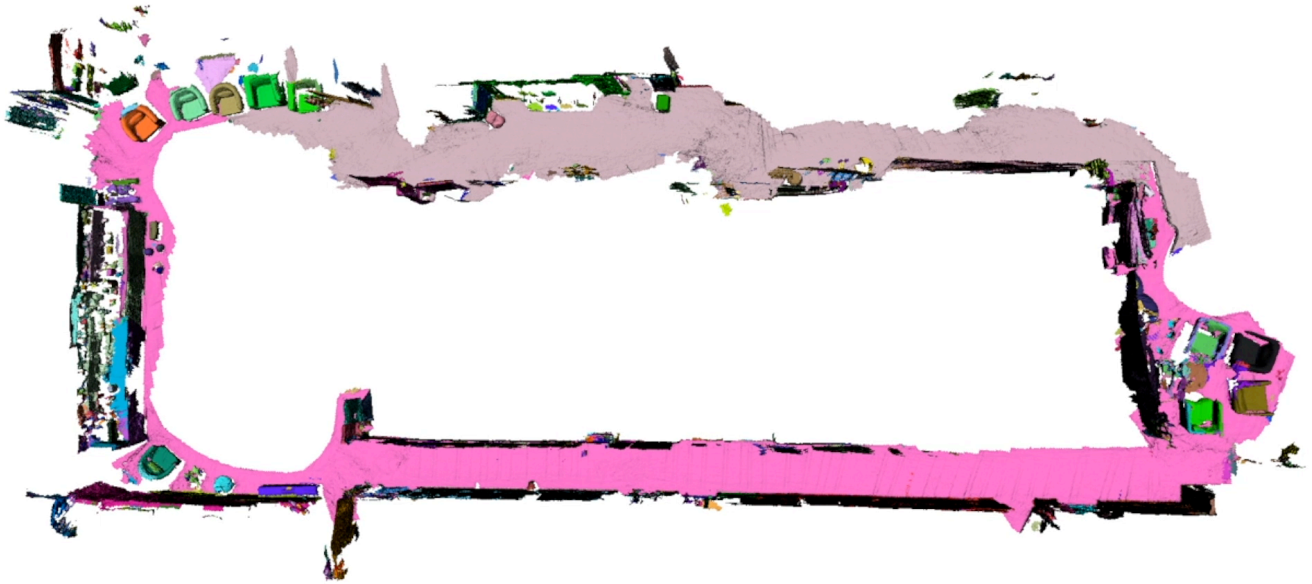
Dense object-level robotic mapping

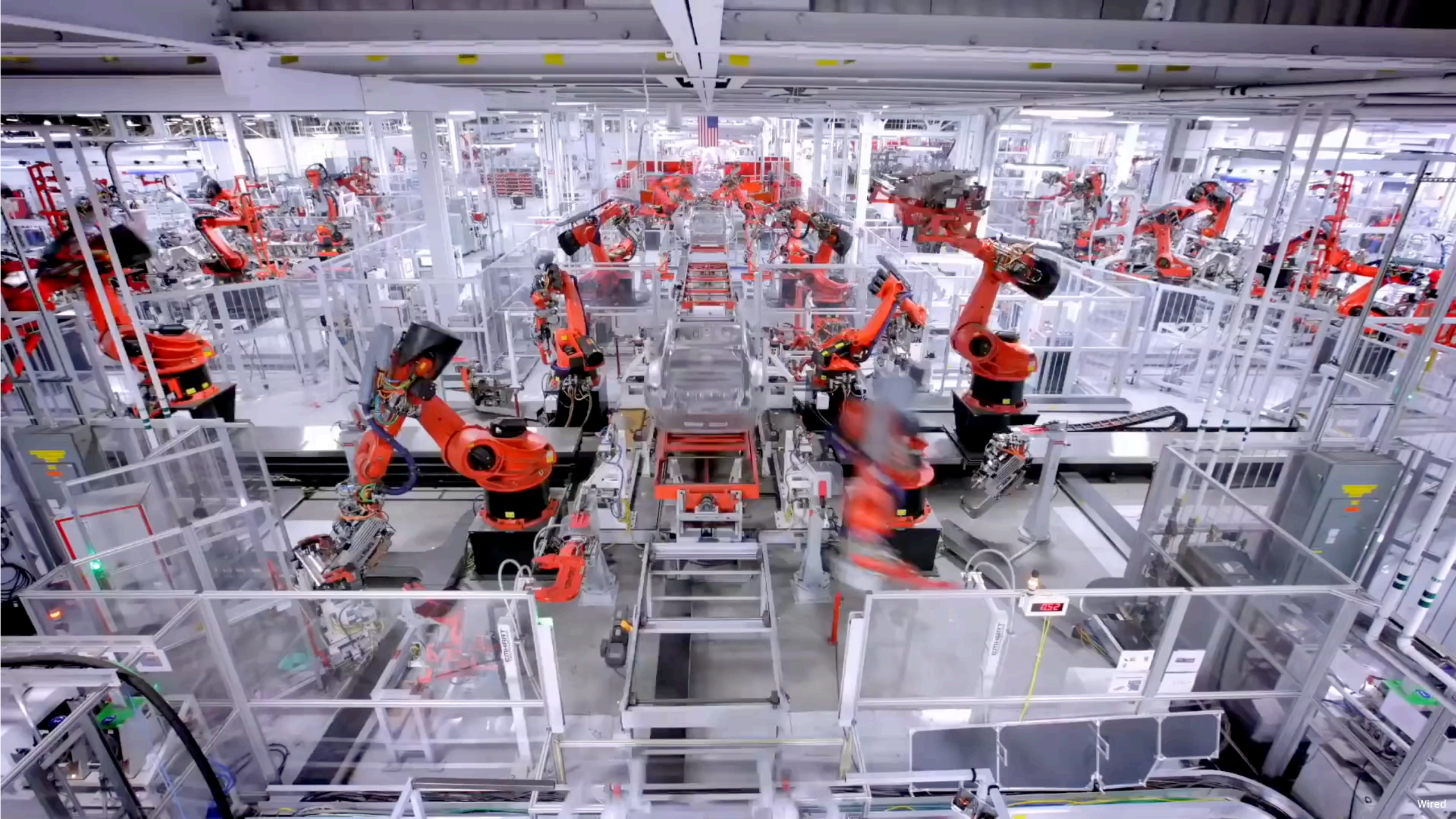


Margarita Grinvald
Autonomous Systems Lab

Doctoral thesis presentation
16 Dec 2021

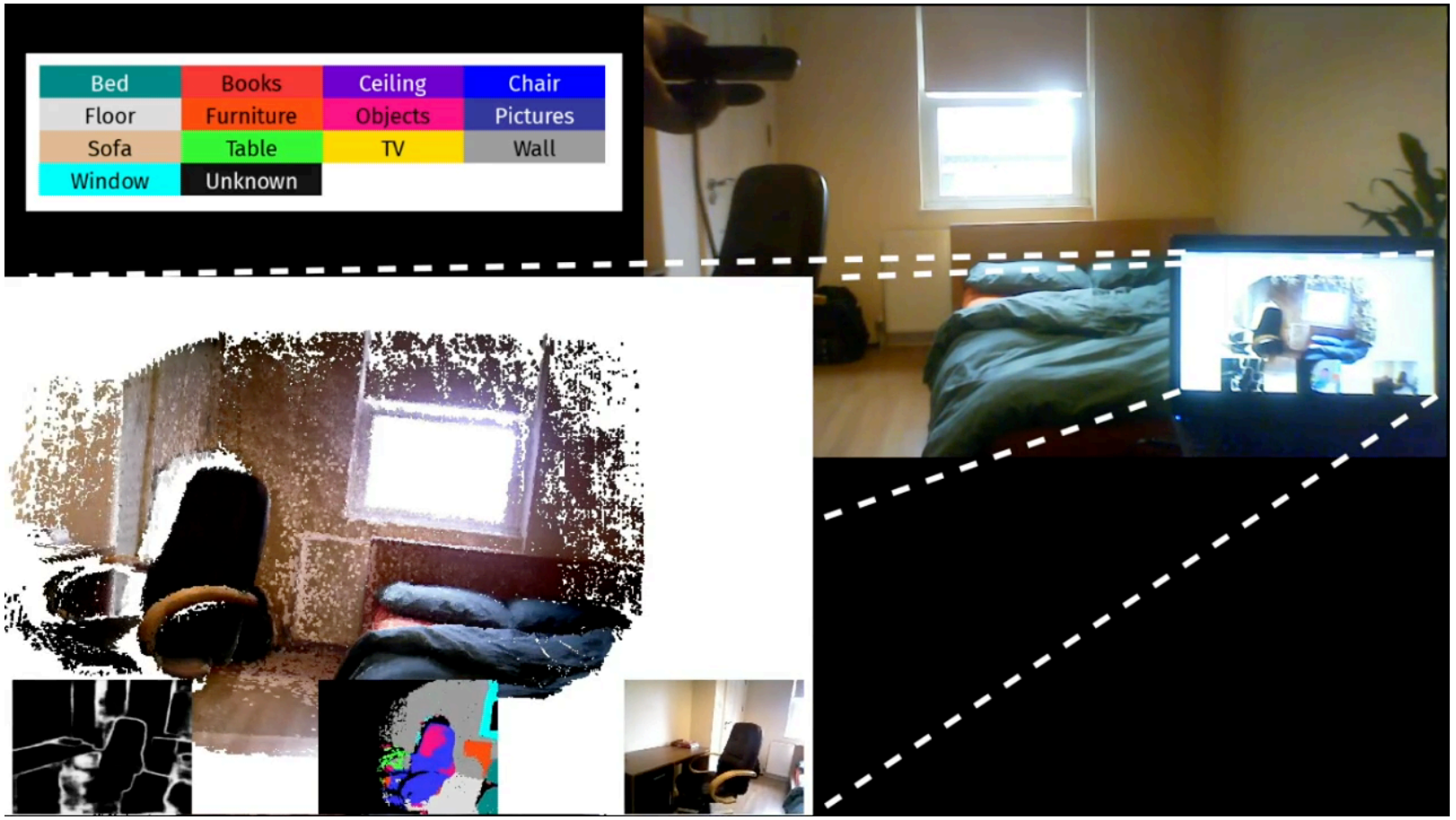
Supervisor: Prof. Dr. Roland Siegwart
Co-examiner: Prof. Dr. Stefan Leutenegger

ETH zürich


Wired

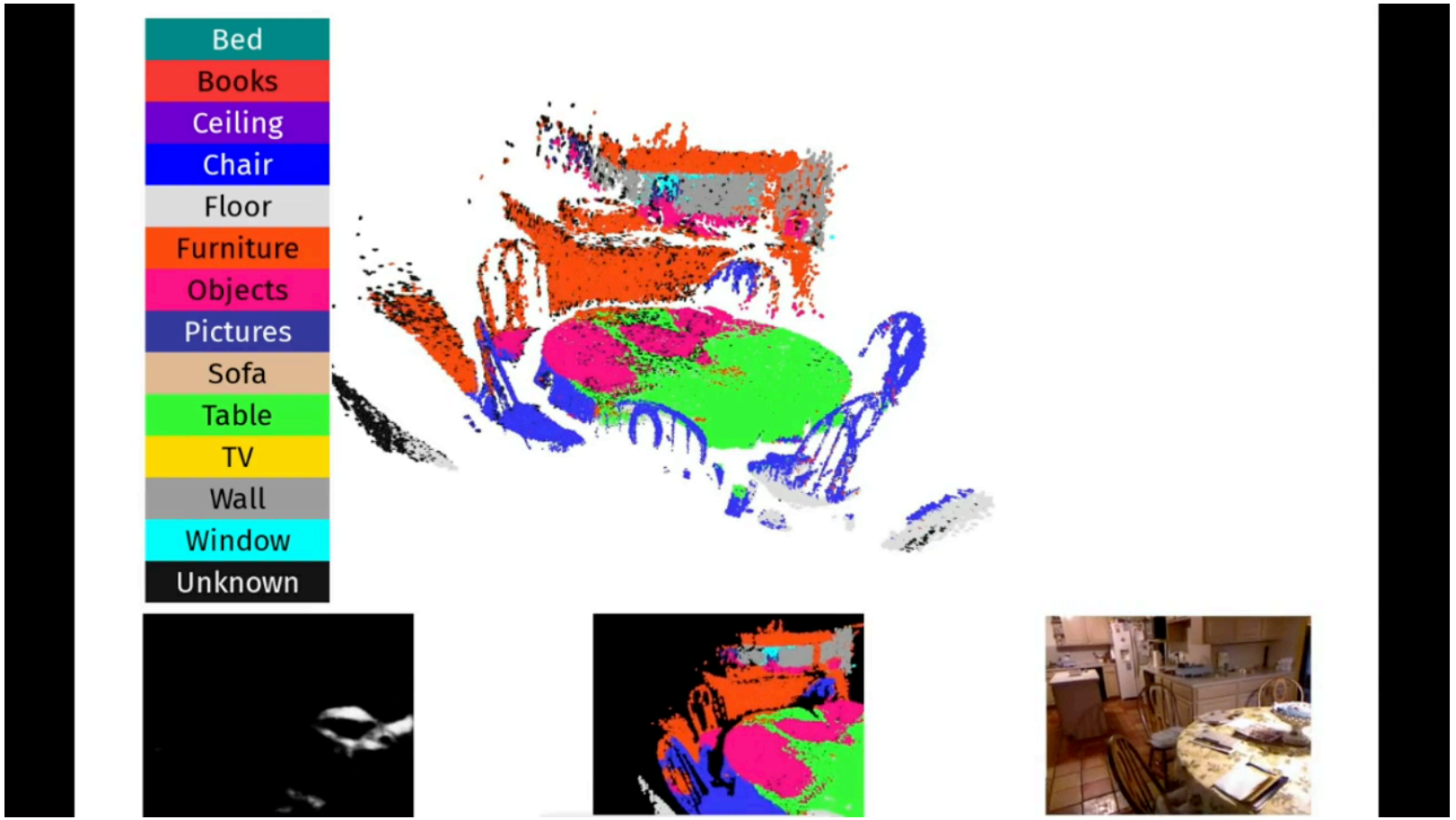
Robots operating autonomously in the wild must perceive and understand the surrounding environment

Autonomous interaction planning requires environment maps that provide a high-level interpretation of the scene



McCormac et al., SemanticFusion, 2016

Traditional semantic mapping classifies scene parts by category but disregards individual object instances



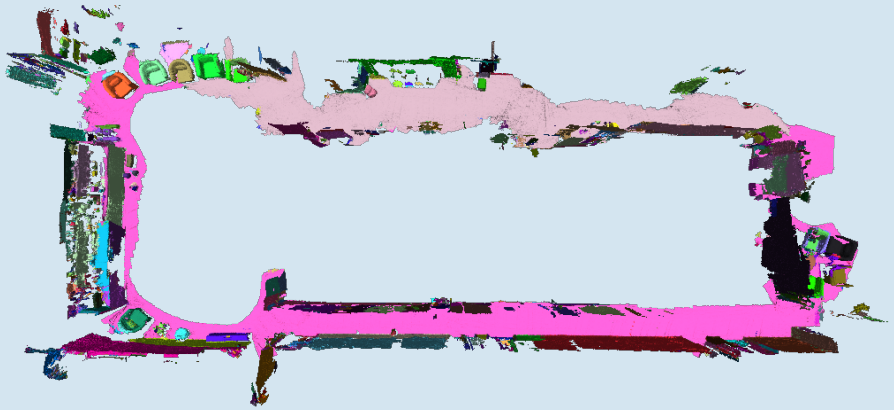
McCormac et al., SemanticFusion, 2016

The presented research is motivated by a set of observations

- Man-made environments are primarily populated by physical objects
- Robotic manipulation typically targets individual object instances
- Upon interaction, objects move as a whole

Thesis: physical objects provide the optimal functional unit for a high-level map of the environment

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

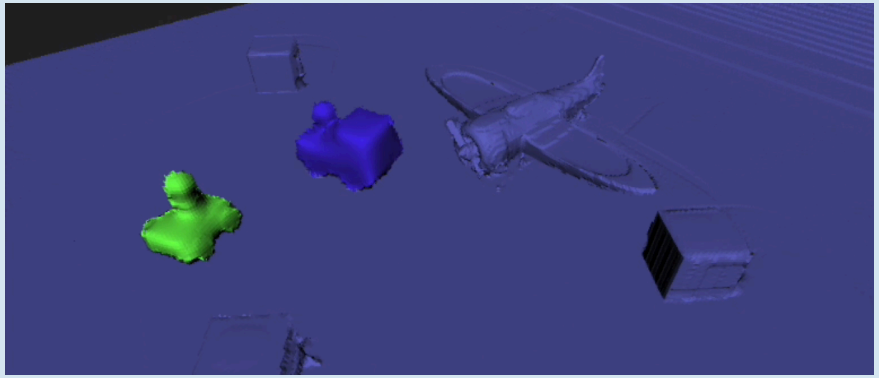
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

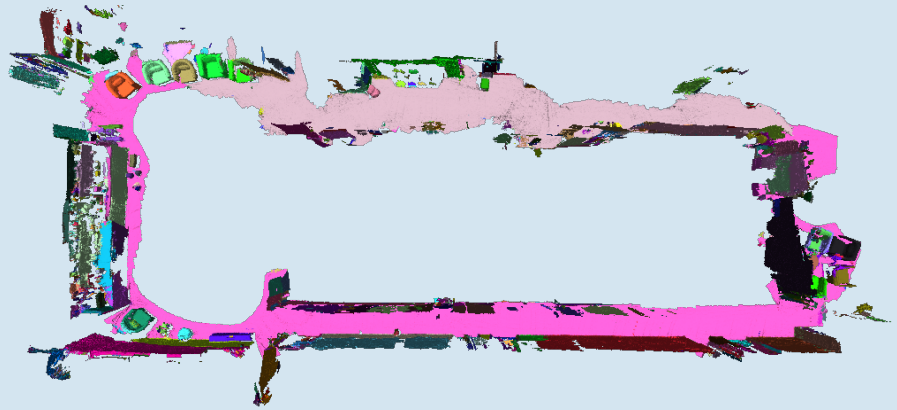


M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Dynamics

Dynamic object tracking and reconstruction

M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

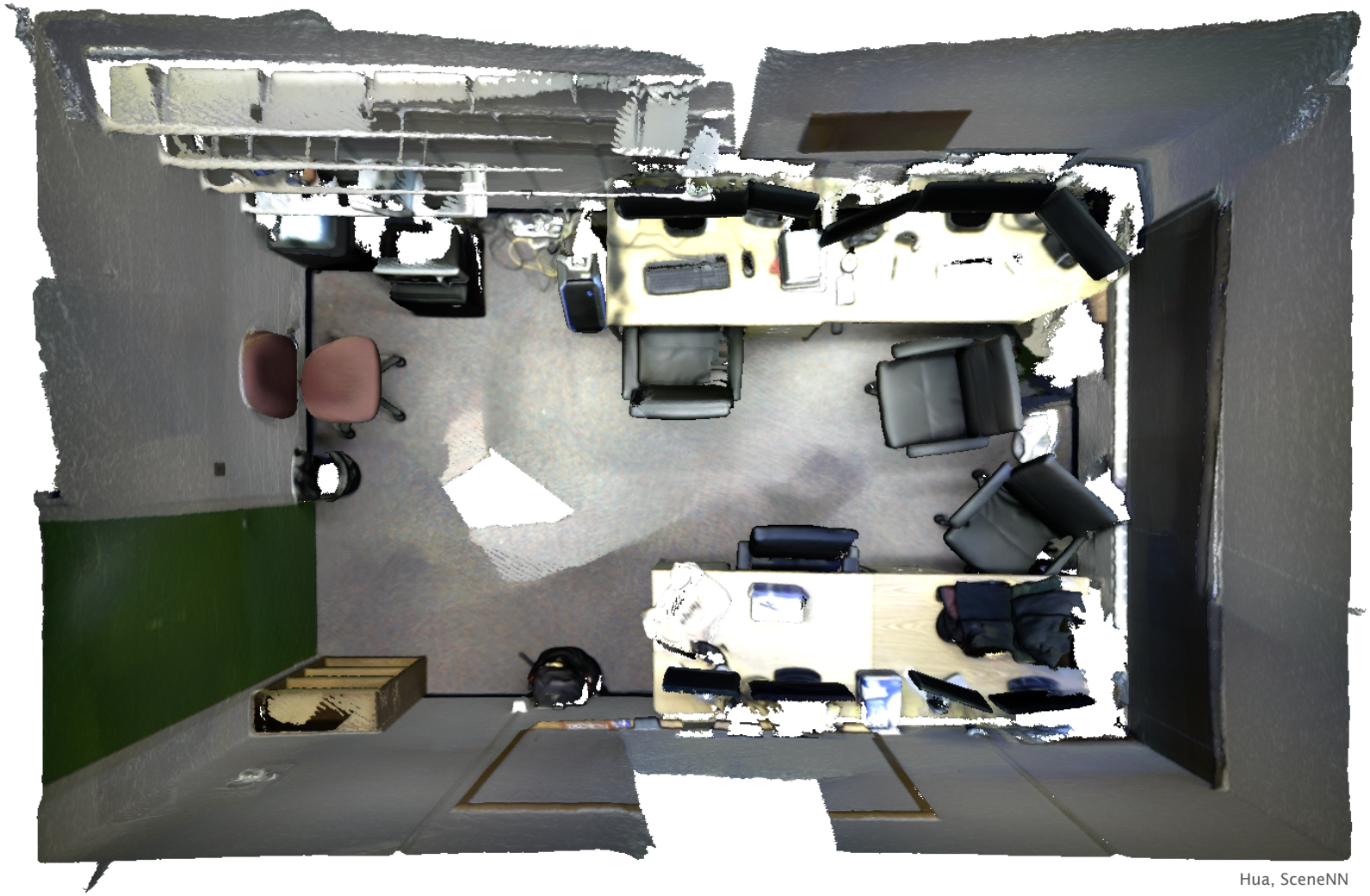
Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Outlook

Future research avenues

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



Hua, SceneNN

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

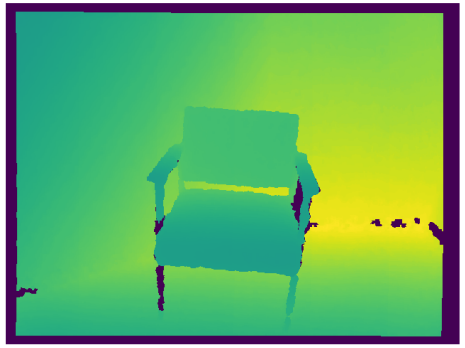
Hua, SceneNN

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

Hua, SceneNN


ASUS

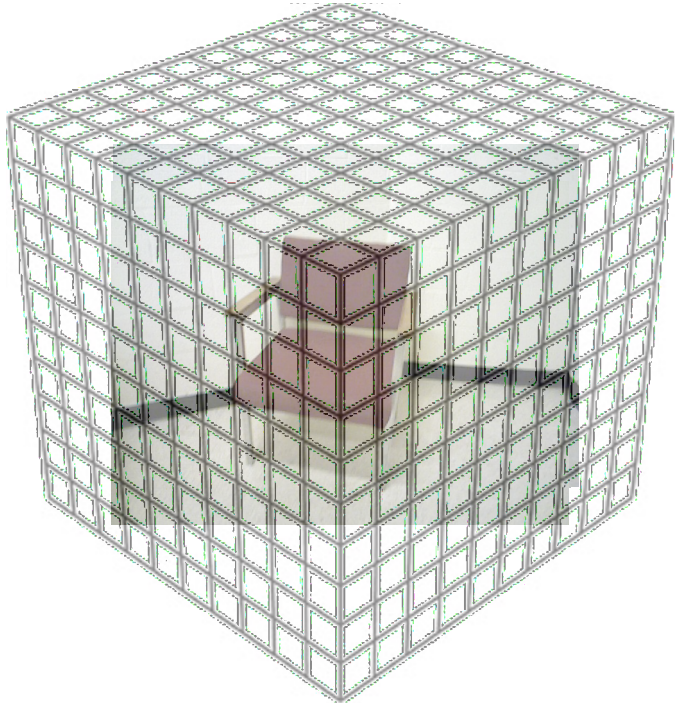
RGB



Depth

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

				1	1	0.2	0.1	-0.2	-0.9
		1	1	0.9	0.5	0.1	-0.5	-0.9	-1
1	1	1	1	0.8	0.2	-0.2	-0.5	-0.9	-1
1	1	1	1	0.7	0.1	-0.2	-0.9	-1	-1
1	1	1	1	0.7	0.2	-0.2	-0.9	-1	-1
1	1	1	0.6	0.2	0.1	-0.5	-1	-1	-1
1	1	1	0.9	0.5	0.2	-0.5	-0.9	-1	-1
	1	1	1	0.6	0.1	-0.2	-0.8	-1	-1
			1	0.5	0.3	-0.1	-0.1	-0.5	-1
					1	1	1	0.5	-1

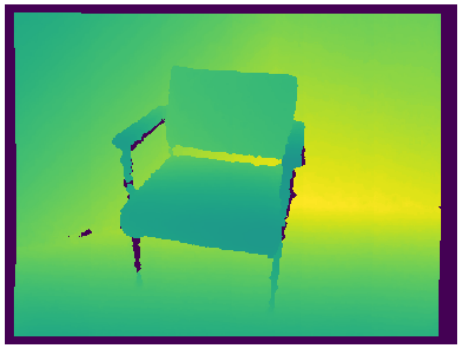


TSDF grid

Oleynikova et al., Voxblox, 2017

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

RGB

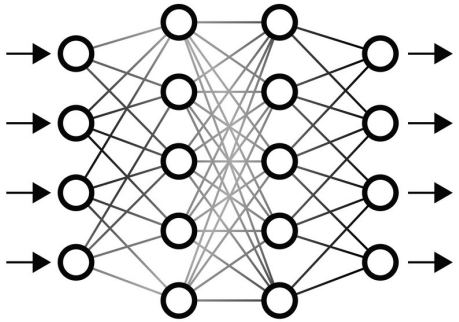


Depth

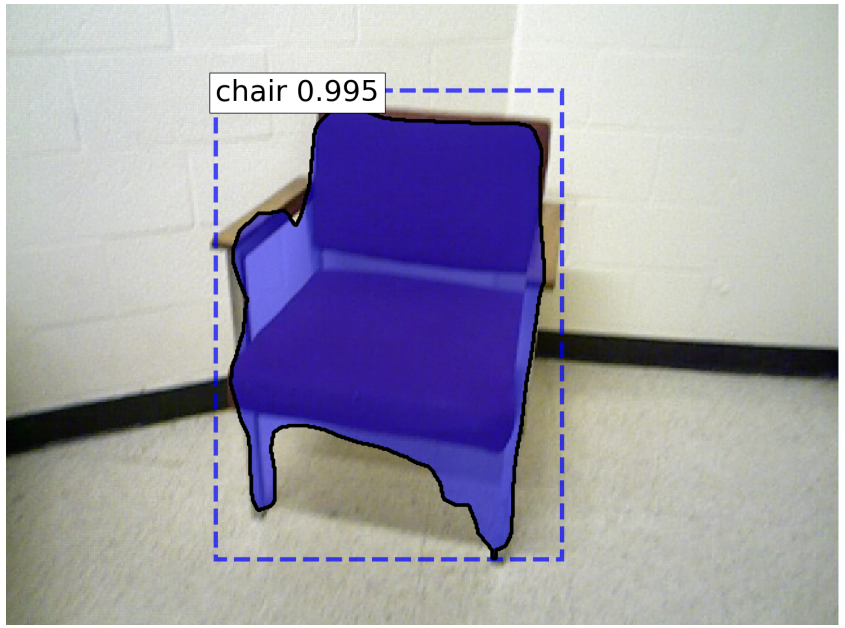
Framewise 2D segmentation

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

RGB

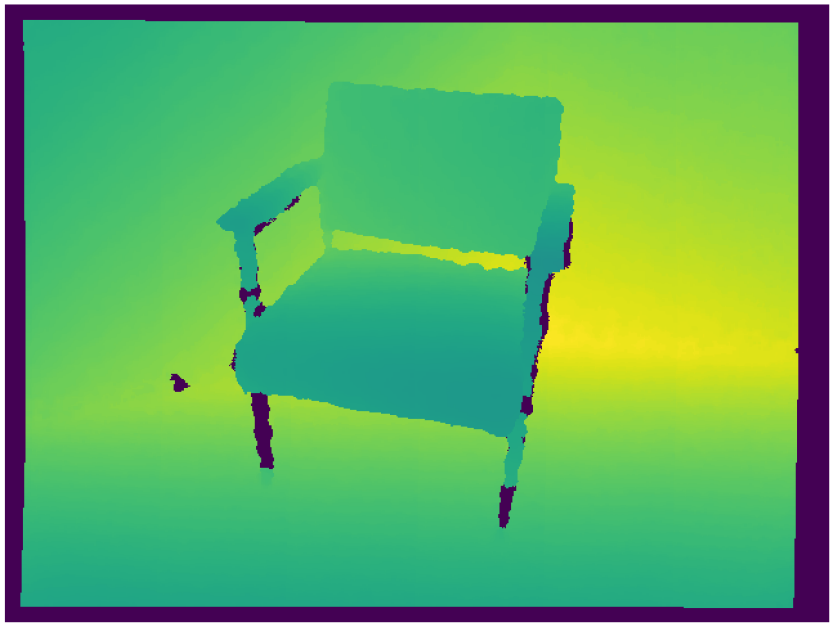


Mask R-CNN

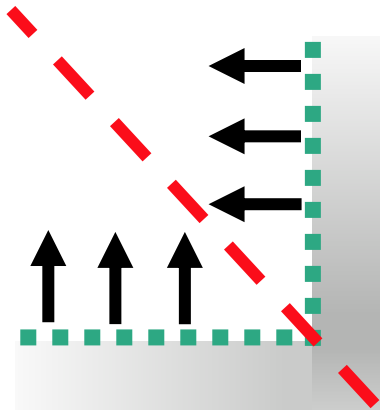


Semantic instance-aware segmentation

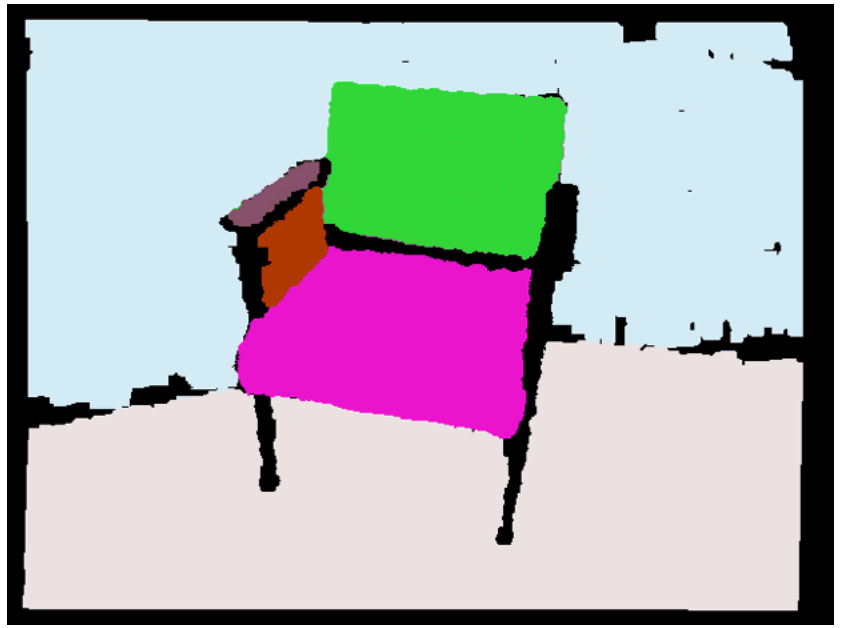
An unsupervised geometric method exhaustively (over)segments the depth frame



Depth



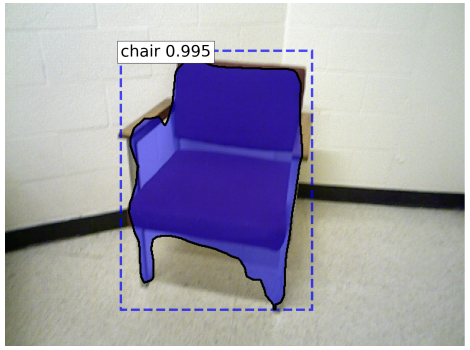
Convexity criterion



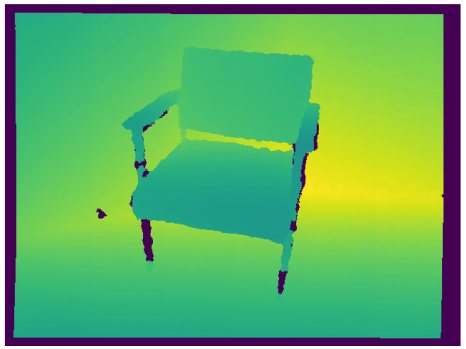
Convexity-based segmentation

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

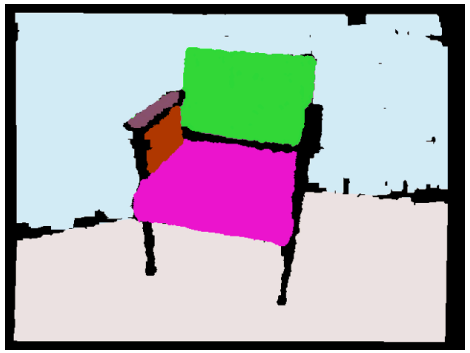
RGB



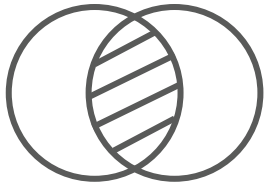
Mask R-CNN



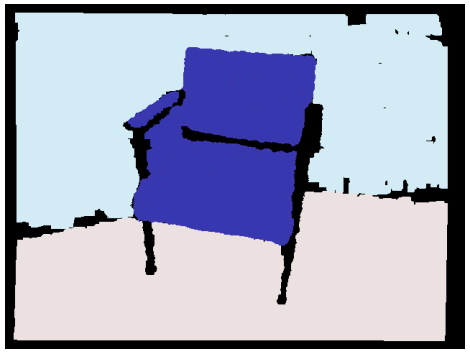
Depth



Geometric segmentation

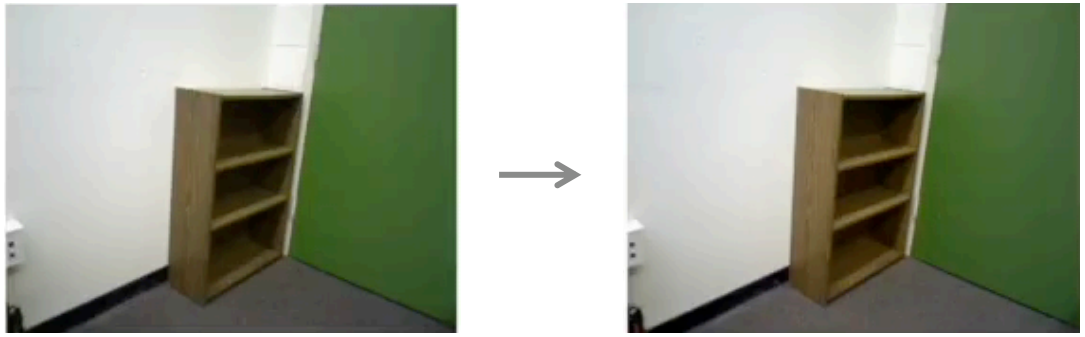


Overlap measure



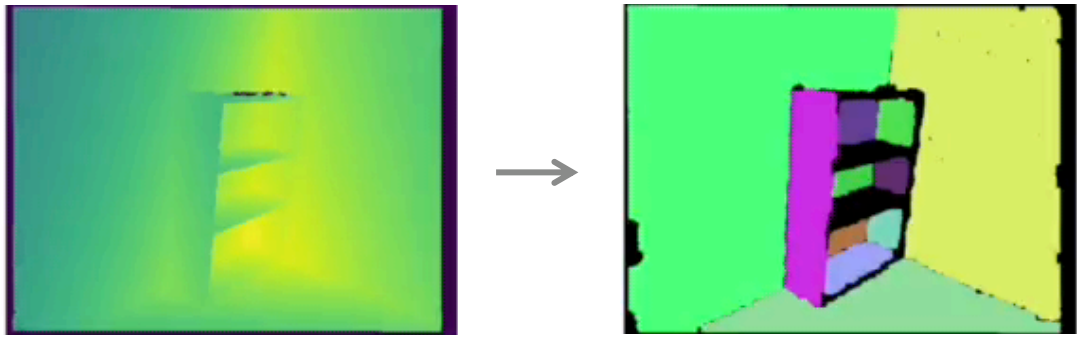
Semantically refined geometric segmentation

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



RGB

Mask R-CNN

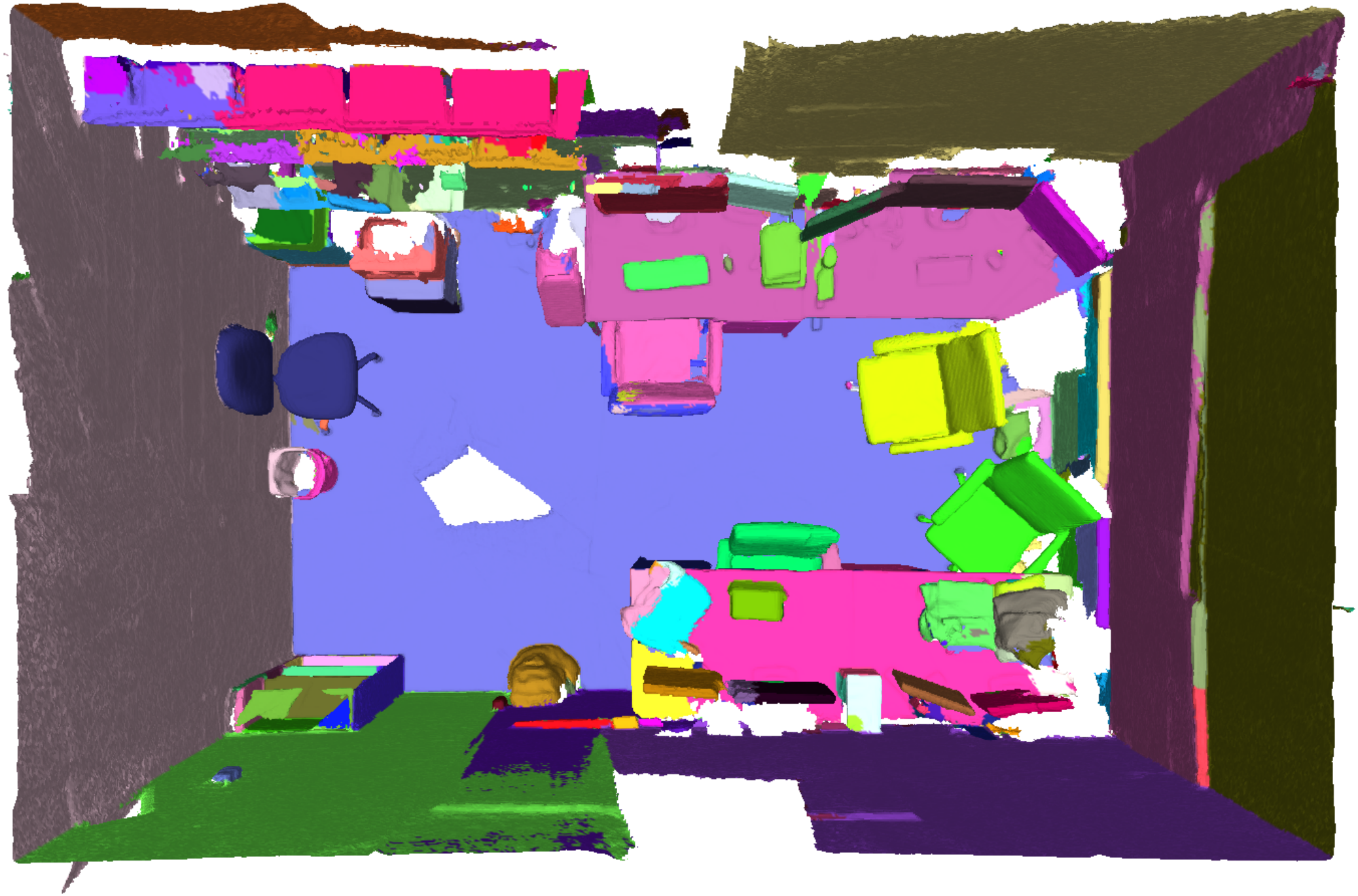


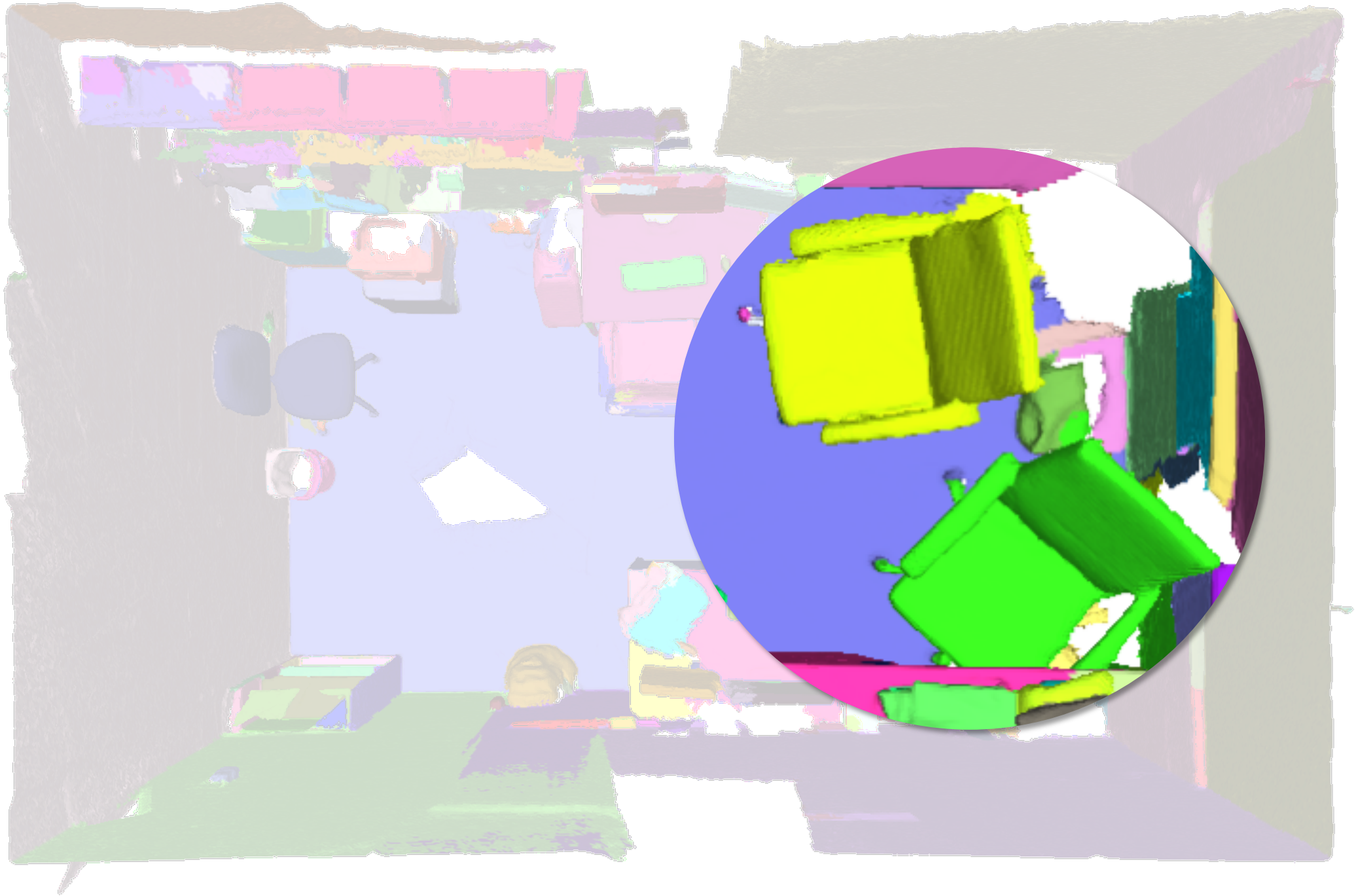
Depth

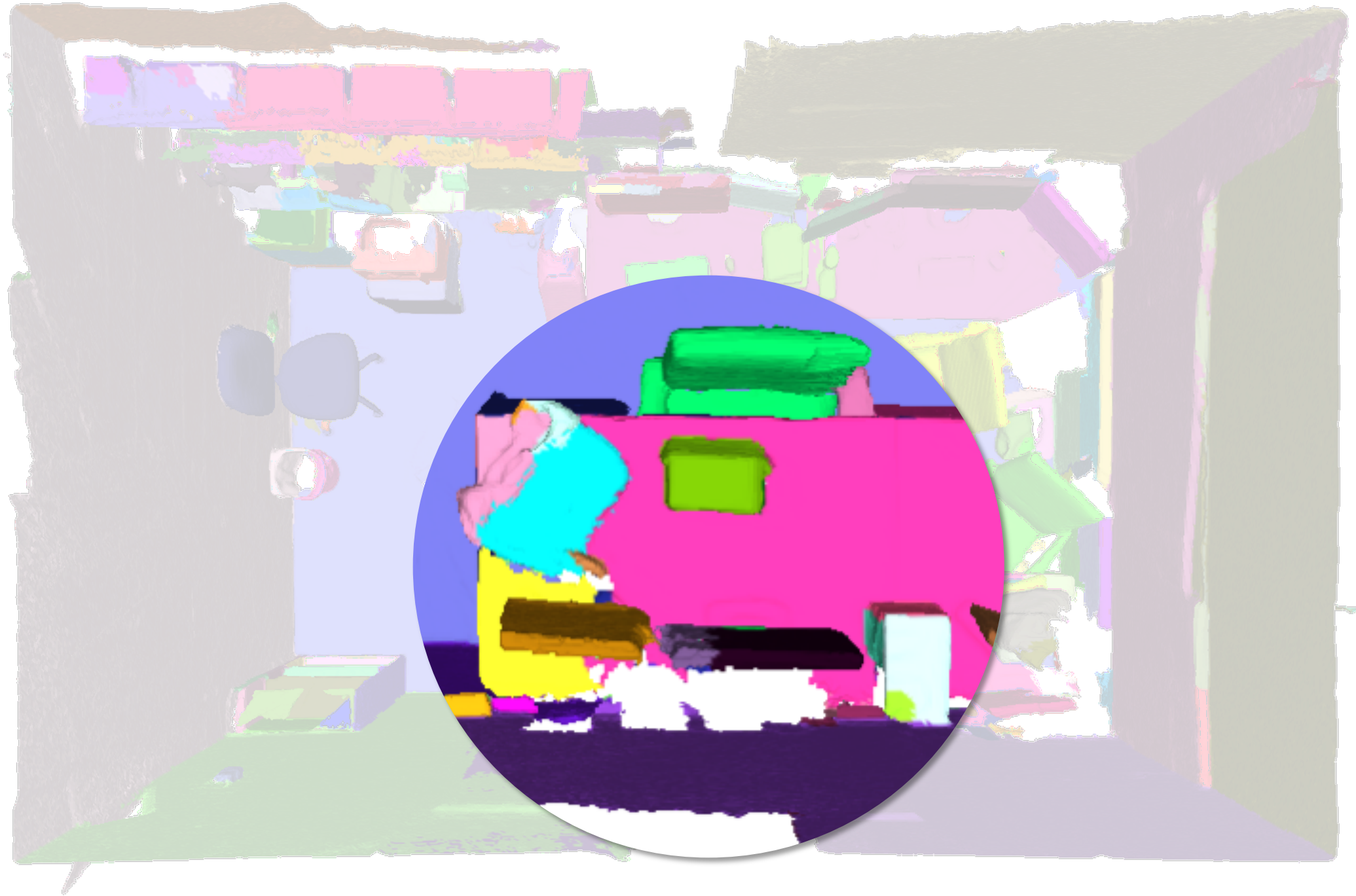
Geometric segmentation

*not actual speed

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





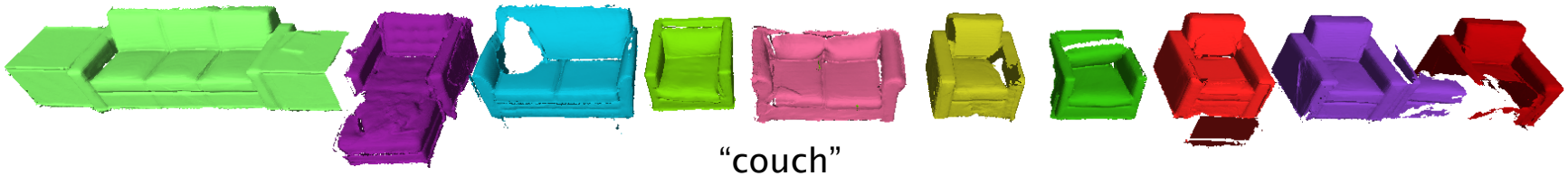



Monitor
Keyboard
Suitcase
Table
Chair
Mouse
Refrigerator
Plant
Backpack
Cup
Microwave
Unknown

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

“chair”

AP_{50} = 51.7

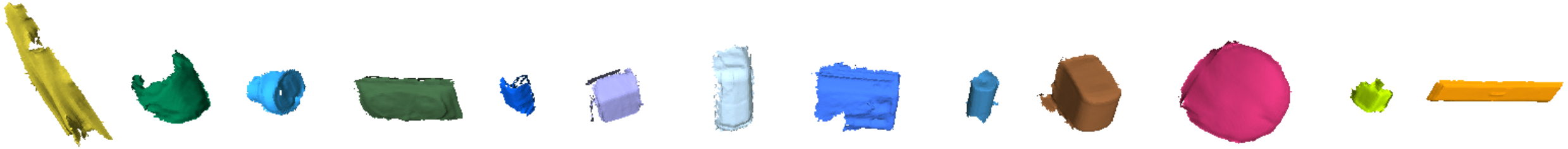


“couch”

AP_{50} = 70.0

“table”

AP_{50} = 58.8



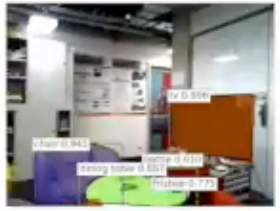
[jacket] [bag] [fan] [fan] [speaker] [box] [case] [heater] [paper roll] [appliance] [pillow] [tissues] [drawer]

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

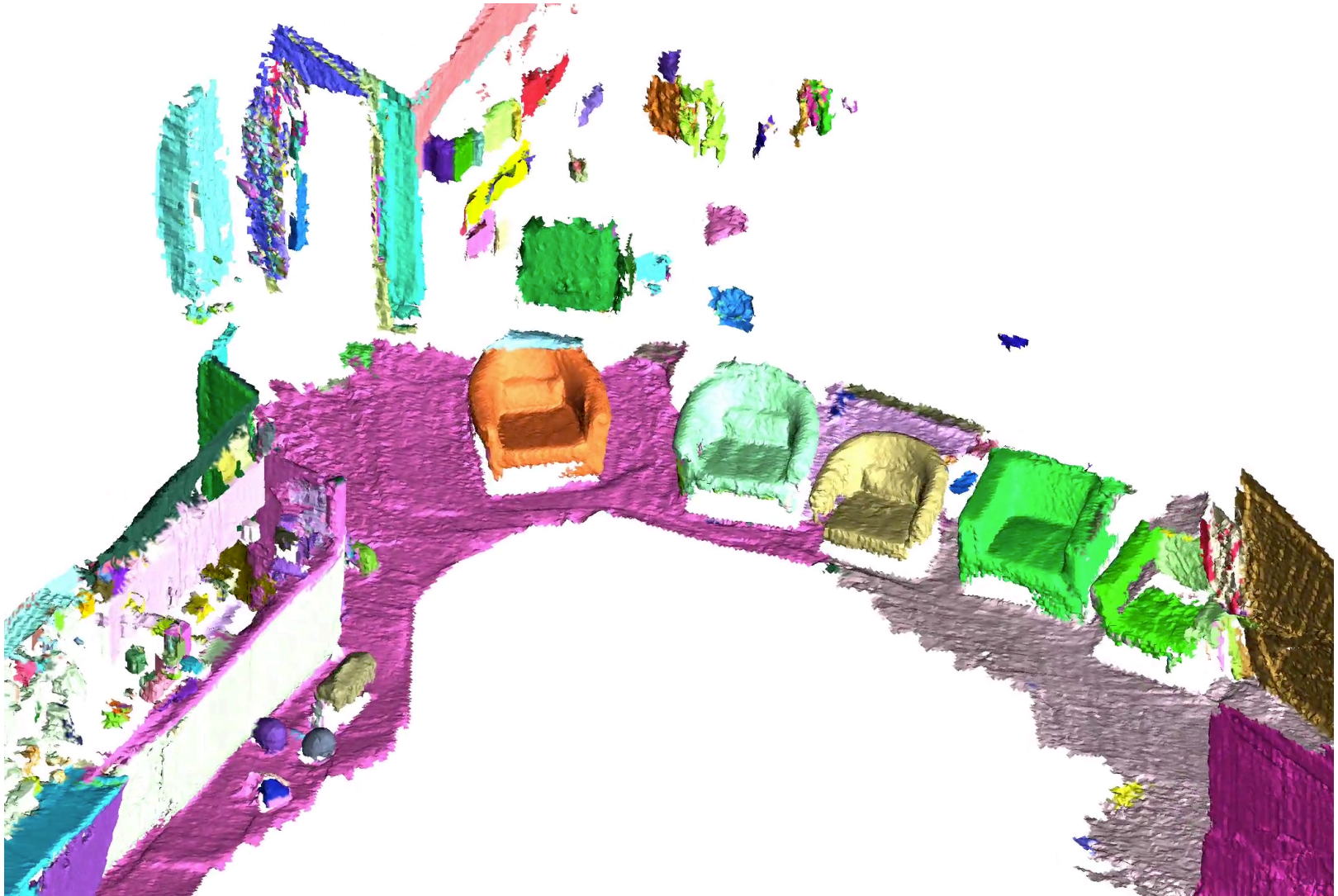
RGB

Mask R-CNN

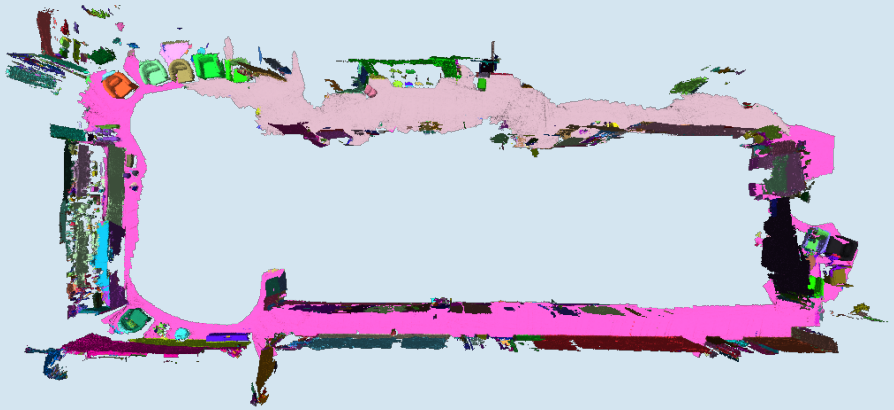
Depth segmentation

10x speed

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



Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

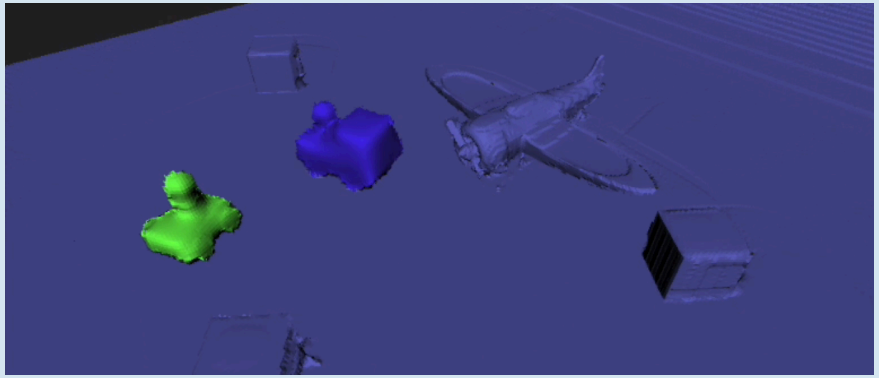
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

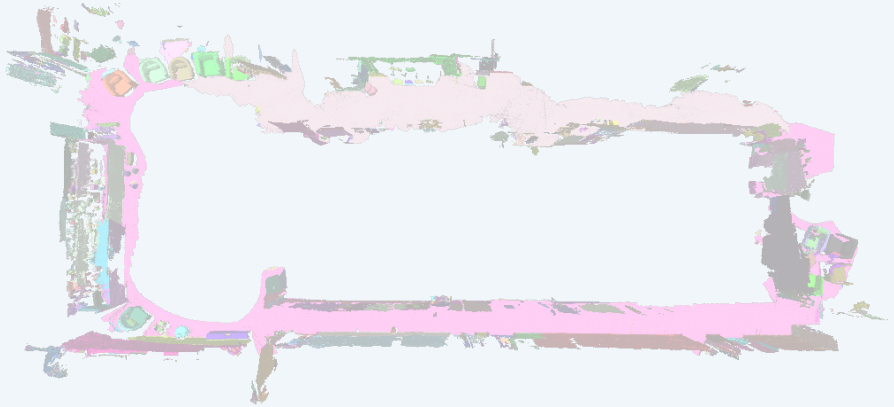


M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Dynamics

Dynamic object tracking and reconstruction

M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

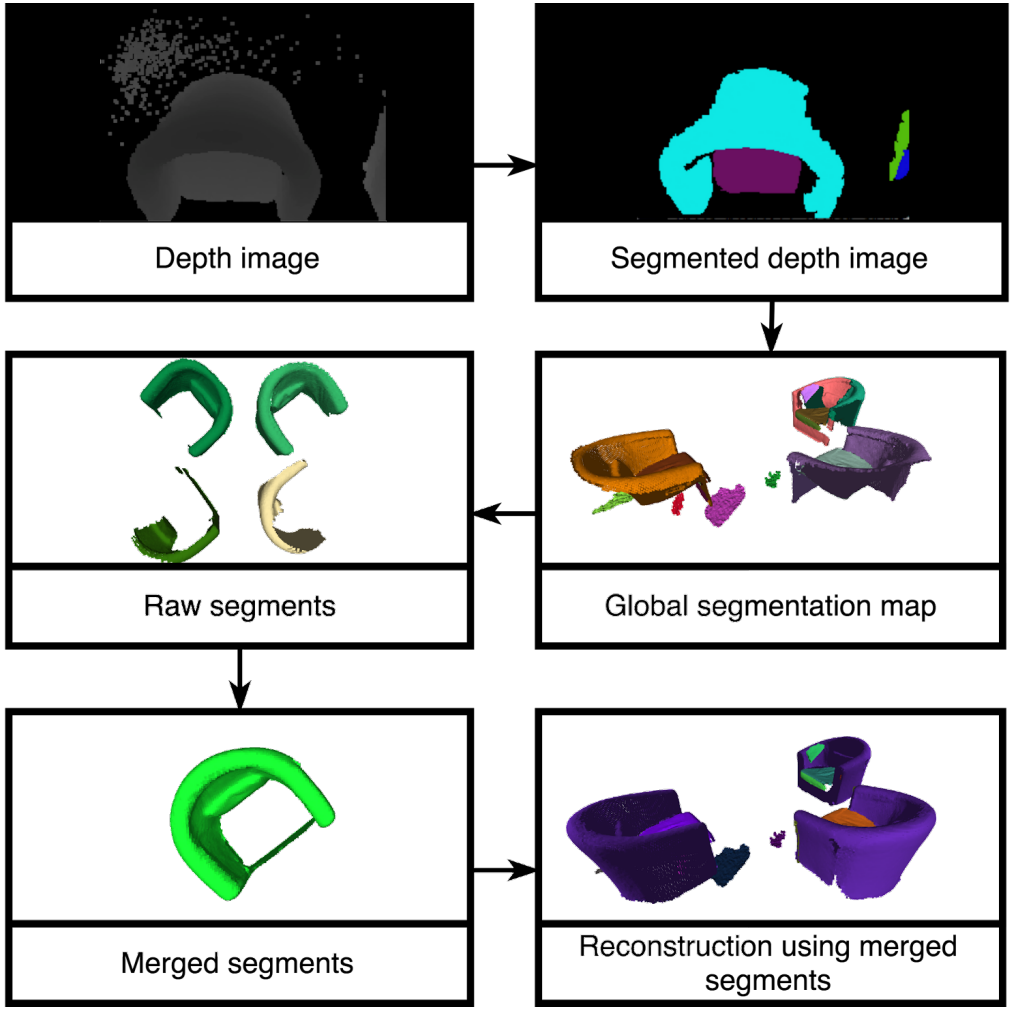
Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

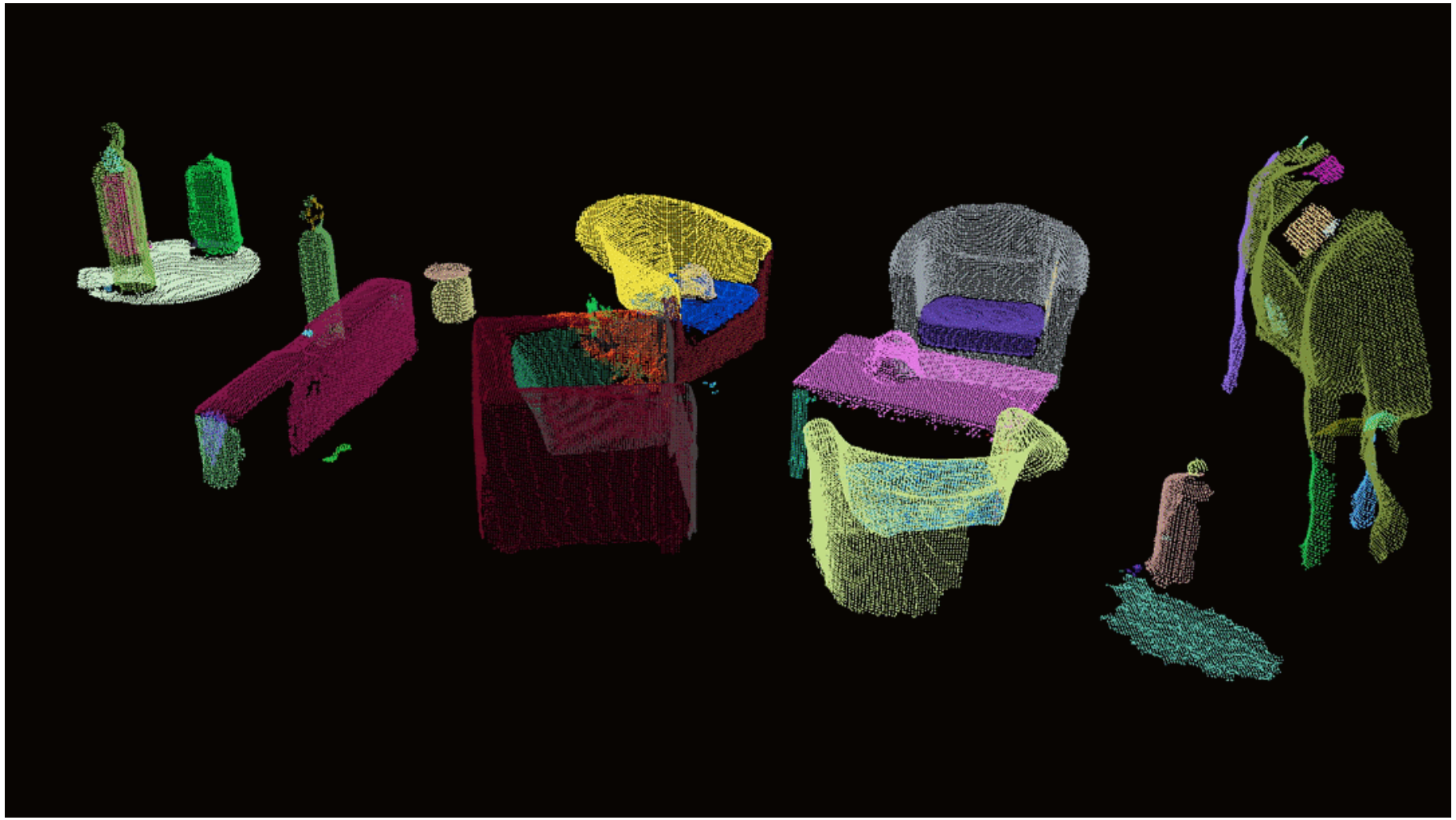
Outlook

Future research avenues

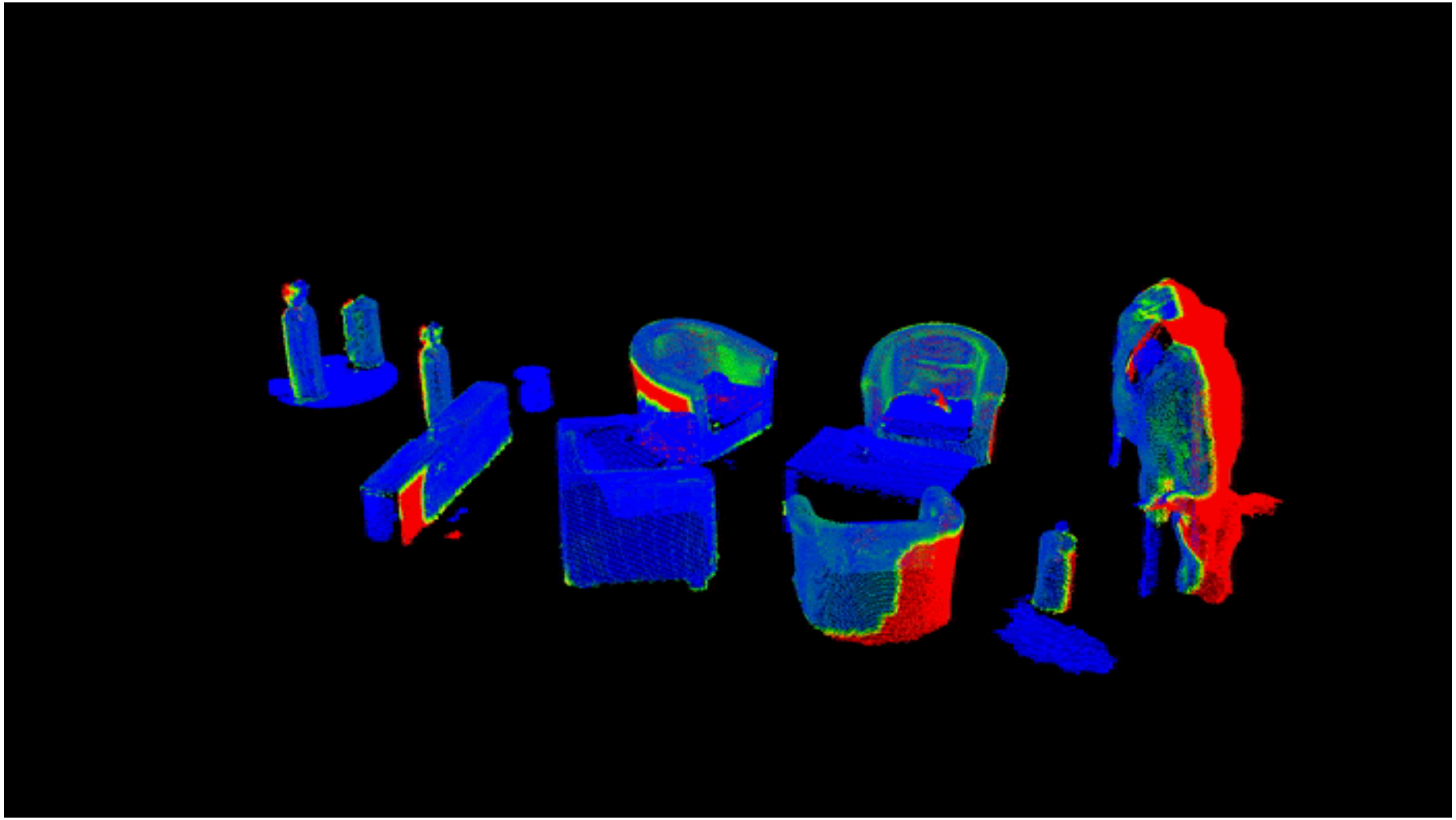
Geometry-only object mapping can be extended to build a persistent database of object models



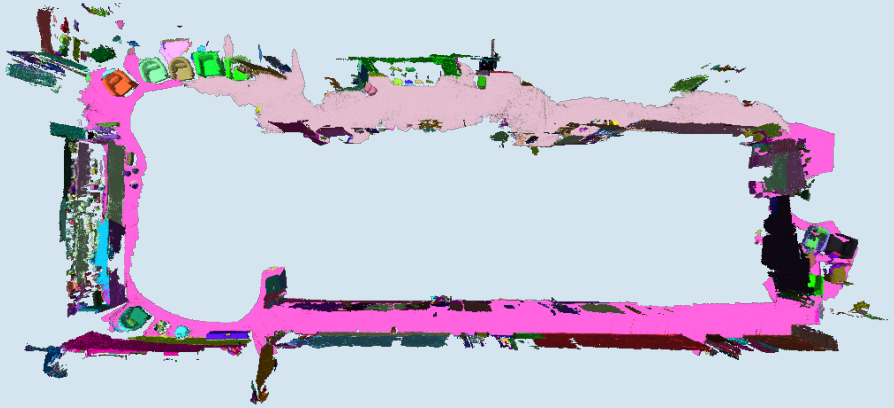
Matched and merged object models can aid reconstruction by replacing the corresponding partial segments in the scene



Matched and merged object models can aid reconstruction by replacing the corresponding partial segments in the scene



Dense object-level mapping

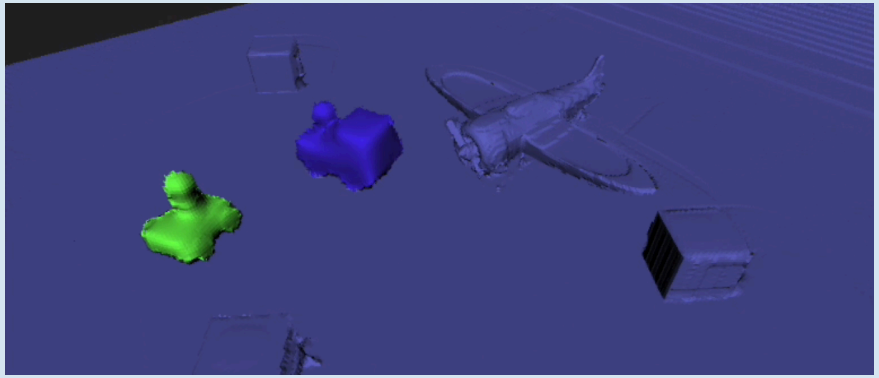


M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Dynamics

Dynamic object tracking and reconstruction



M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Outlook

Future research avenues

Dense object-level mapping

M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

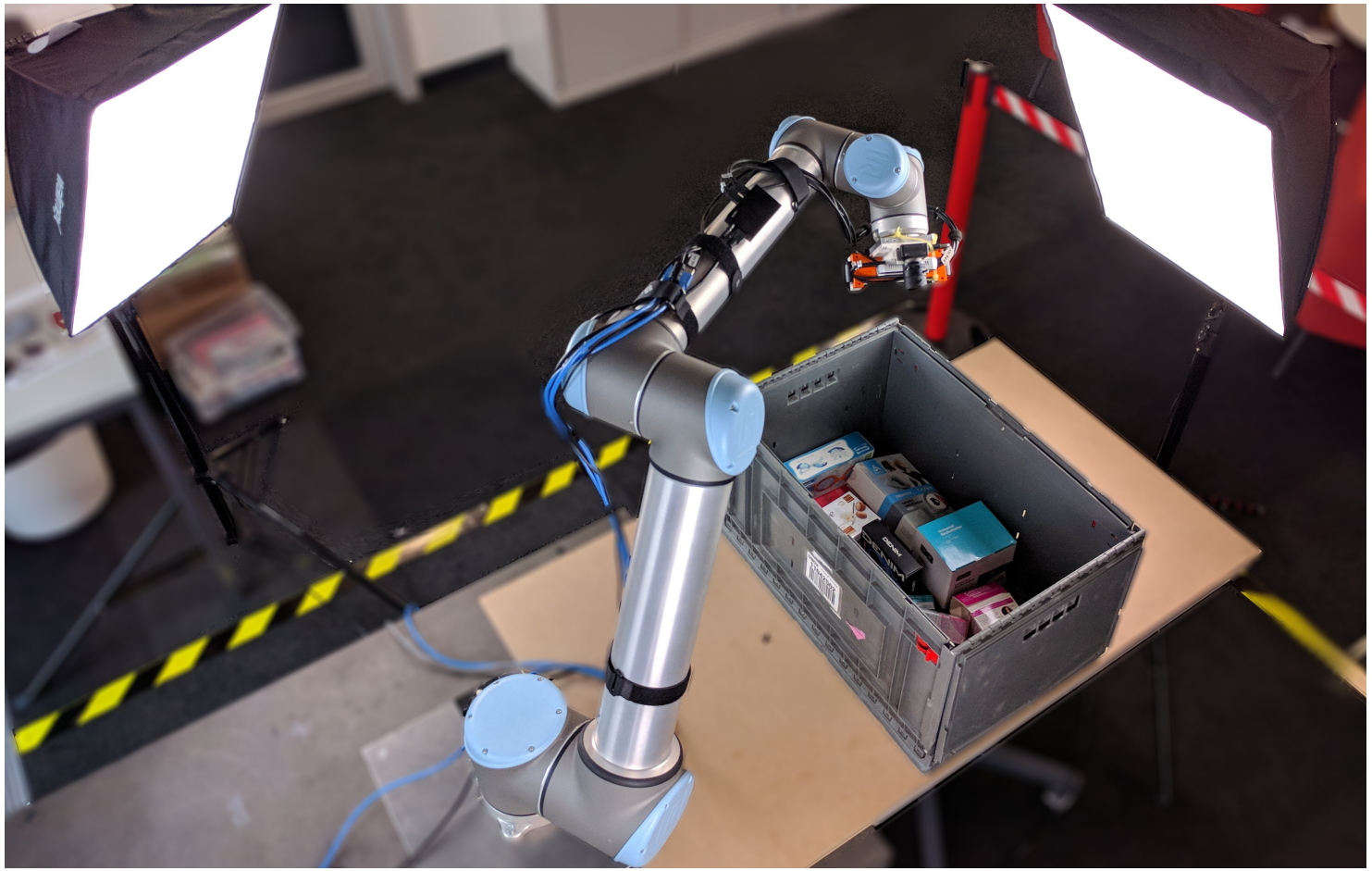
M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

The CLUBS dataset captures warehouse-like cluttered boxes containing common household items

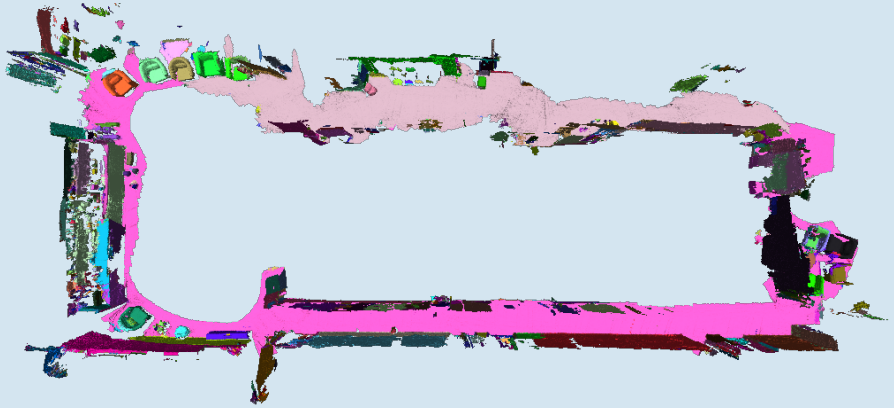
Data is collected with a robotic arm and manually annotated with 2D and 3D object information



3D reconstruction

Instance segmentation

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

↓

Online object database building

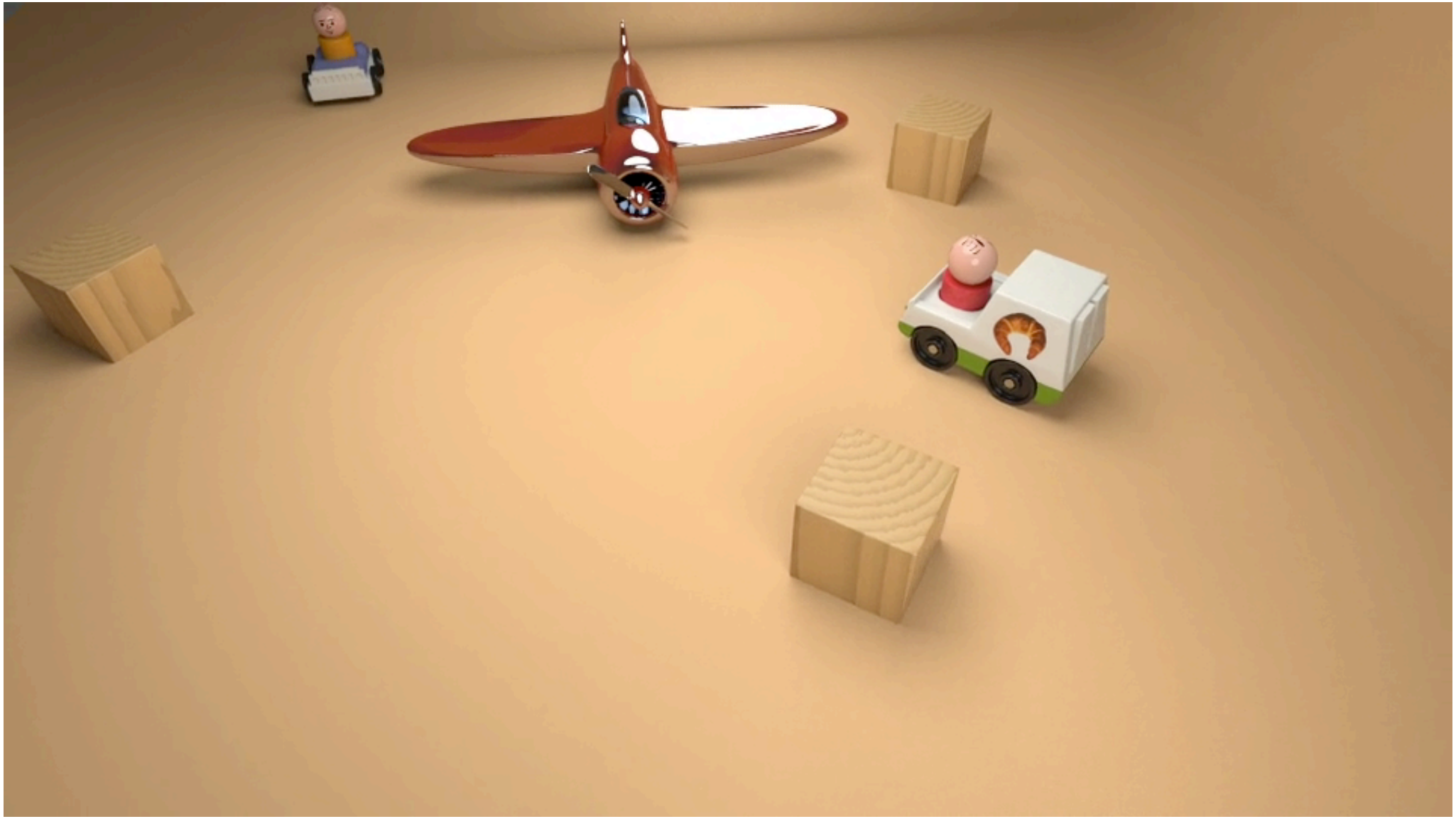
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

↓

Dataset of common household objects

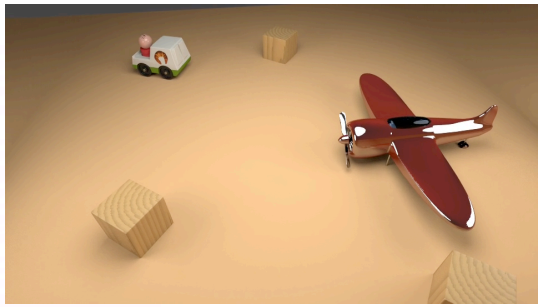
T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

What would happen if we process a dynamic scene?

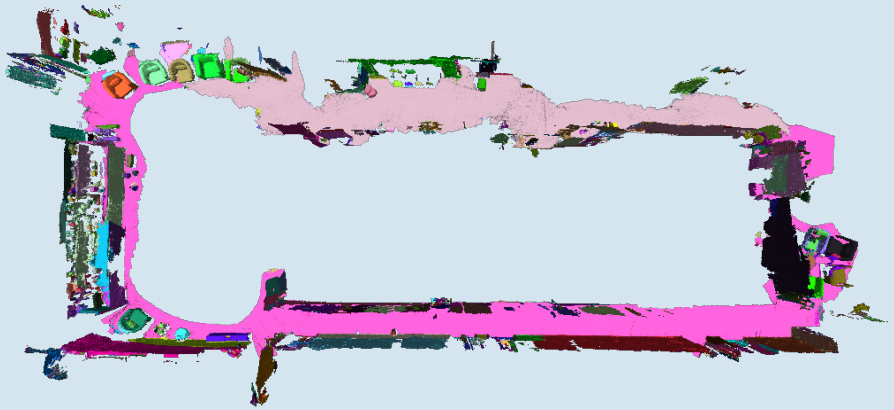


Rünz et al., Co-Fusion, 2017

Fusing dynamic content into a stationary map introduces ghost artifacts in the reconstruction



Dense object-level mapping

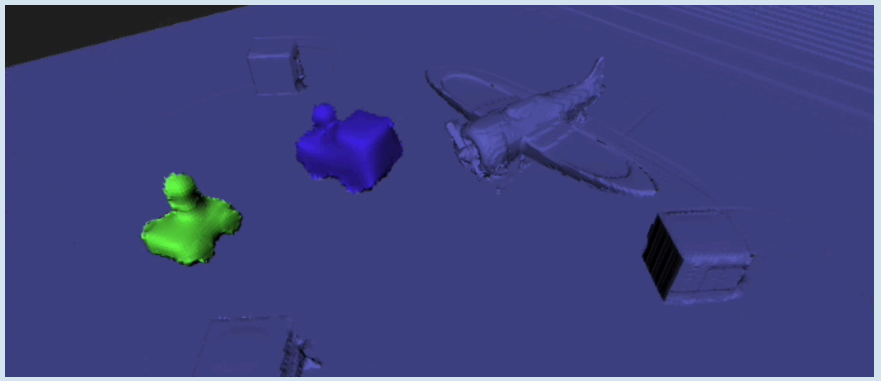


M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Dynamics

Dynamic object tracking and reconstruction



M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Outlook

Future research avenues

Dense object-level mapping

M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

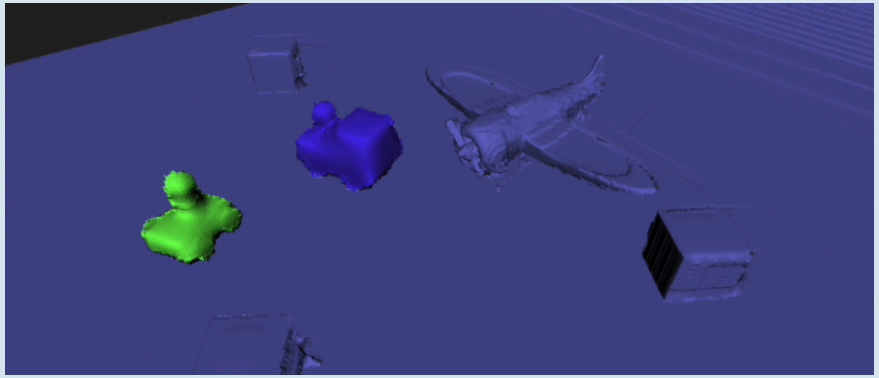
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

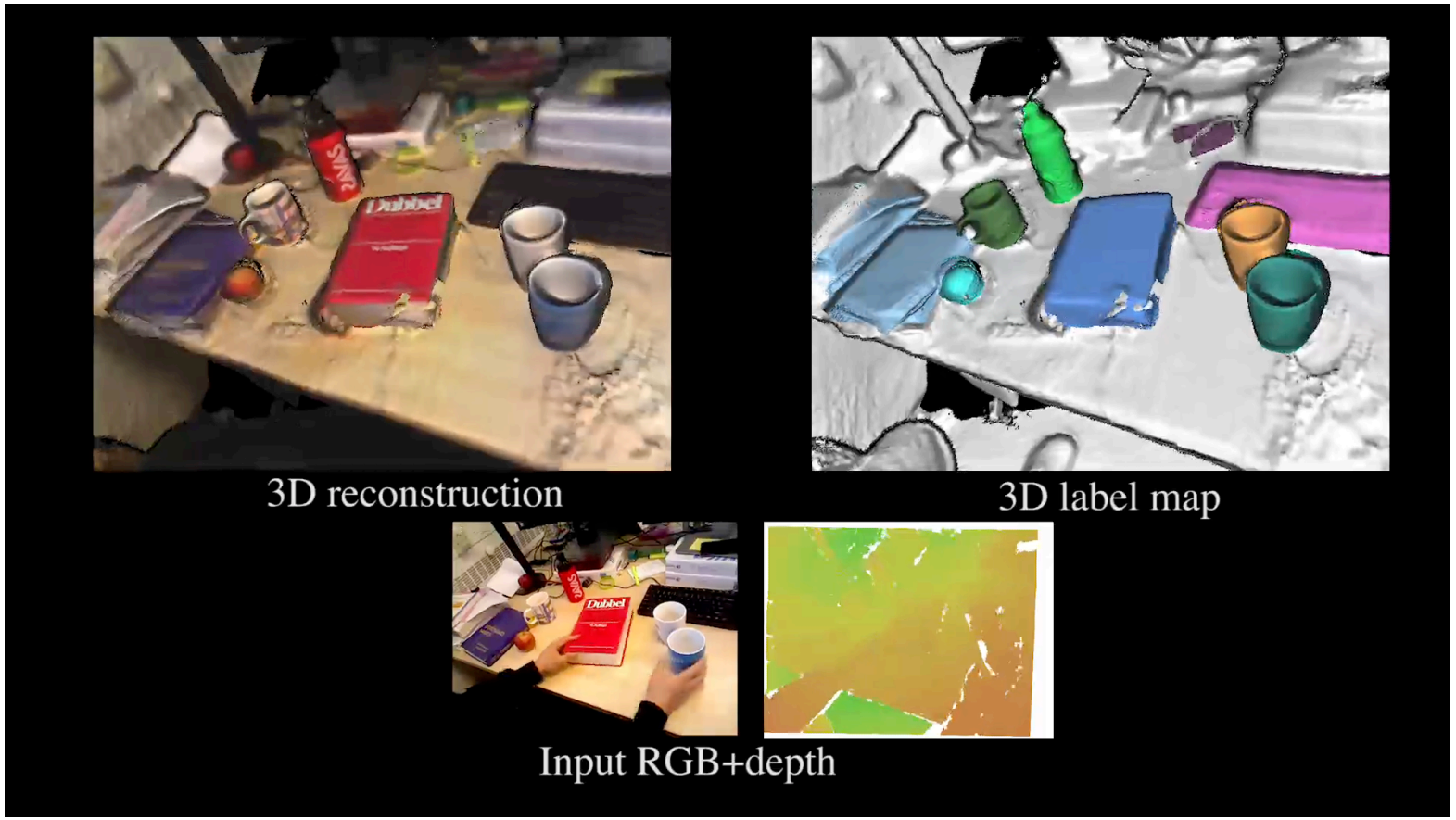


M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

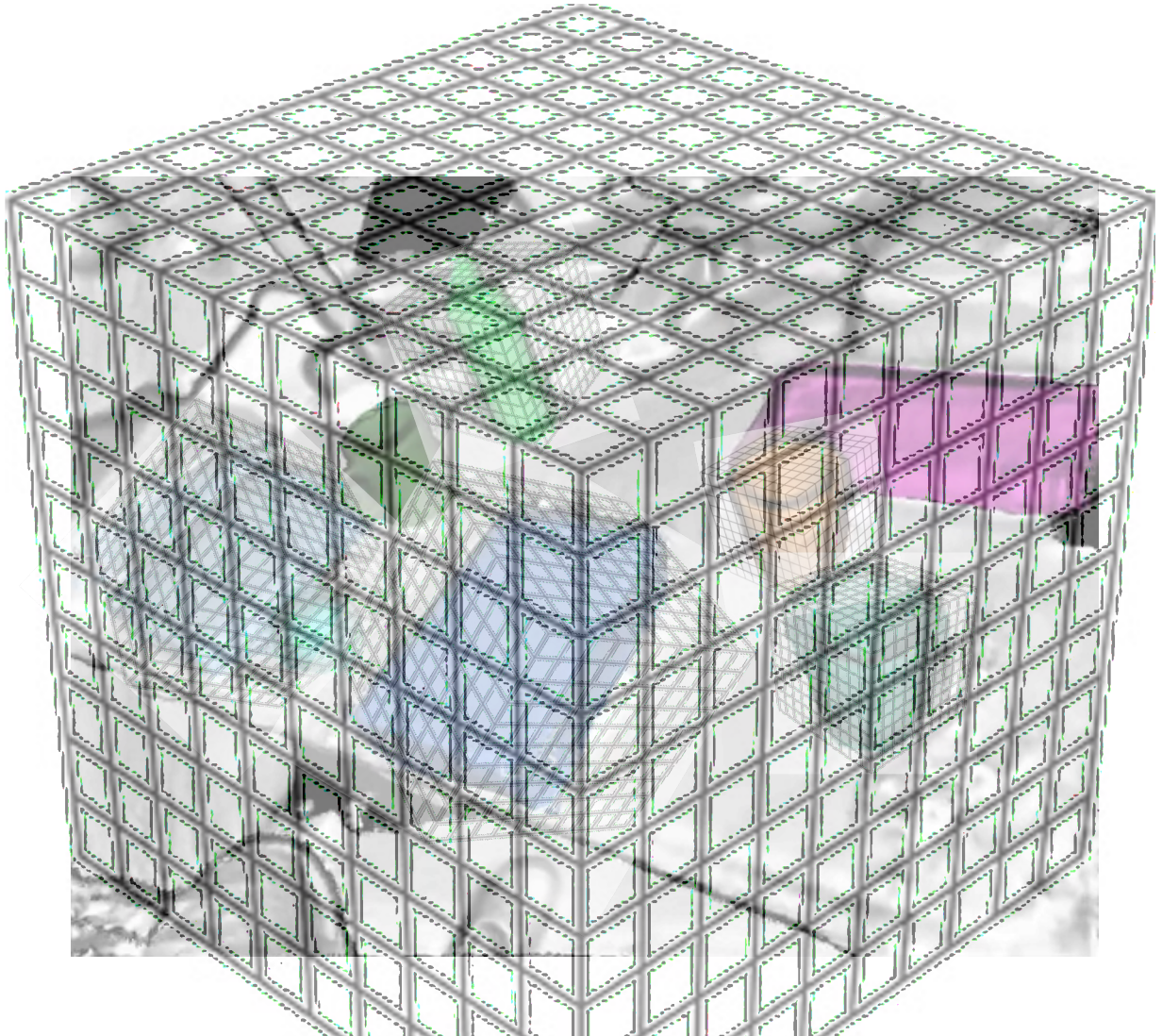
Future research avenues

MID-Fusion tracks multiple dynamic objects and reconstructed them in separate octrees



Xu et al., MID-Fusion, 2018

All previous works store individual objects in separate volumes and track the relative pose between them



The map as a collection of 3D volumes is simple and intuitive but it carries several limitations

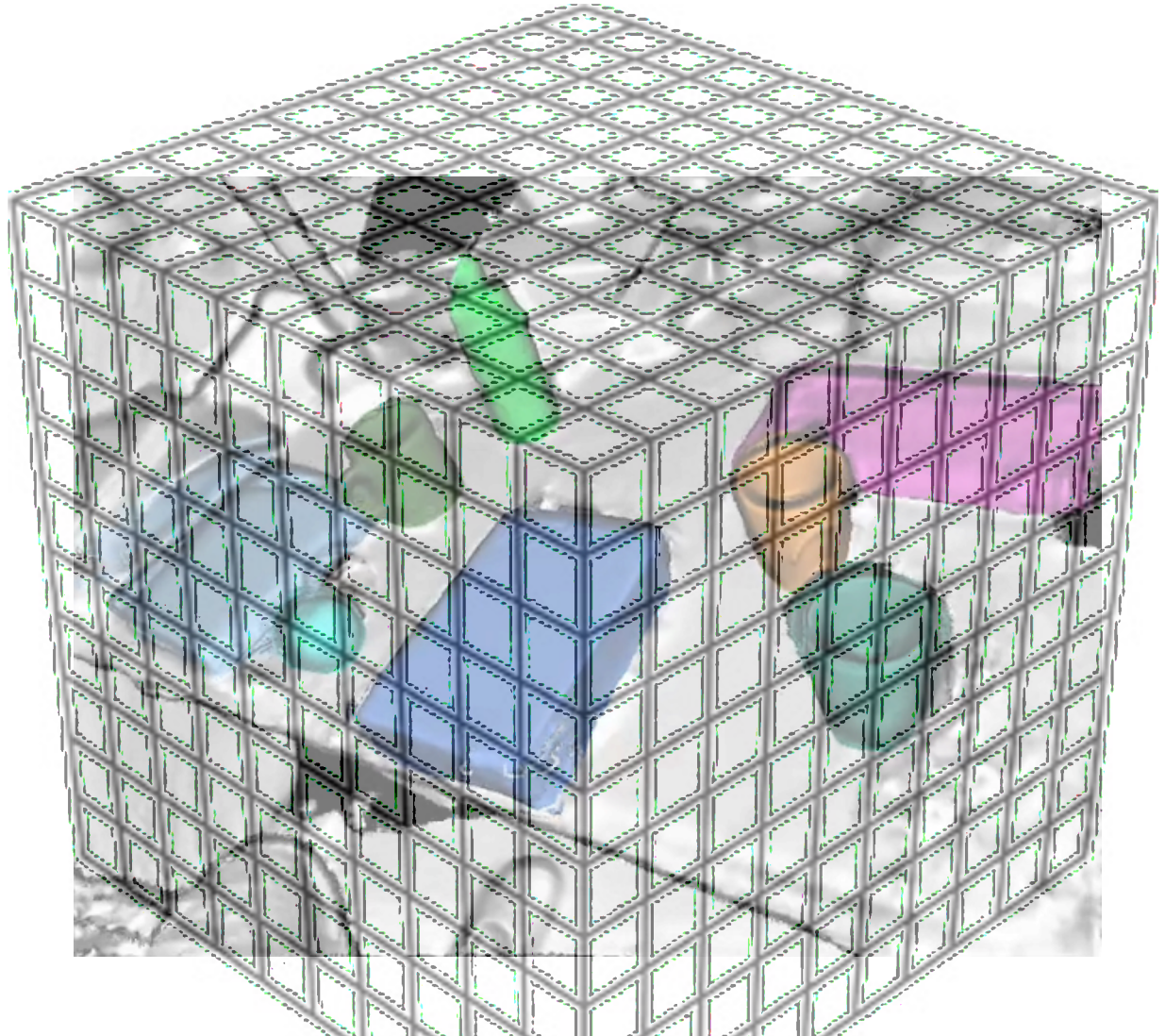
Problem #1
scalability

Each ray casting operation must be performed multiple times
once for every object model volume stored in the map

Problem #2
occlusion solving

An explicit occlusion handling strategy is required
when ray casting through multiple independent volumes

An alternative is a single volume in which TSDF values of moving objects are translated to reflect their tracked pose



A single standard TSDF volume storing the entire scene…

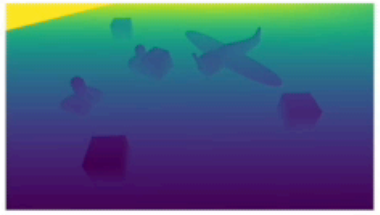
- avoids having to ray cast several volumes at each time step
- removes the need for an explicit occlusion handling strategy

but the translated TSDF values inevitably overwrite the surface previously reconstructed at the destination location

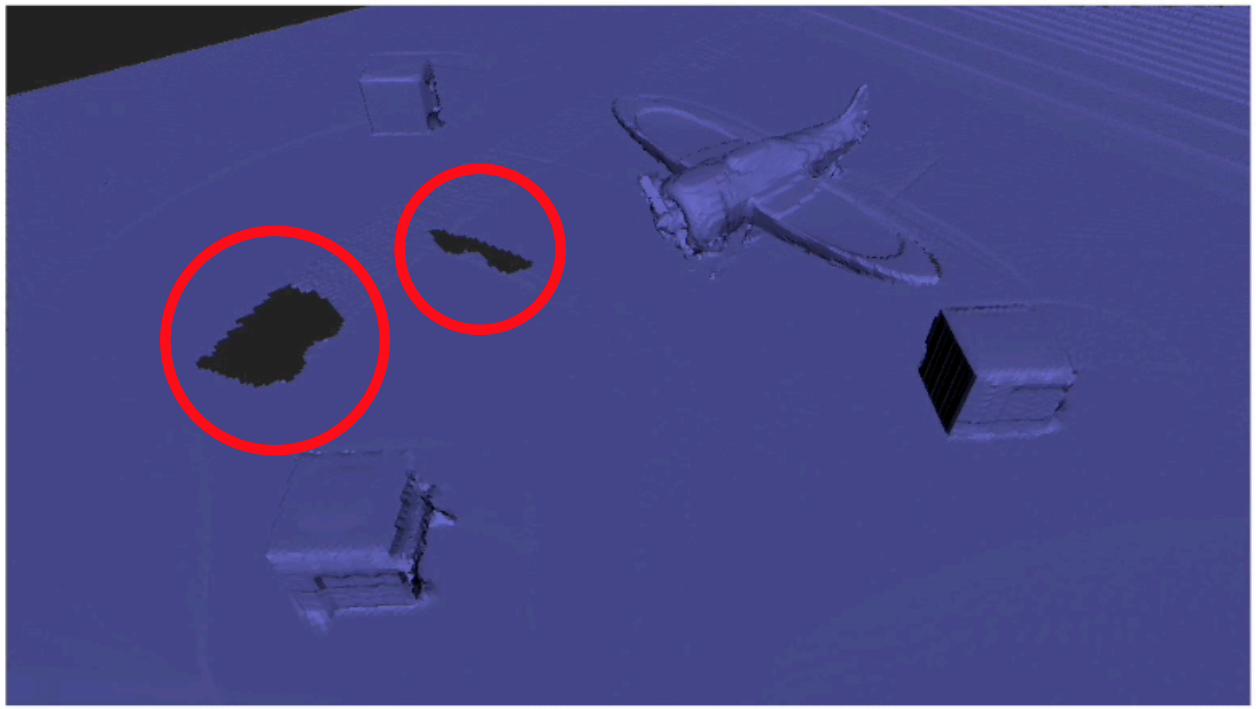
A standard TSDF map cannot withstand occlusion in a dynamic object tracking and reconstruction scenario

RGB

Depth



Mask



Traditional TSDF map

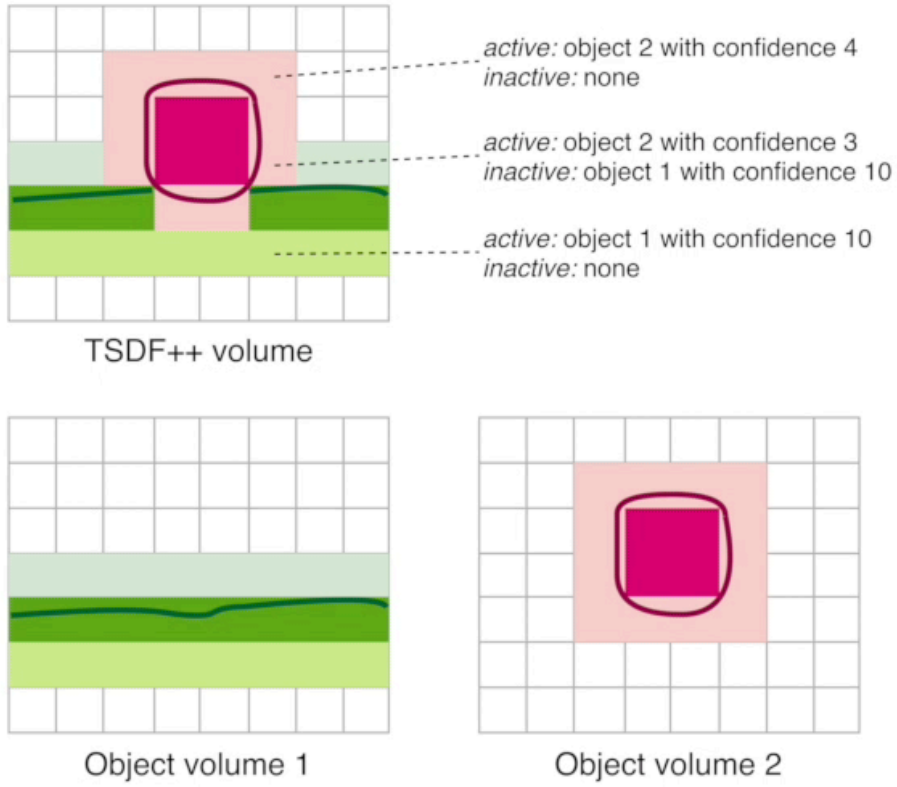
*not actual speed

Can we design a volumetric representation that…

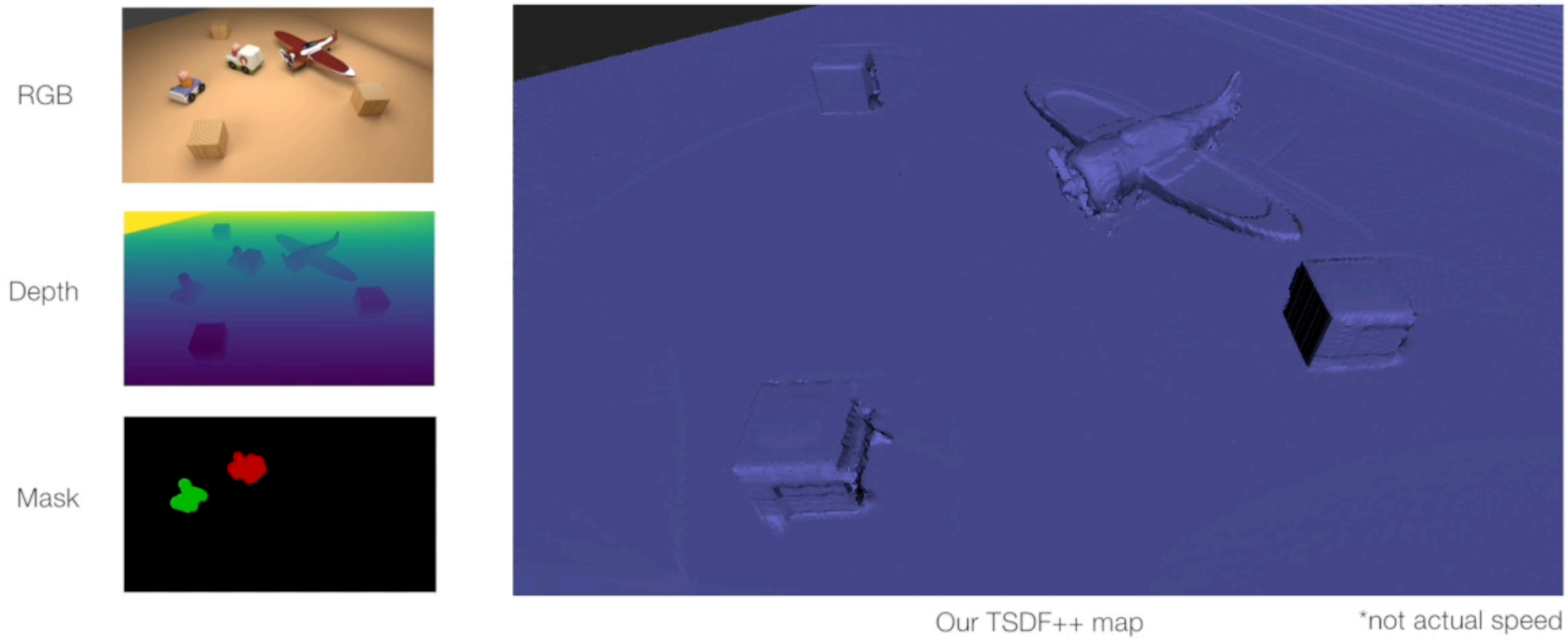
- indexes the entire scene within a single volume
- removes the need for repeated ray casting and occlusion handling
- can preserve surface reconstructions throughout occlusion?

TSDF++ is a novel multi-object formulation
for dynamic object tracking and reconstruction

TSDF++ allows simultaneous encoding at each voxel of multiple implicit object surfaces



TSDF++ preserves surface reconstructions even while they become temporarily occluded by objects moving nearby

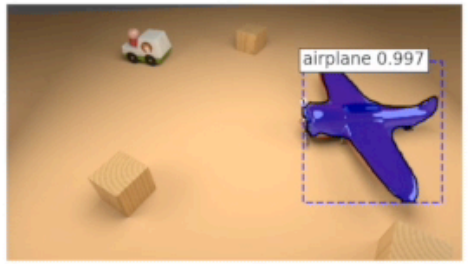


The novel formulation suffers from a few limitations

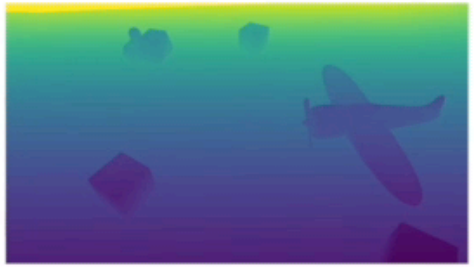
Problem #1	Reconstruction accuracy deteriorates with interpolation
interpolation	when TSDF values are translated within the object volumes
Problem #2	Memory dedicated to storing the multiple object surfaces at each voxel
robustness	rather than modeling the segmentation uncertainty

The reconstruction results are impacted by the inaccuracies or missed detections in the per-frame segmentation scheme

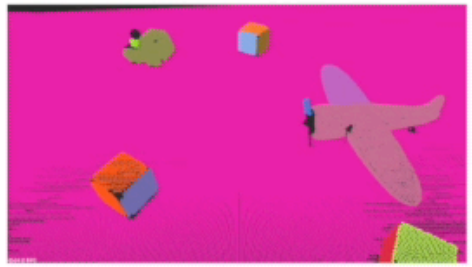
RGB



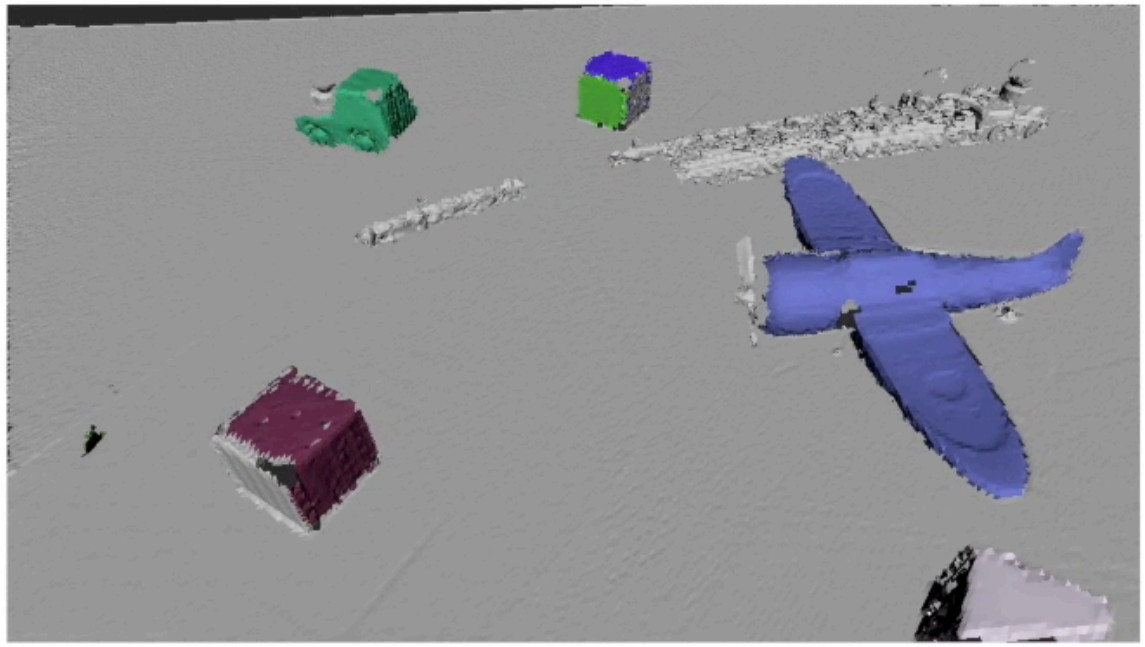
Mask R-CNN



Depth

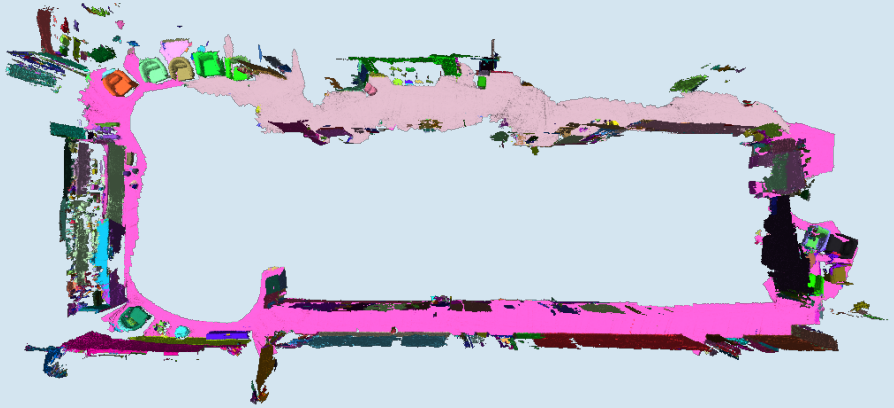


Depth segmentation



*not actual speed

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

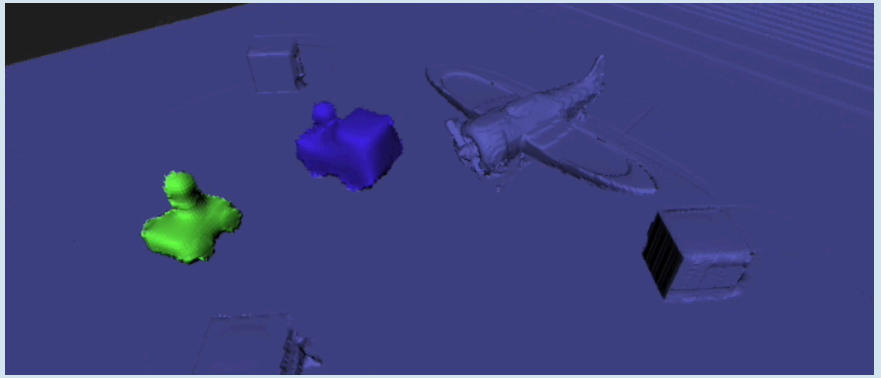
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

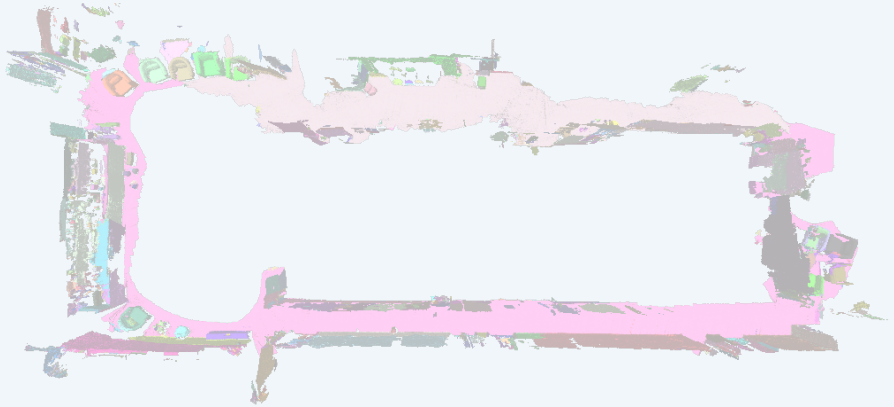


M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction

M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

The presented findings have implications for future research in the field

#1 Online object-level mapping can empower next-level robot autonomy

#2 TSDF-based depth fusion offers unparalleled advantages

Future research avenues enabled by the presented work

efficiency and robustness — Validating “objectness” and robust temporal tracking

affordances

transparent objects

Future research avenues enabled by the presented work

efficiency and robustness

affordances — Parsing of actionable attributes of objects

transparent objects

Future research avenues enabled by the presented work

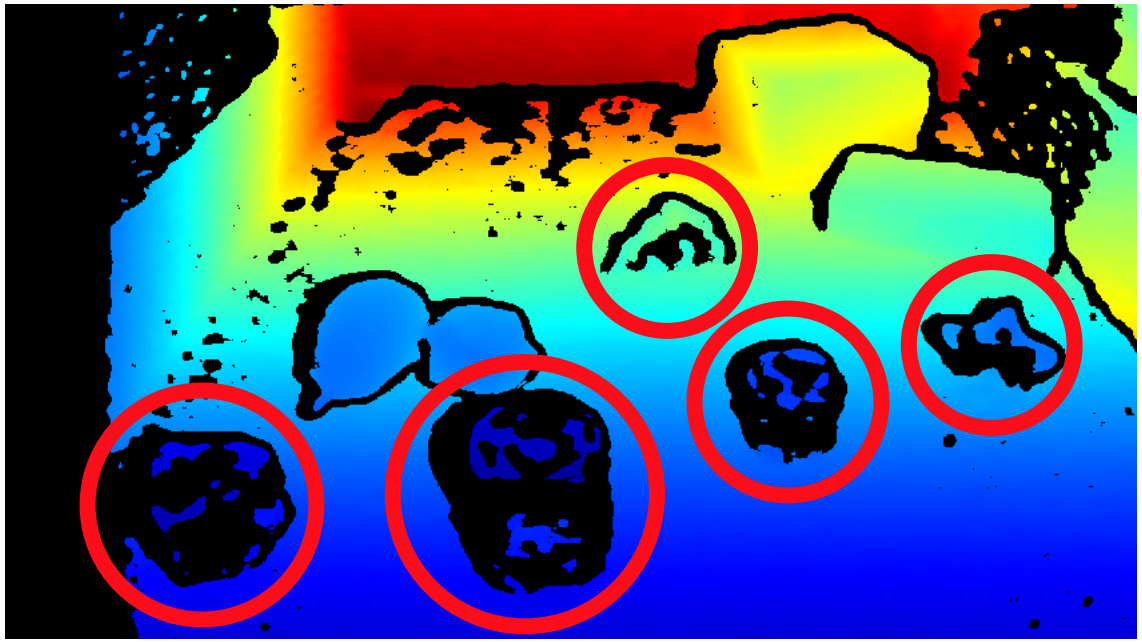
efficiency and robustness

affordances

transparent objects

Object-aware 3D reconstruction of transparent surfaces

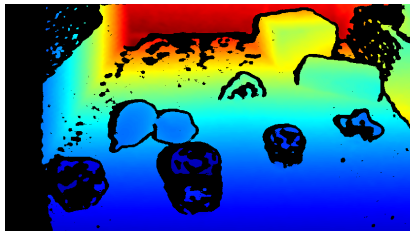
RGB



Depth

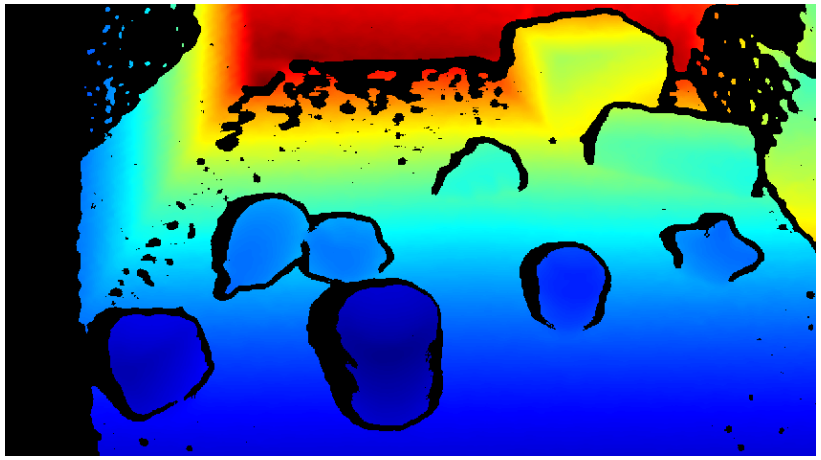
Beyond reconstructing the visible geometry, is it possible to reconstruct “invisible” transparent surfaces?

RGB

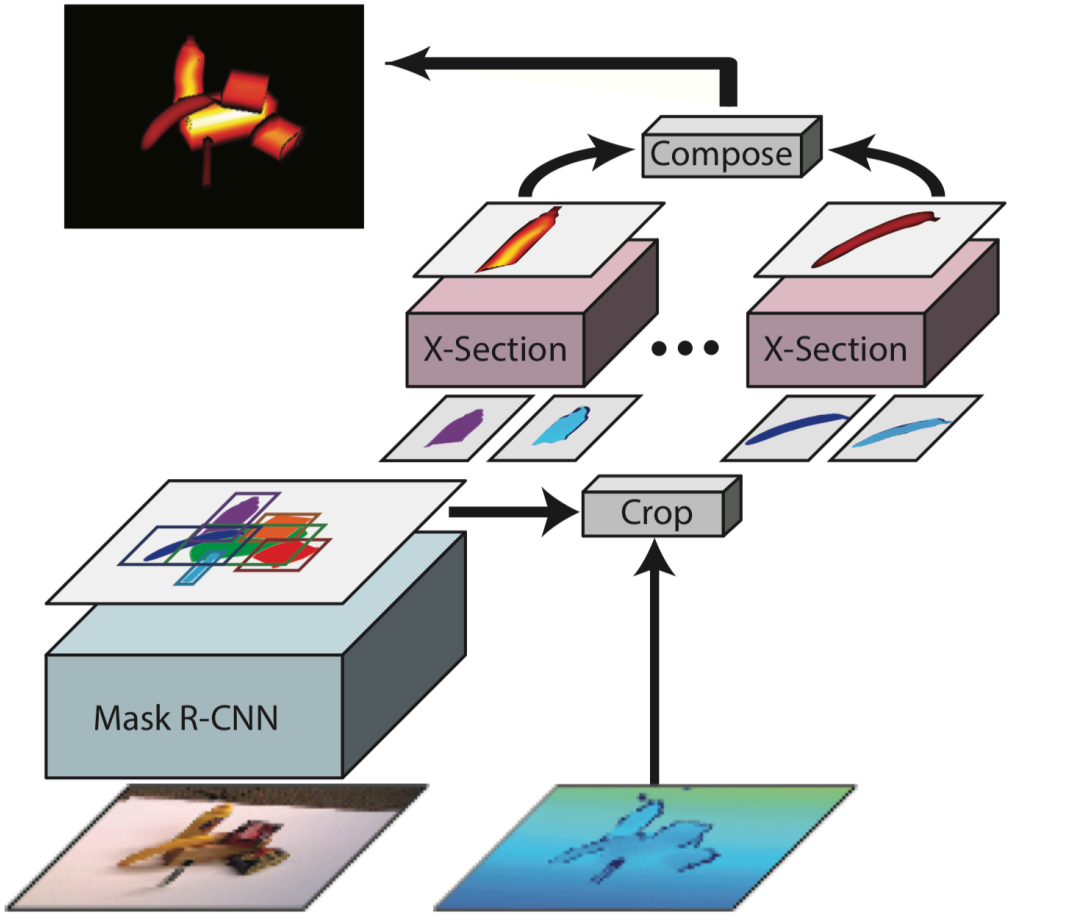


Depth

Sajjan et al., ClearGrasp, 2019



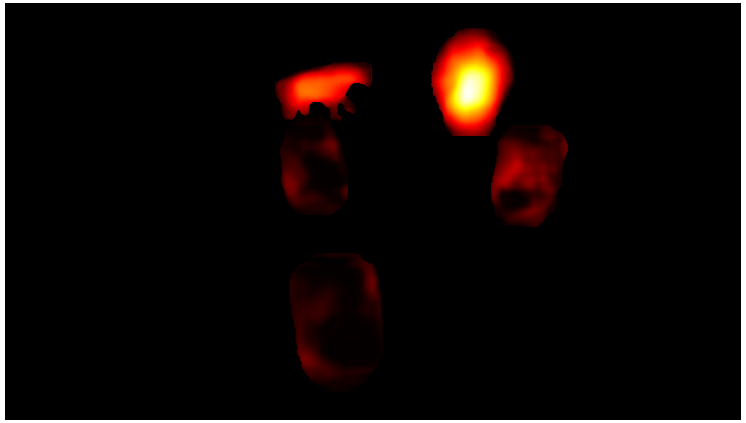
Refined depth



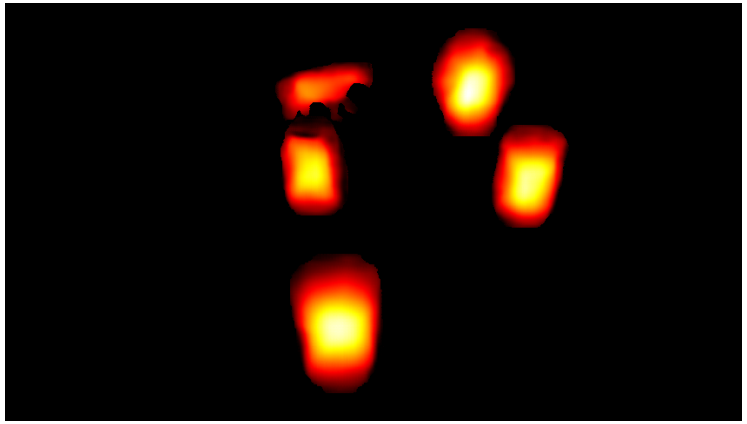
Nicastro et al., X-Section, 2019

Preliminary results hint at the fact that thickness prediction can be solved solely from monocular RGB

Mask R-CNN



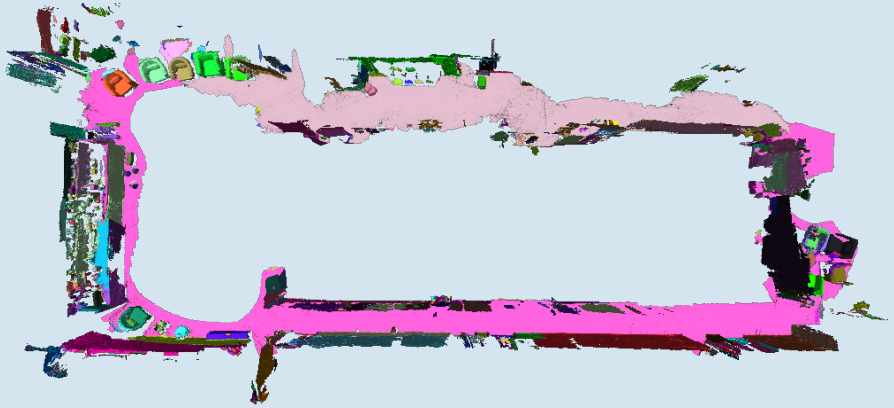
Thickness prediction from mask and transparent depth



Thickness prediction from mask and ground truth depth

Thickness prediction from mask only

Dense object-level mapping



M. Grinvald, F. Furrer, T. Novkovic, J. J. Chung, C. Cadena, R. Siegwart, and J. Nieto. Volumetric instance-aware semantic mapping and 3D object discovery. RA-L, 2019

★ IROS Best Paper award on Cognitive Robotics Finalist ★

Online object database building

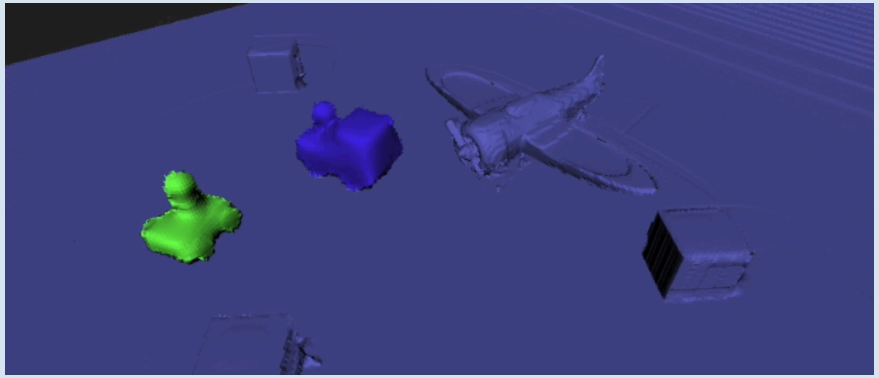
F. Furrer, T. Novkovic, M. Fehr, A. Gawel, M. Grinvald, T. Sattler, R. Siegwart, and J. Nieto. Incremental Object Database: Building 3D models from multiple partial observations. IROS, 2018

Dataset of common household objects

T. Novkovic, F. Furrer, M. Panjek, M. Grinvald, R. Siegwart, and J. Nieto. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. IJRR, 2019

Dynamics

Dynamic object tracking and reconstruction



M. Grinvald, F. Tombari, R. Siegwart, and J. Nieto. TSDF++: A multi-object formulation for dynamic object tracking and reconstruction. ICRA, 2021

Outlook

Future research avenues

It takes a village to raise a PhD

Prof. Dr. Roland Siegwart

Prof. Dr. Stefan Leutenegger

Dr. Cesar Cadena

Dr. Jen Jen Chung

Dr. Juan Nieto

PD Dr. Federico Tombari

Luciana Borsatti

Cornelia Della Casa

Michael Riner

Michel Breyer

Kenneth Blomqvist

Andrei Cramariuc

Marius Fehr

Julian Förster

Fadri Furrer

Abel Gawel

Tonci Novkovic

Lionel Ott

Marko Panjek

Florian Tschopp

Alberto Dall'Olio

Pietro Griffa

Cyrill Hedinger

Timothy Marvel

Mattia Segù

…

ABB

Amazon

Facebook

Google