Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

MIN Faculty
Department of Informatics

Fusion++
Volumetric Object-Level SLAM

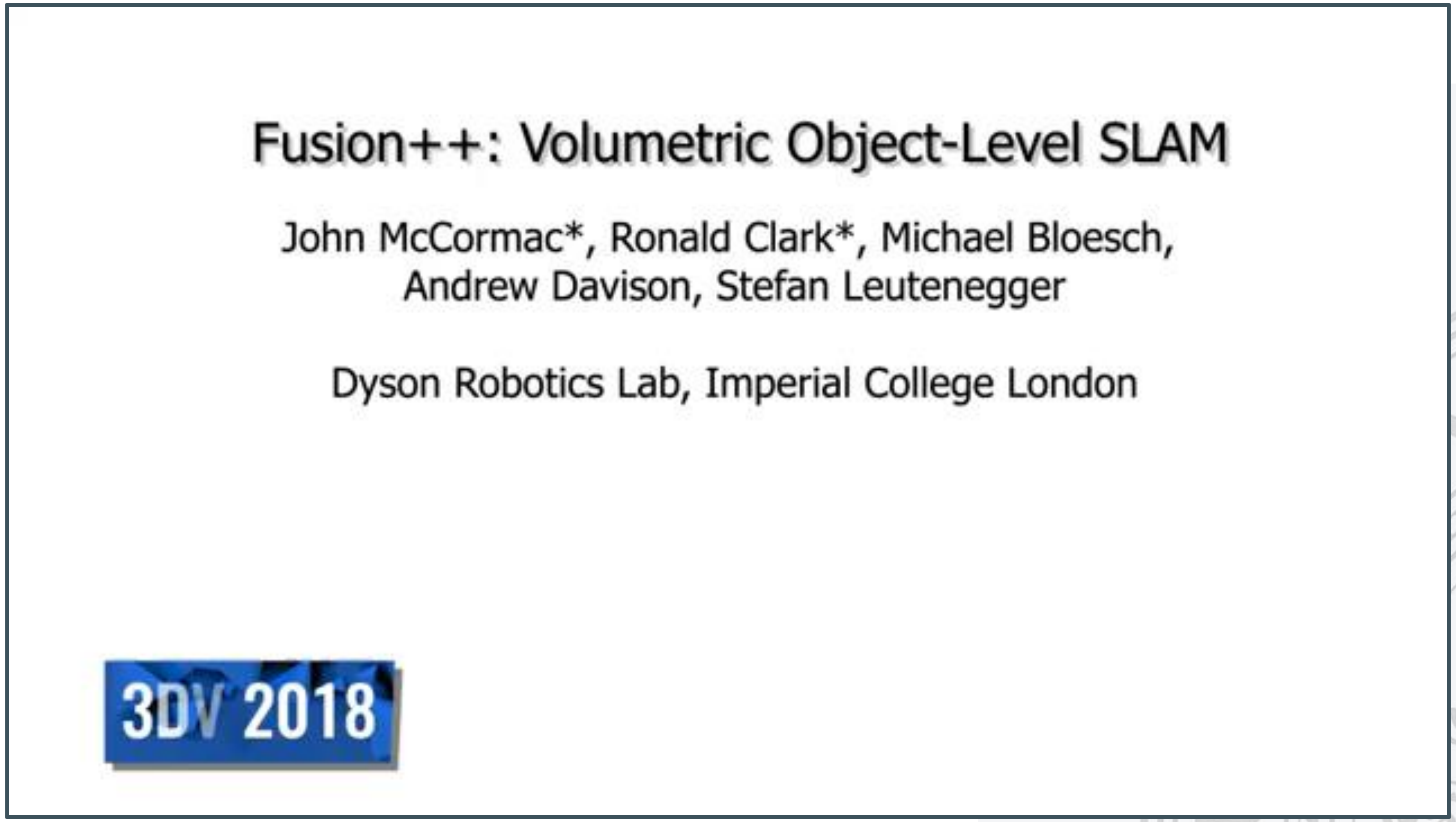
Lasse Haffke

University of Hamburg
Faculty of Mathematics, Informatics and Natural Sciences
Department of Informatics

Technical Aspects of Multimodal Systems

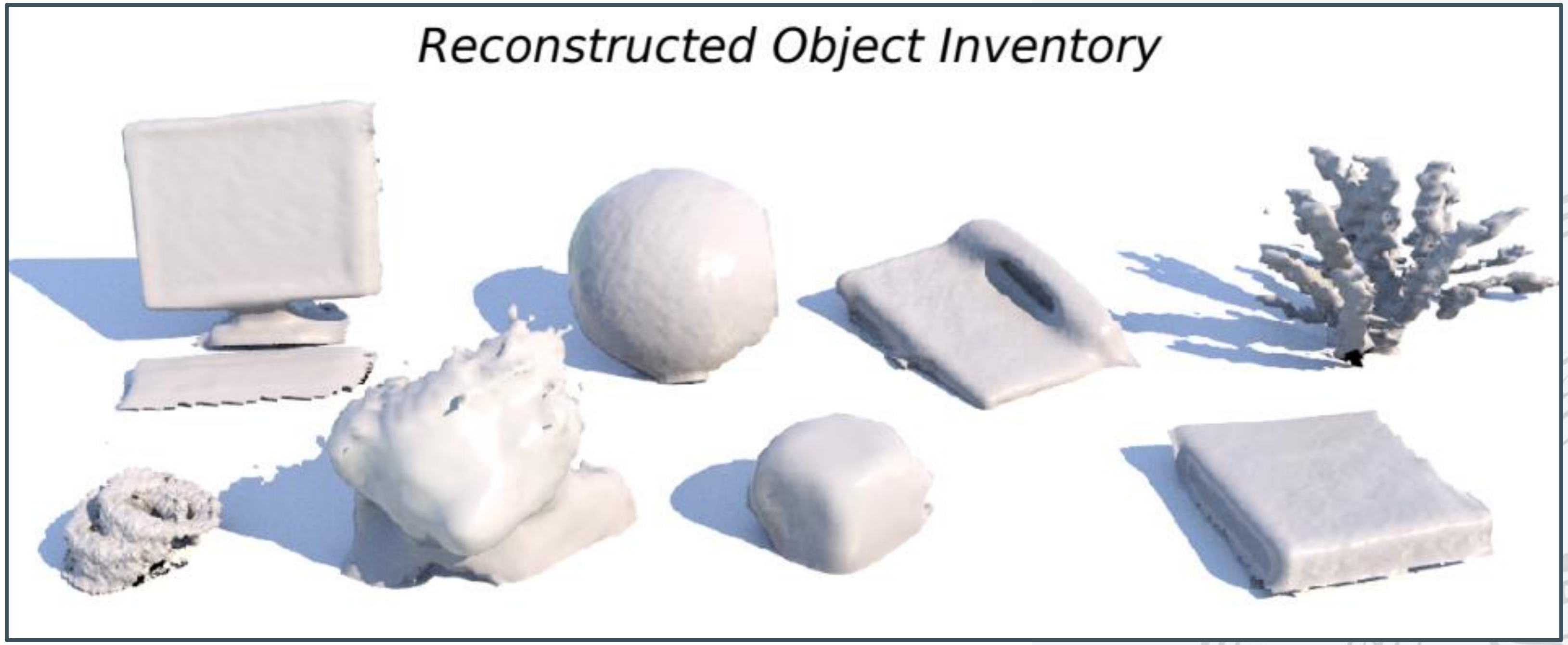
25.06.2020

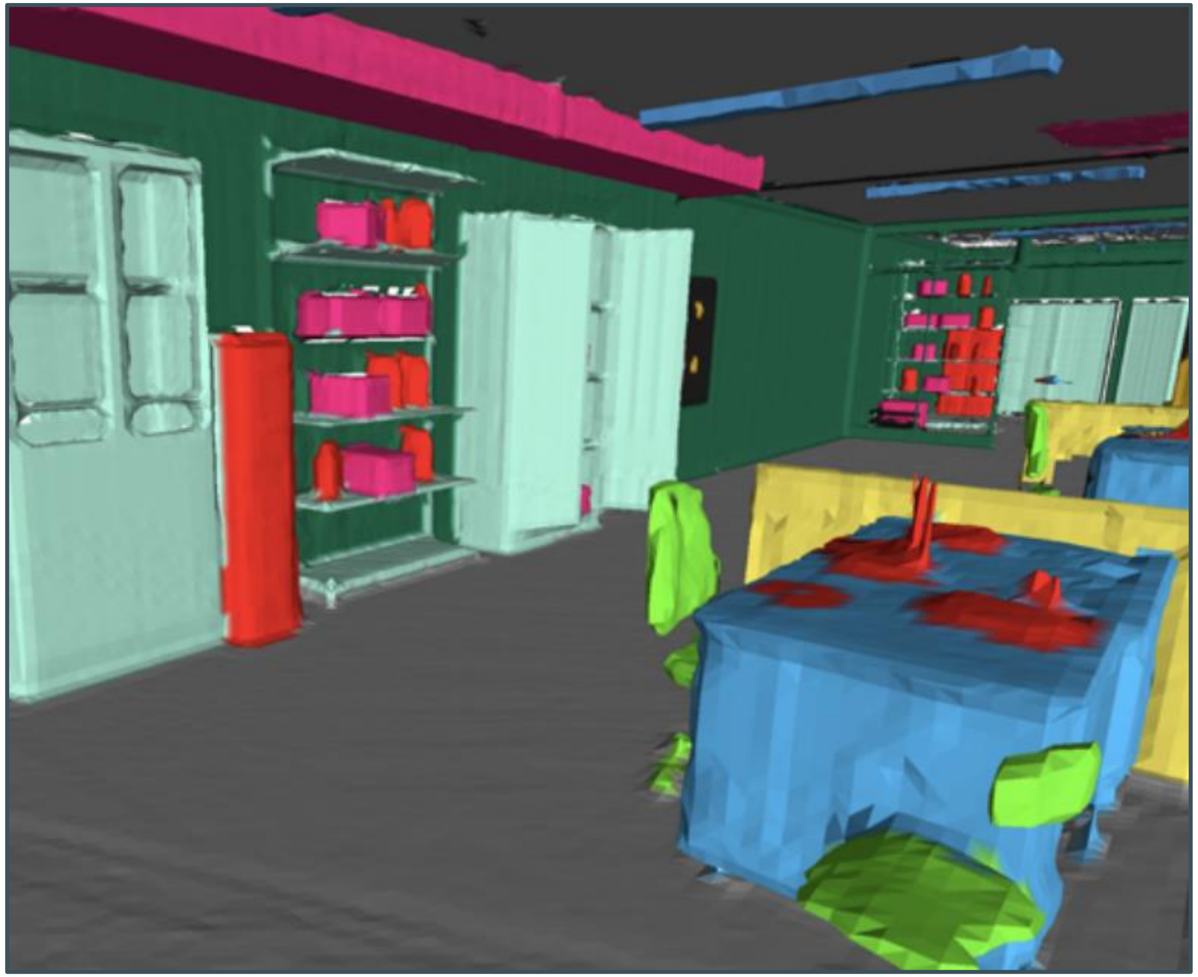
Intro



https://www.youtube.com/watch?v=2luKNC03x4k

Fusion++





Kimera

Fusion++

Kimera:

One dense mesh

Pixel to object label

Navigation, Obstacle avoidance

Fusion ++:

Seperate object instances

Object-centric map

Communication, Interaction

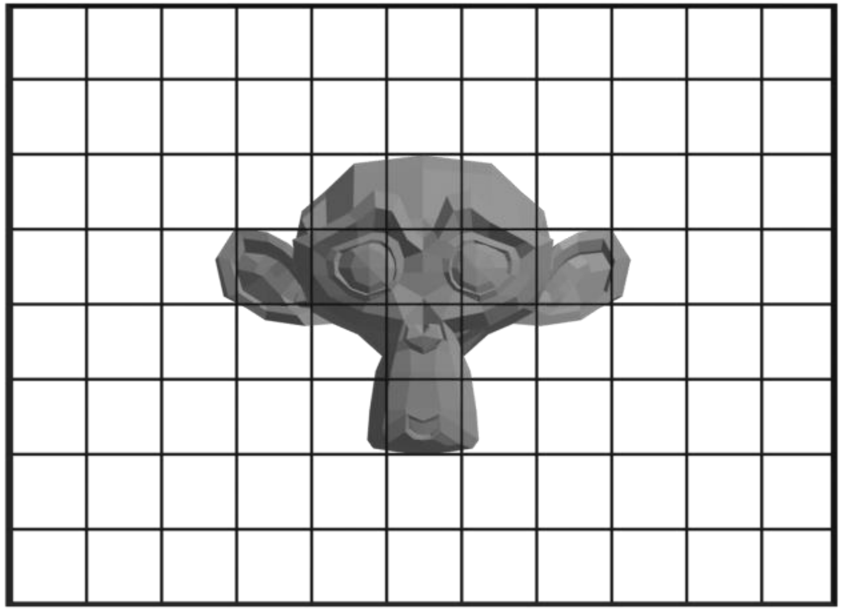
TSDF

Raycasting

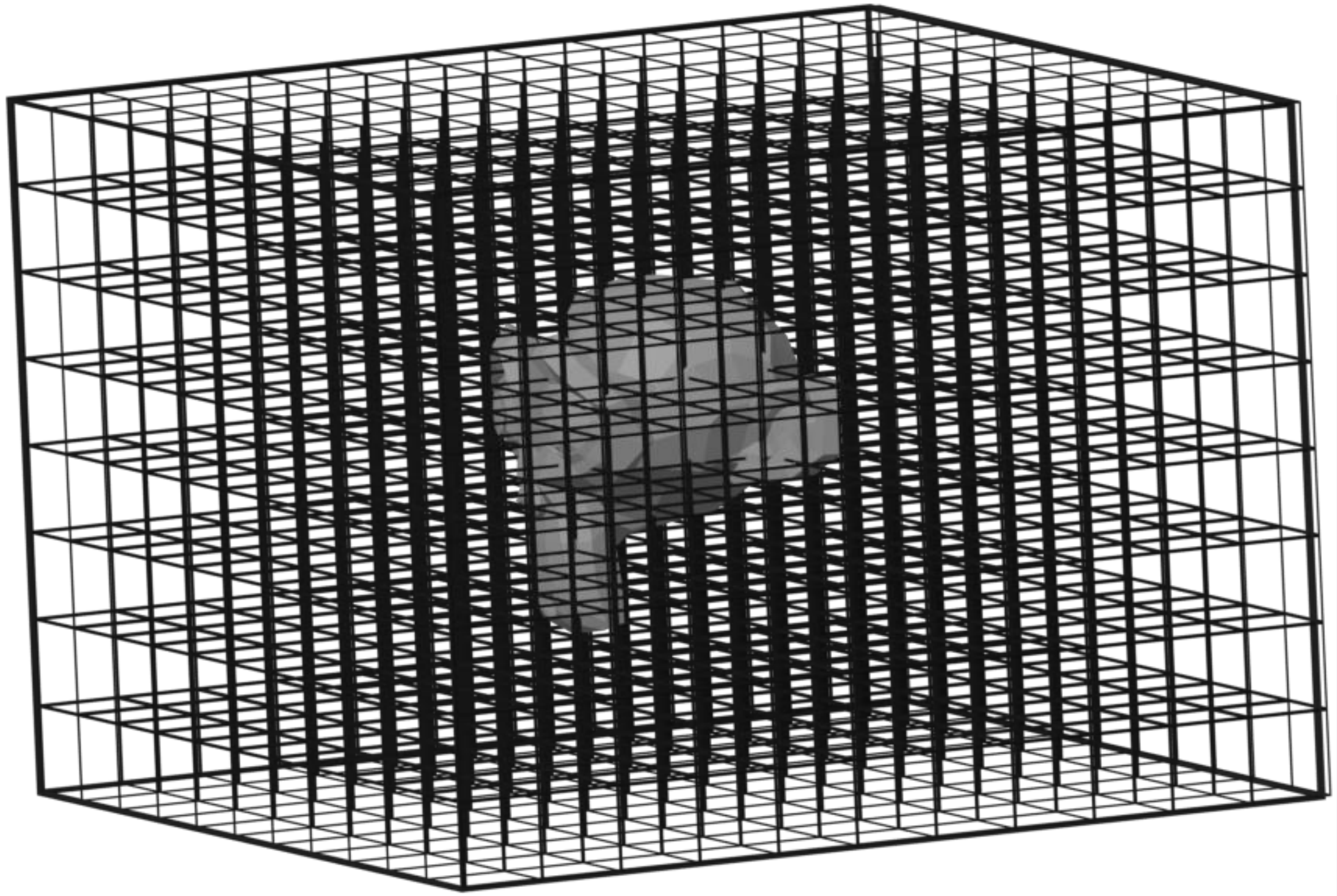
Object detection

TSDF

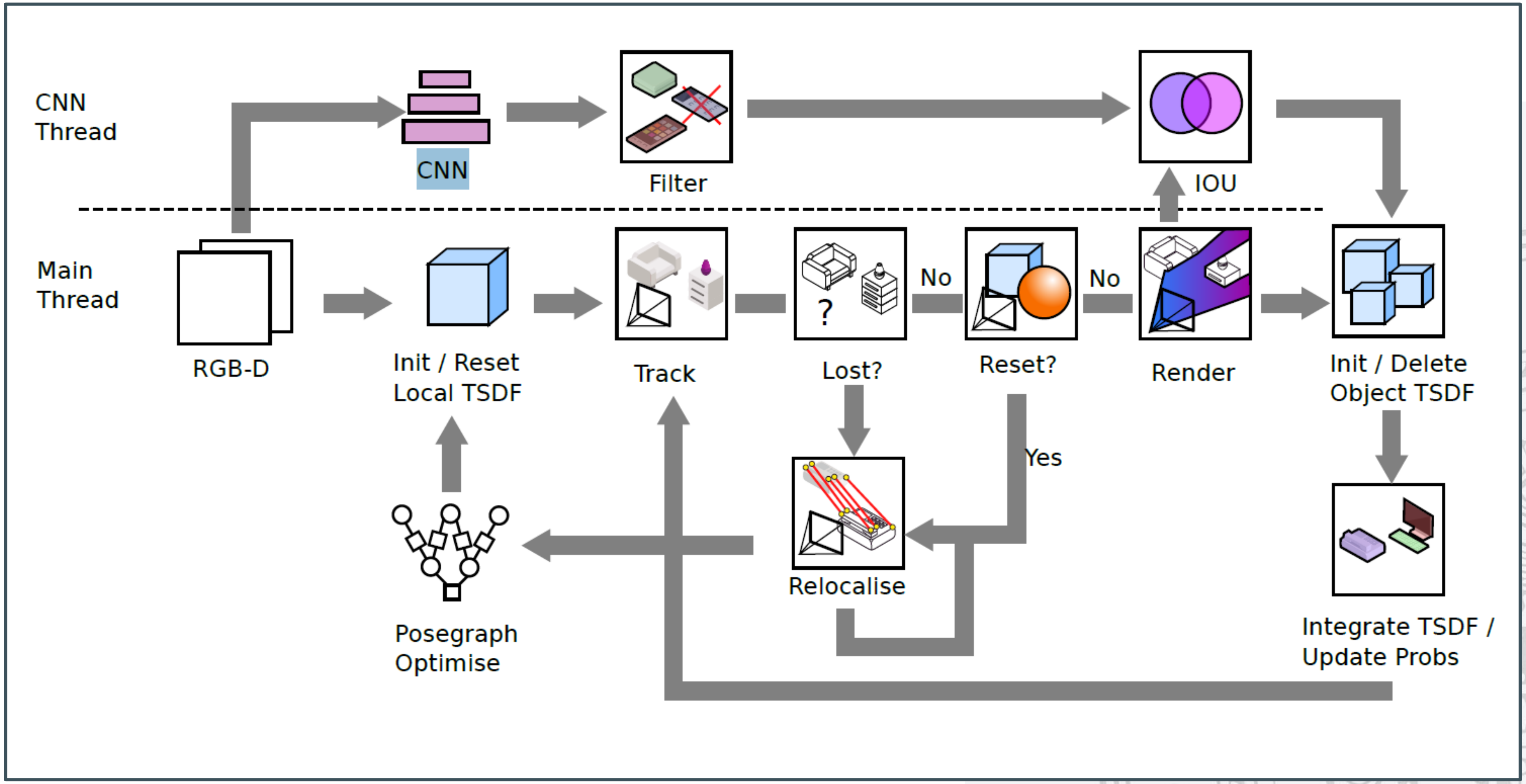
Truncated signed distance function:



TSDF



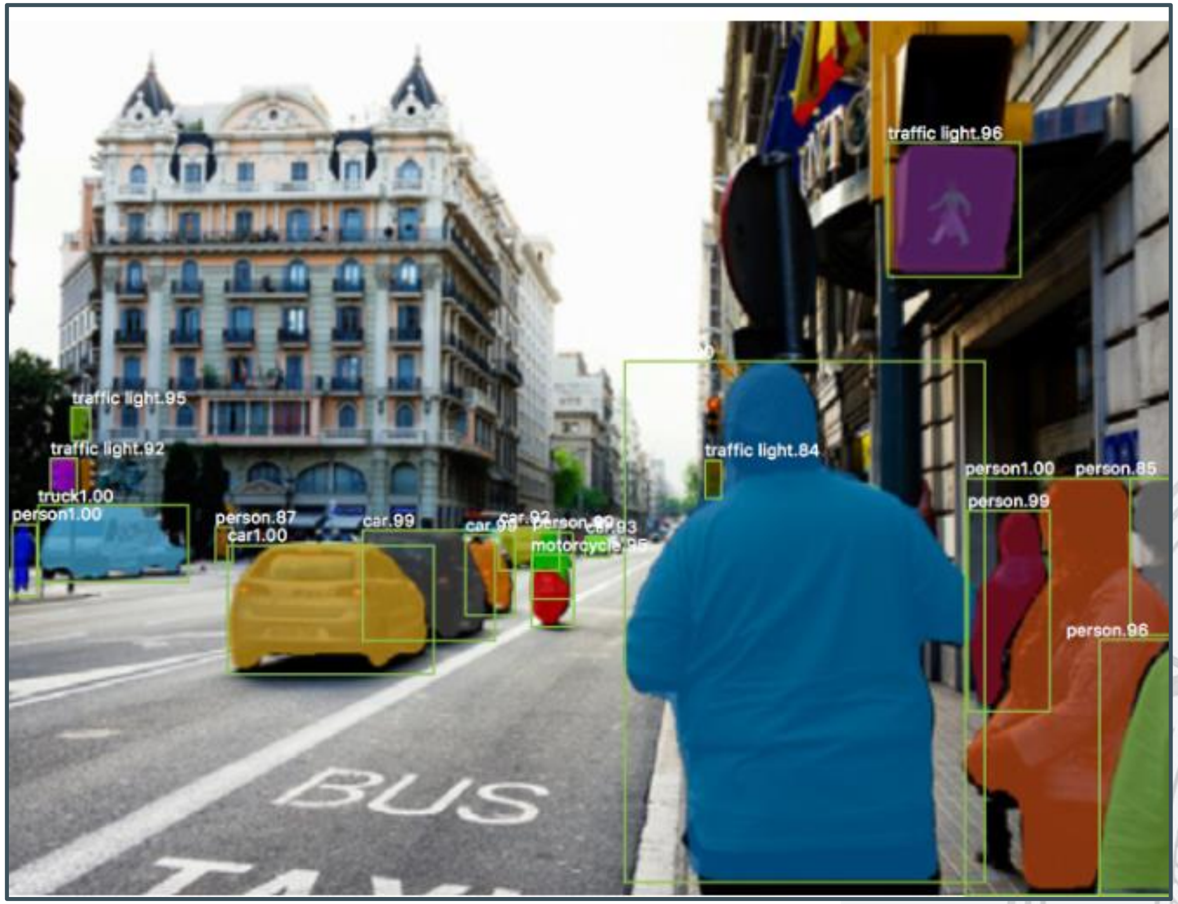
Method


CNN
Thread
CNN
Filter
IOU
Main
Thread
RGB-D
Init / Reset
Local TSDF
Track
Lost?
No
Reset?
No
Render
Init / Delete
Object TSDF
?
Yes
Relocalise
Posegraph
Optimise
Integrate TSDF /
Update Probs

Method

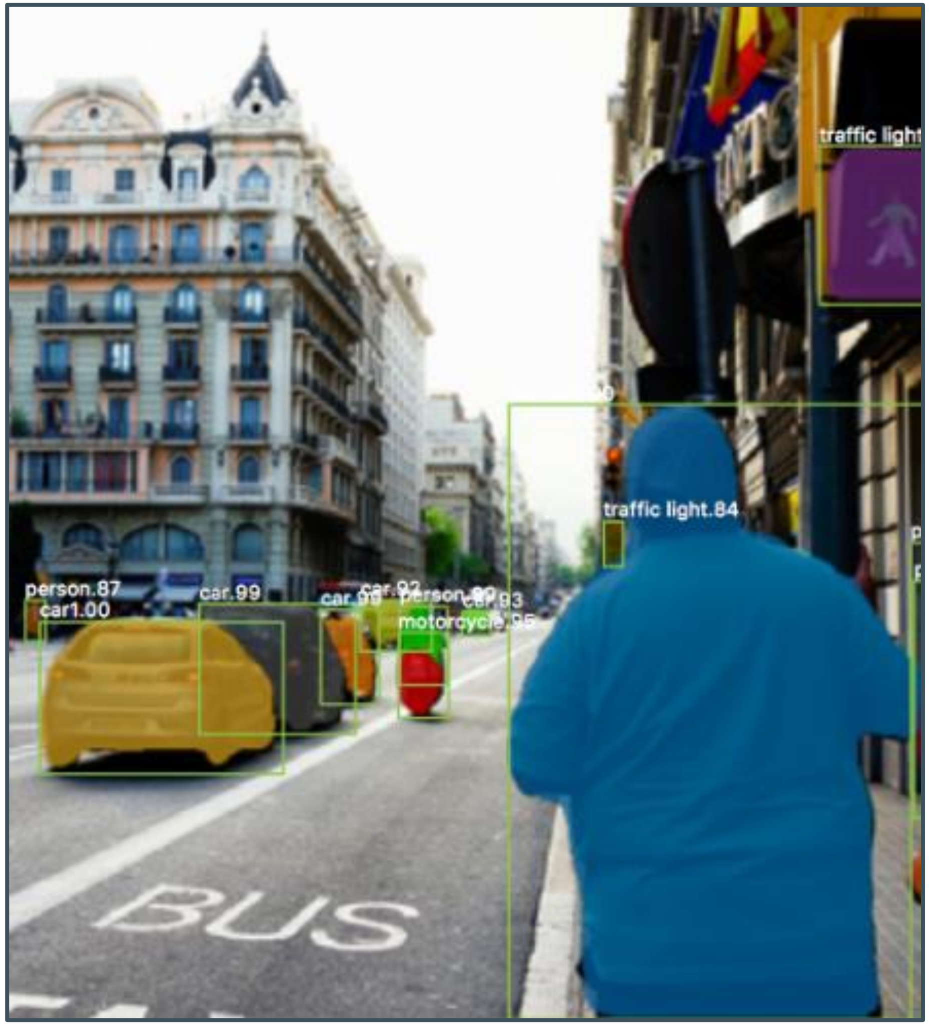
1. Mask-RCNN on a seperate thread
2. Track background
3. Localize camera
4. Initialize new objects
5. Integrate into / update existing objects

Mask-RCNN



HE, Kaiming, et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. 2017. S. 2961-2969.

Mask-RCNN



Mask-RCNN:

- Extension of FASTER-RCNN
- Output exclusive binary masks
- Bottleneck

TSDF Instances

Overlay binary mask into depth map to get a point cloud

Take 10-90 percentile. Create a cube from that edge length.

Sample that cube at 64 steps per edge.

Squirrel 89

Take the [10-90] percentile

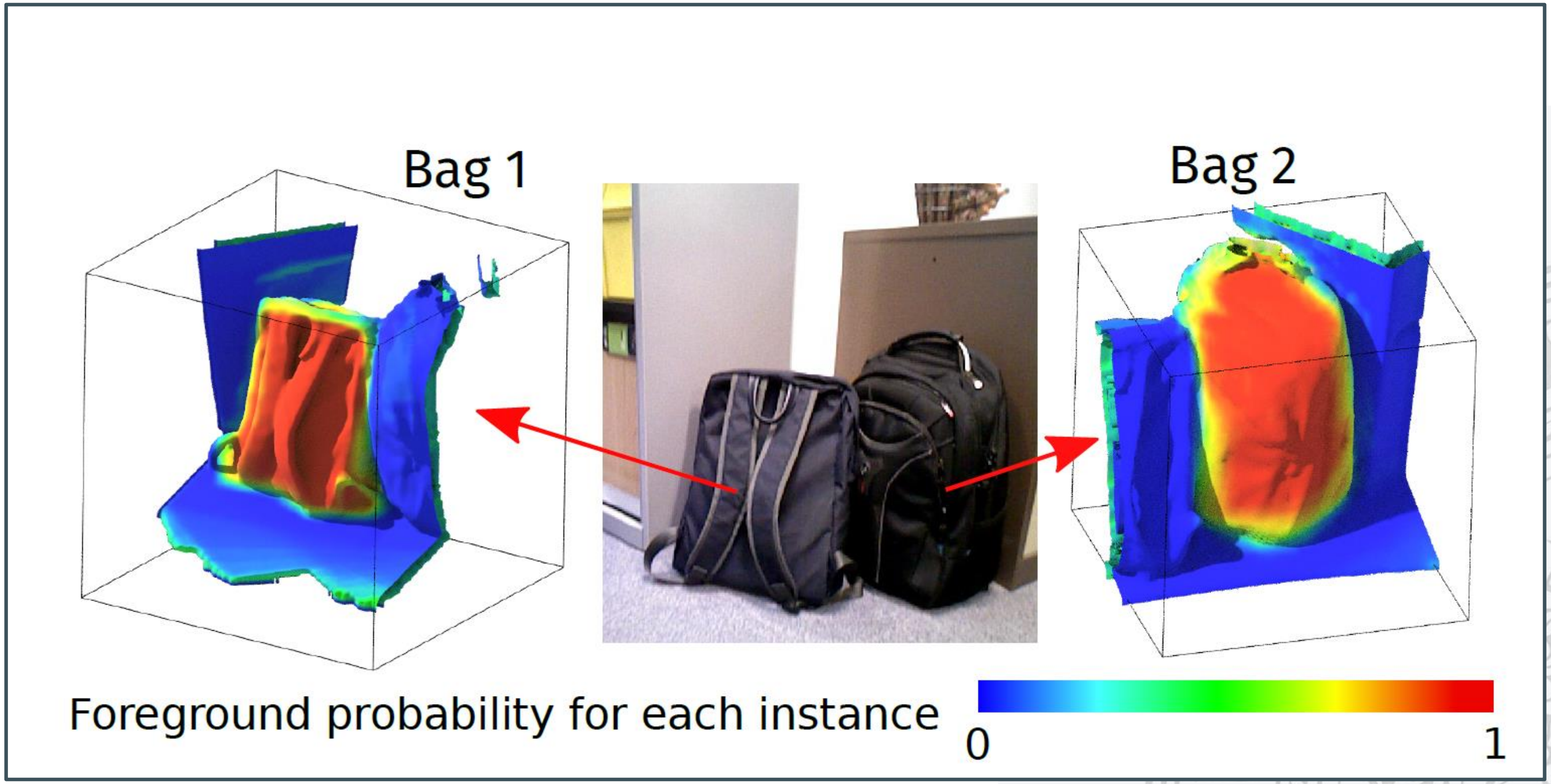
[30-245]

Foreground probability

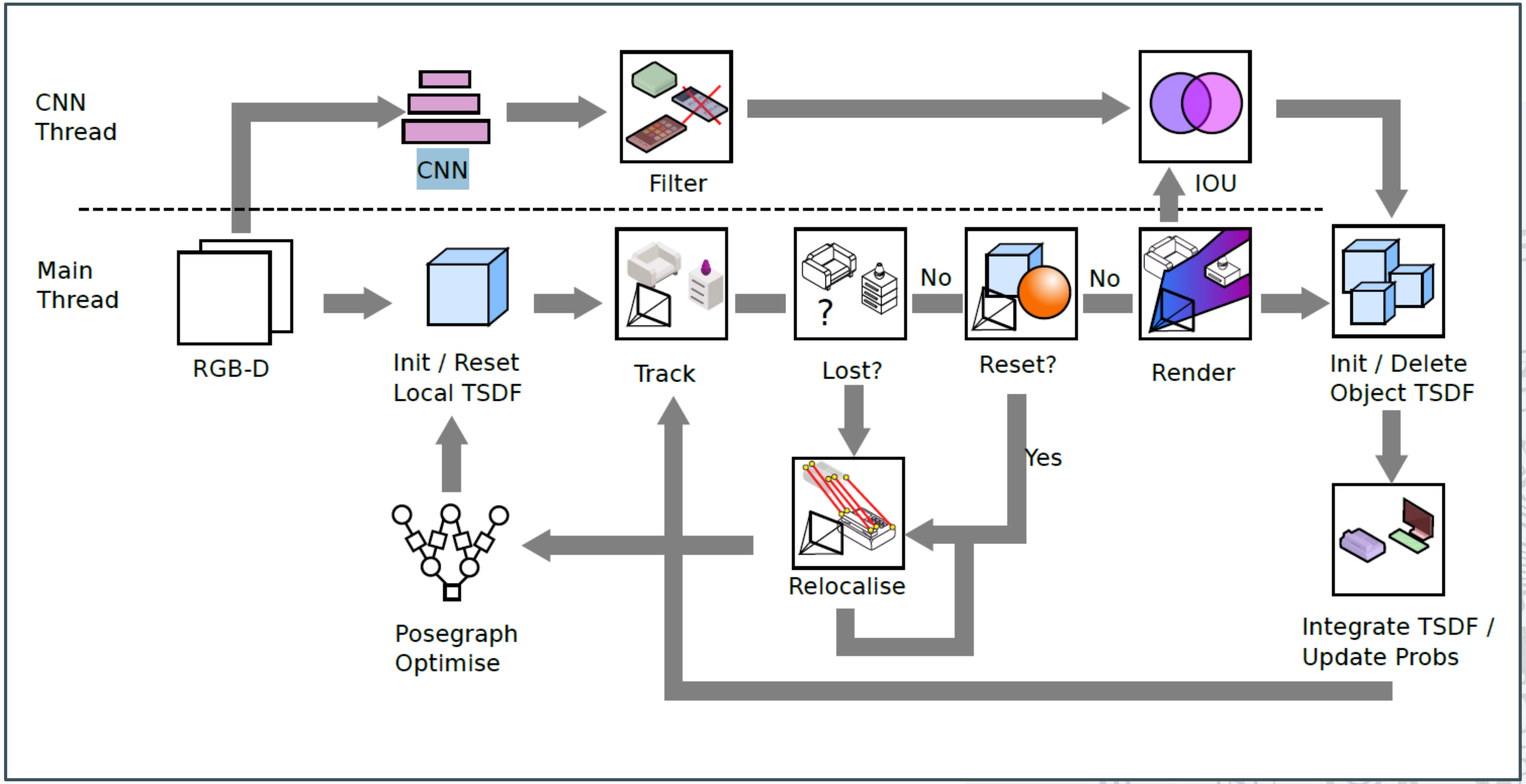
Surface integration is performed over the entire volume

Overlay masks from different frames as foreground probability

Foreground probability

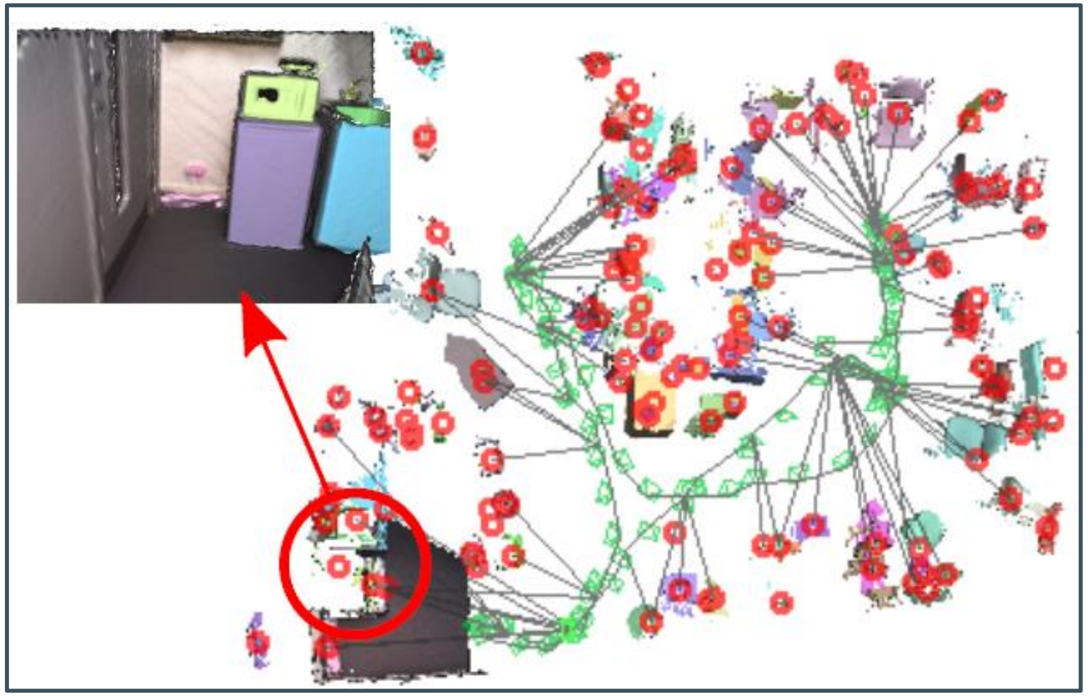


Method


CNN
Thread
CNN
Filter
IOU
Main
Thread
RGB-D
Init / Reset
Local TSDF
Track
Lost?
No
Reset?
No
Render
Init / Delete
Object TSDF
?
Yes
Relocalise
Posegraph
Optimise
Integrate TSDF /
Update Probs

How do we position our objects?

Posegraph



Object-centric map:

Pose for each ask-RCNN frame

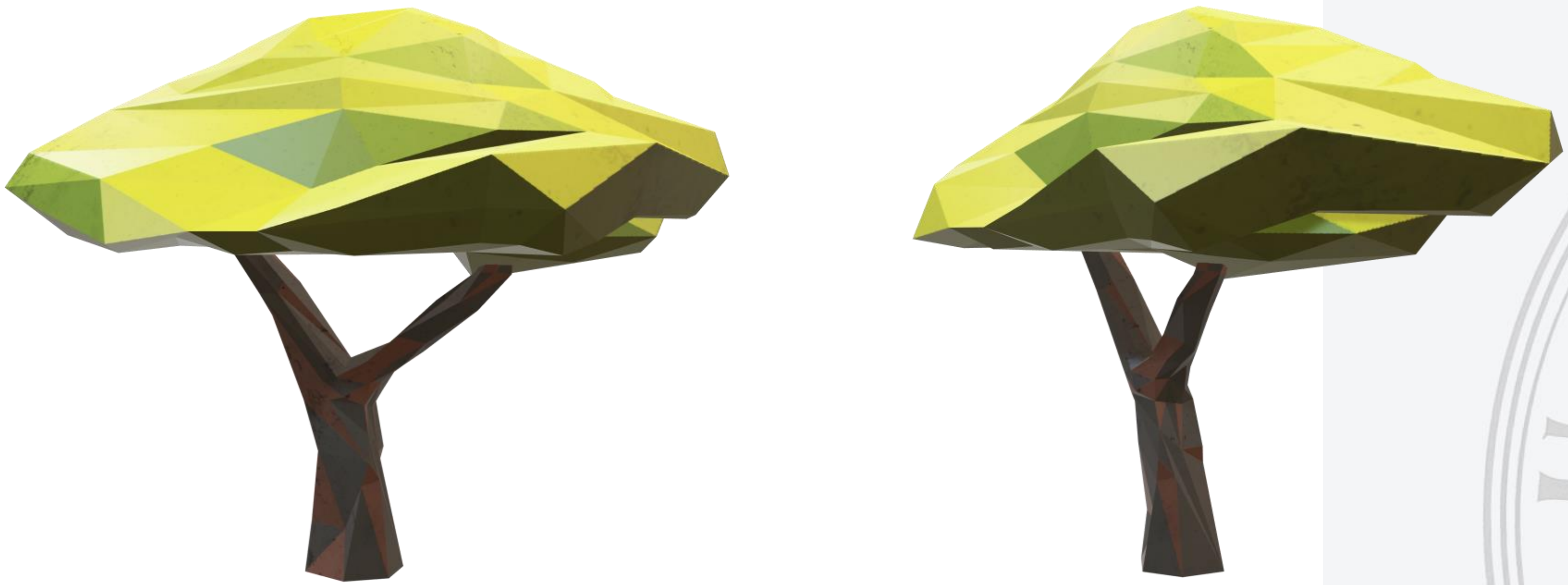
Red circle for each object instance

Loop closure after full loop

ICP-Tracking

Iterative closest point tracking:

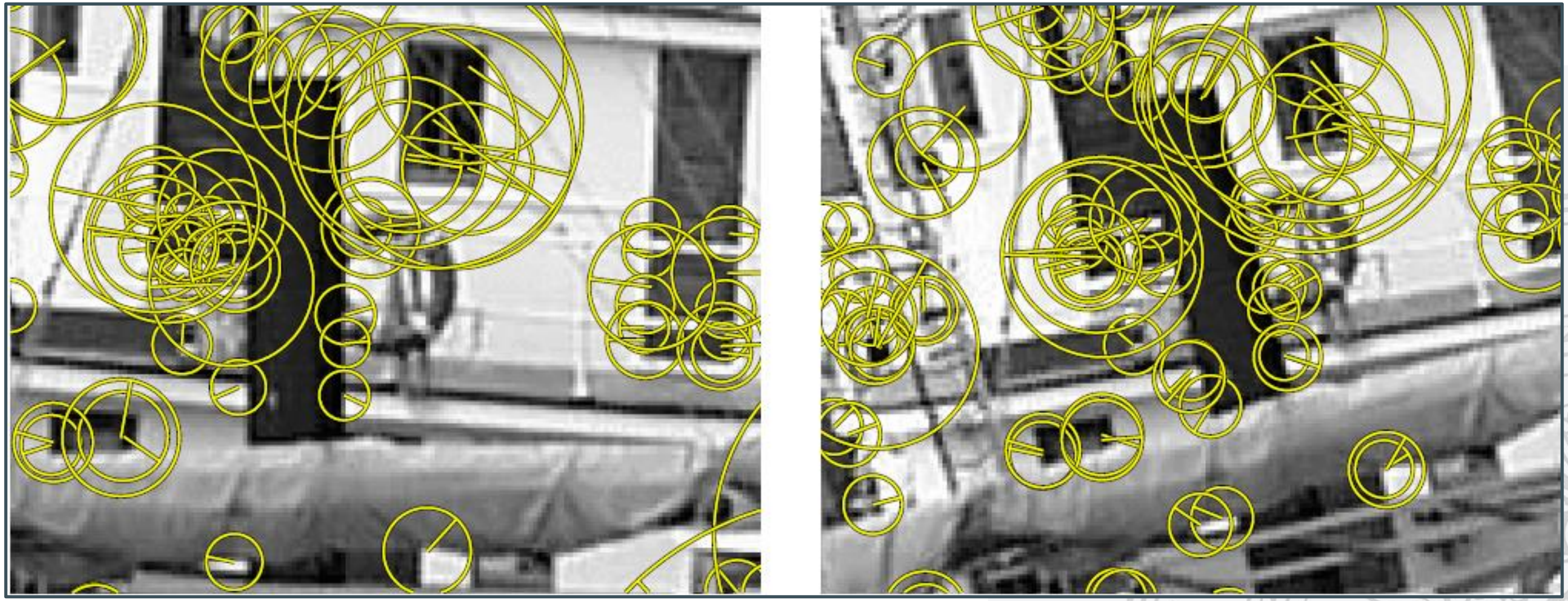
Match two point clouds



BRISK: Binary Robust Invariant Scalable Keypoints

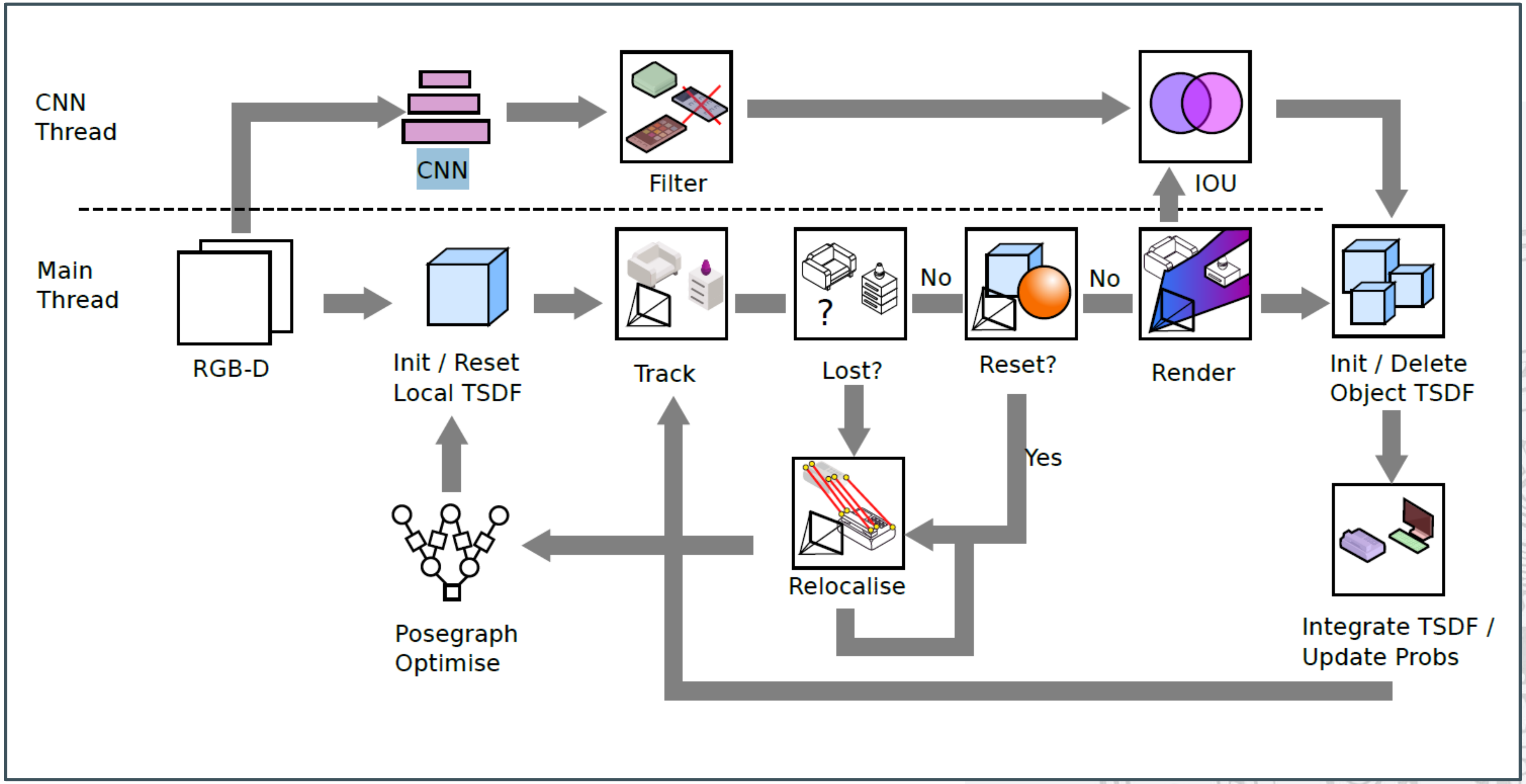
- Detect possible keypoints by 9-16
- Apply to every octave
- Pool nearest maxima
- Sample the location of the keypoint

BRISK



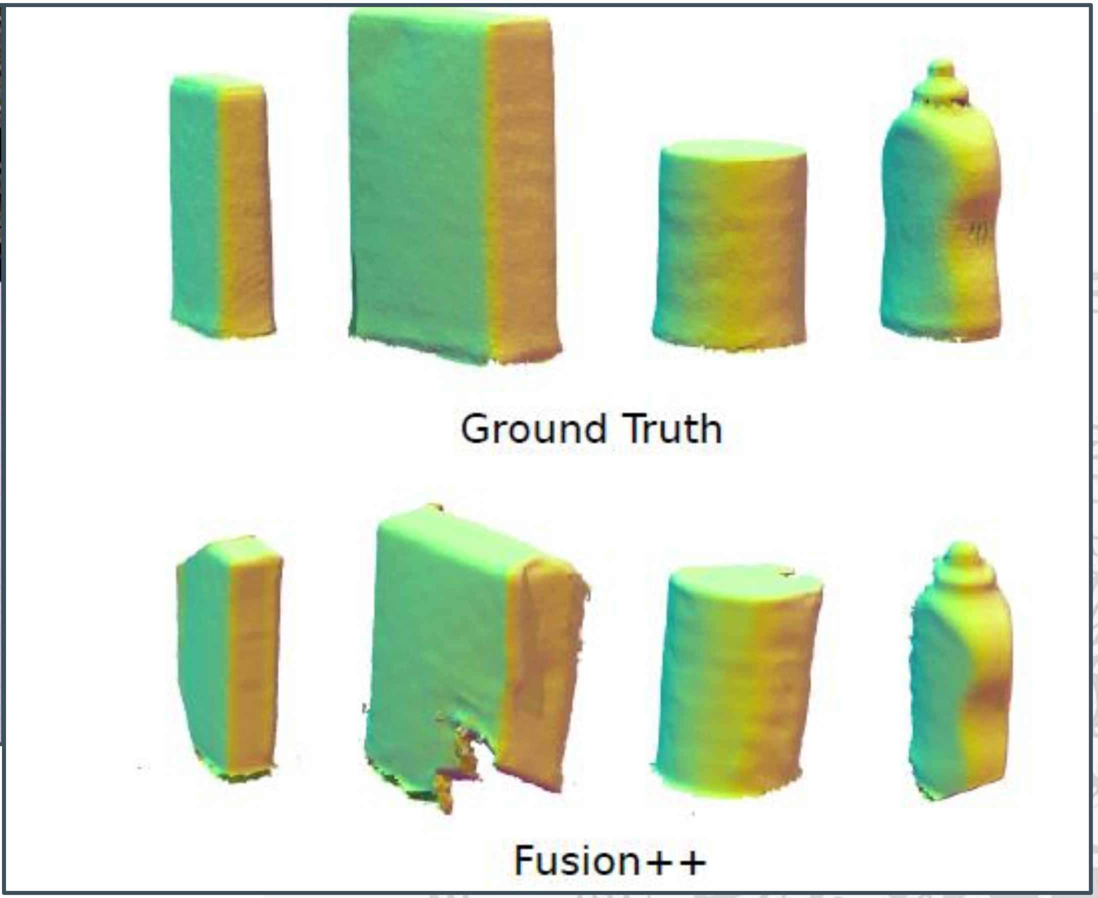
S. Leutenegger, M. Chli, and R. Siegwart. BRISK: Binaryrobust invariance scalable keypoints. In Proceedings of theInternational Conference on Computer Vision (ICCV), 2011.

Method


CNN
Thread
CNN
Filter
IOU
Main
Thread
RGB-D
Init / Reset
Local TSDF
Track
Lost?
No
Reset?
No
Render
Init / Delete
Object TSDF
?
Yes
Relocalise
Posegraph
Optimise
Integrate TSDF /
Update Probs

Comparison

YCB-Dataset Sequence 0001



Shortcomings

Sub-realtime performance

Static environment

Occlusion failures

Framerate dependency

Conclusion

Used techniques:

- Mask R-CNN
- TSDF
- ICP
- BRISK

This is one of the first approach to object-centric SLAM.

In terms of not strictly object-centric metrics, but general RGB-D SLAM metrics, it performs worse than state-of-the-art systems.

Thank you for your attention

References

B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar. The YCB object and Model set: Towards common benchmarks for manipulation research. In International Conference on Advanced Robotics (ICAR), pages 510–517, 2015.

B. Curless and M. Levoy. A volumetric method for building complex models from range images. In Proceedings of SIGGRAPH, 1996.

S. Leutenegger, M. Chli, and R. Siegwart. BRISK: Binary Robust Invariance Scalable Keypoints. In Proceedings of the International Conference on Computer Vision (ICCV), 2011.

R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon. KinectFusion: Real-Time Dense Surface Mapping and Tracking. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), 2011.

Rosinol, A., Abate, M., Chang, Y., and Carlone, L. (2019a). Kimera: an Open-Source Library for Real-Time Metric-Semantic Localization and Mapping.

HE, Kaiming, et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*. 2017. S. 2961-2969.

TSDF - Integration

RCNN