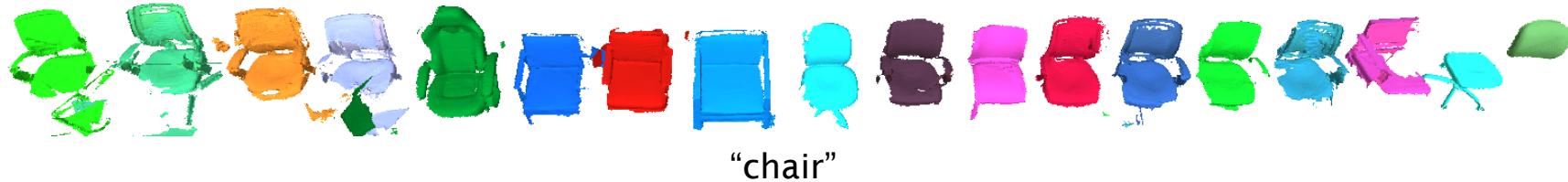
Suitcase
Refrigerator
Microwave

Table
Plant
Unknown

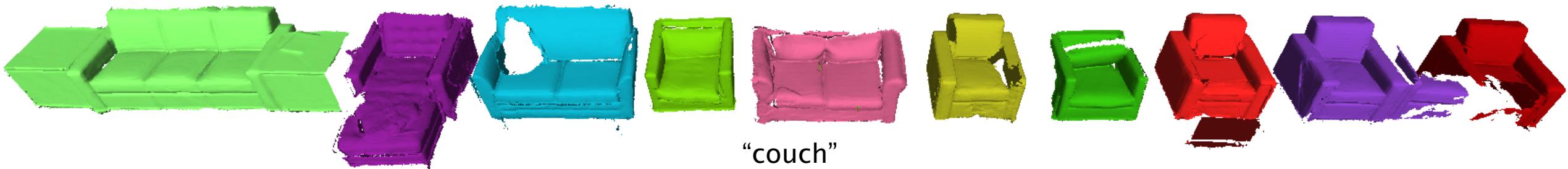
A sample inventory of objects includes recognized instances as well as previously unseen, discovered elements



A sample inventory of objects includes recognized instances as well as previously unseen, discovered elements

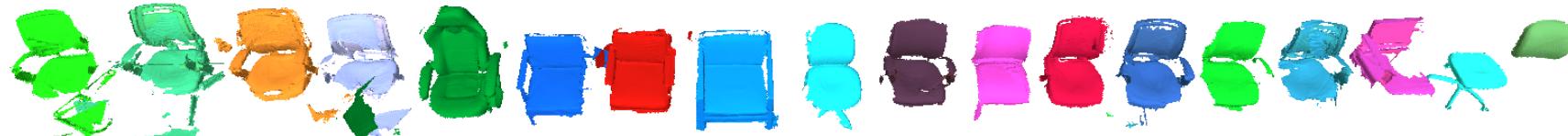


"chair"

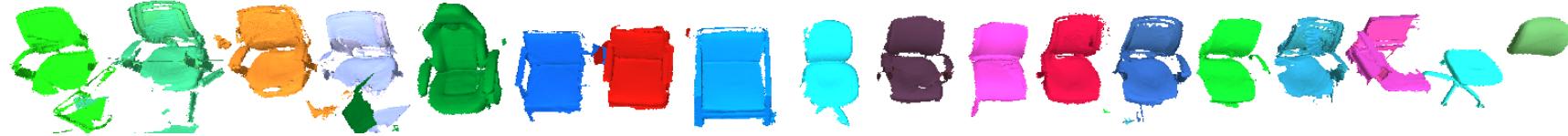


"couch"

A sample inventory of objects includes recognized instances as well as previously unseen, discovered elements



A sample inventory of objects includes recognized instances as well as previously unseen, discovered elements



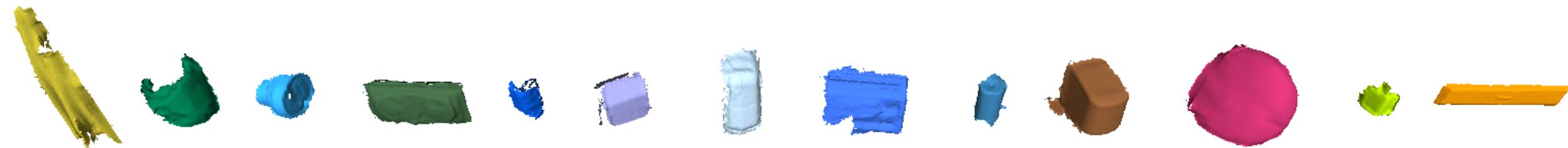
“chair”



“couch”



“table”



[jacket] [bag] [fan] [fan] [speaker] [box] [case] [heater] [paper roll] [appliance] [pillow] [tissues] [drawer]

Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup

Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup

The framework has been validated within a real-world setup



The object-level map of an office floor is built in an online fashion

Front-facing camera



RGB



Mask R-CNN



Depth segmentation

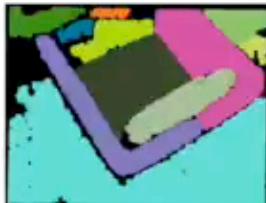
Downward-facing camera



RGB



Mask R-CNN



Depth segmentation



10x speed

The final map densely describes individual scene objects without introducing a significant memory overhead



Volumetric instance-aware semantic mapping and 3D object discovery



Dense object-level mapping
with a localized RGB-D camera

Object detection in an open-set world
by fusing classic and modern computer vision

Efficient online framework
well-suited for a real-world robotic setup

arxiv.org

1

Volumetric Instance-Aware Semantic Mapping and 3D Object Discovery

Margarita Grinvald, Fadri Furrer, Tonci Novkovic, Jen Jen Chung, Cesar Cadena, Roland Siegwart, Juan Nieto

Abstract—To autonomously navigate and plan interactions in real-world environments, robots require the ability to robustly perceive and map complex, unstructured surrounding scenes. Besides building an internal representation of the observed scene geometry, the key insight toward a truly functional understanding of the environment is the usage of higher-level entities during mapping, such as individual object instances. This work presents an approach to incrementally build volumetric object-centric maps during online scanning with a localized RGB-D camera. First, a per-frame segmentation scheme combines an unsupervised geometric approach with instance-aware semantic predictions to detect both recognized scene elements as well as previously unseen objects. Next, a data association step tracks the predicted instances across the different frames. Finally, a map integration strategy fuses information about their 3D shape, location, and, if available, semantic class into a global volume. Evaluation on a publicly available dataset shows that the proposed approach for building instance-level semantic maps is competitive with state-of-the-art methods, while additionally able to discover objects of unseen categories. The system is further evaluated within a real-world robotic mapping setup, for which

The figure displays four 3D point clouds arranged in a 2x2 grid, illustrating the results of semantic mapping. The top row shows a 'Object-centric Map' (left) and a 'Ground Truth Instance Map' (right). The bottom row shows another 'Object-centric Map' (left) and a 'Ground Truth Instance Map' (right). Each point cloud represents a room with various objects like chairs, tables, and boxes. Colored segments indicate different semantic instances. Red circles highlight specific objects for comparison between the two types of maps.

(a) Object-centric Map

(b) Ground Truth Instance Map

GitHub

 [ethz-asl / voxblox-plusplus](#)



Volumetric instance-aware semantic mapping and 3D object discovery



Margarita Grinvald, Fadri Furrer,
Tonci Novkovic, Jen Jen Chung,
Cesar Cadena, Roland Siegwart,
Juan Nieto

IROS, 5 Nov 2019

ETH zürich