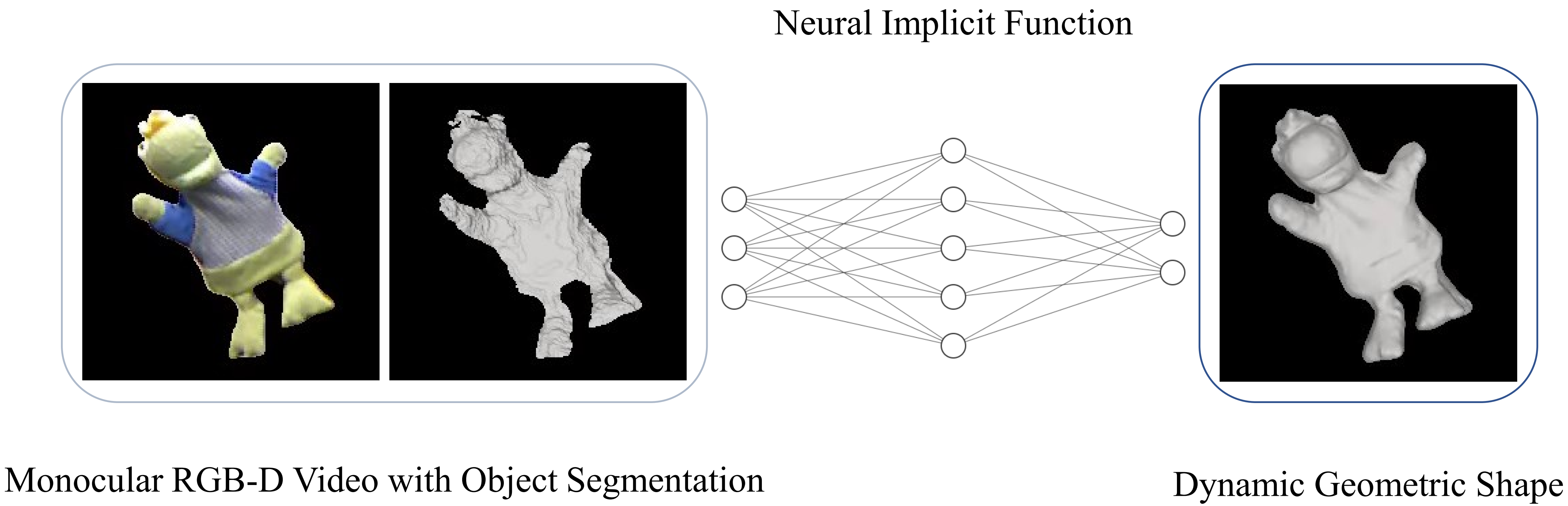
NEURAL INFORMATION PROCESSING SYSTEMS

Neural Surface Reconstruction of Dynamic Scenes with Monocular RGB-D Camera

Hongrui Cai, Wanquan Feng, Xuetao Feng, Yan Wang, Juyong Zhang*

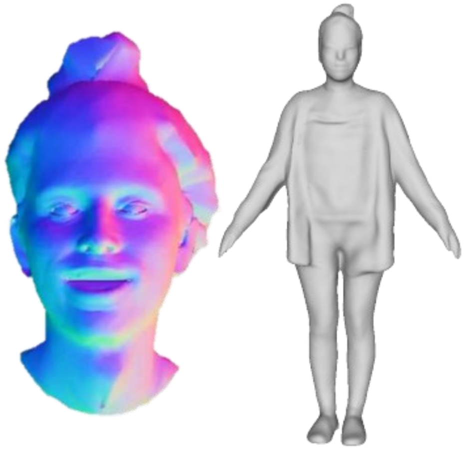
NeurIPS 2022

Overview


Neural Implicit Function
Monocular RGB-D Video with Object Segmentation
Dynamic Geometric Shape

Prior works

- Template-based methods



[3DMM. 1999.]
[SMPL. 2015.]
[FML. 2019.]
[A-NeRF. 2021.]
[NeuralHeadAvatar. 2022.]
[SelfRecon. 2022.]
… …

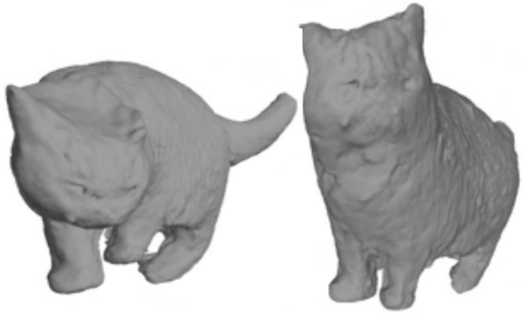
Ours: Various object categories

- RGB-D based methods

[DynamicFusion. 2015.]
[VolumeDeform. 2016.]
[MonoFVV. 2017.]
[KillingFusion. 2017.]
[DeepDeform. 2020.]
[OcclusionFusion. 2022.]
… …

Ours: Neural SDF and rendering

- RGB based methods

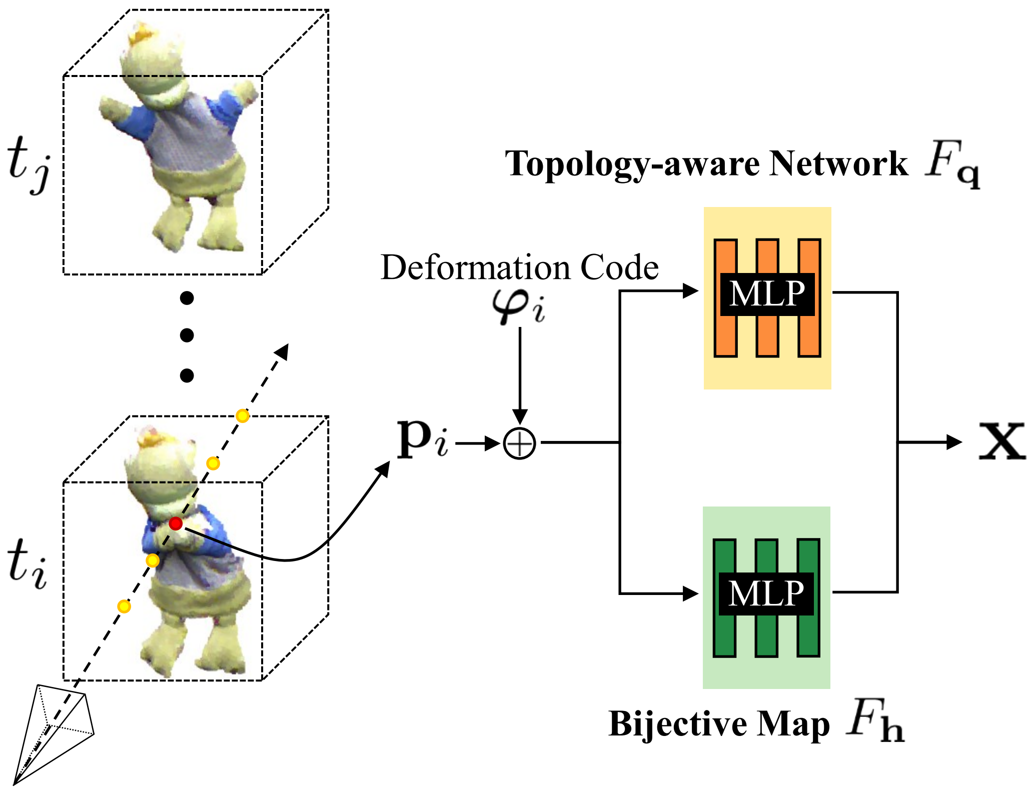


[LASR. 2021.]
[ViSER. 2021.]
[BANMo. 2022.]
… …

Ours: Cycle consistency
Topology-aware

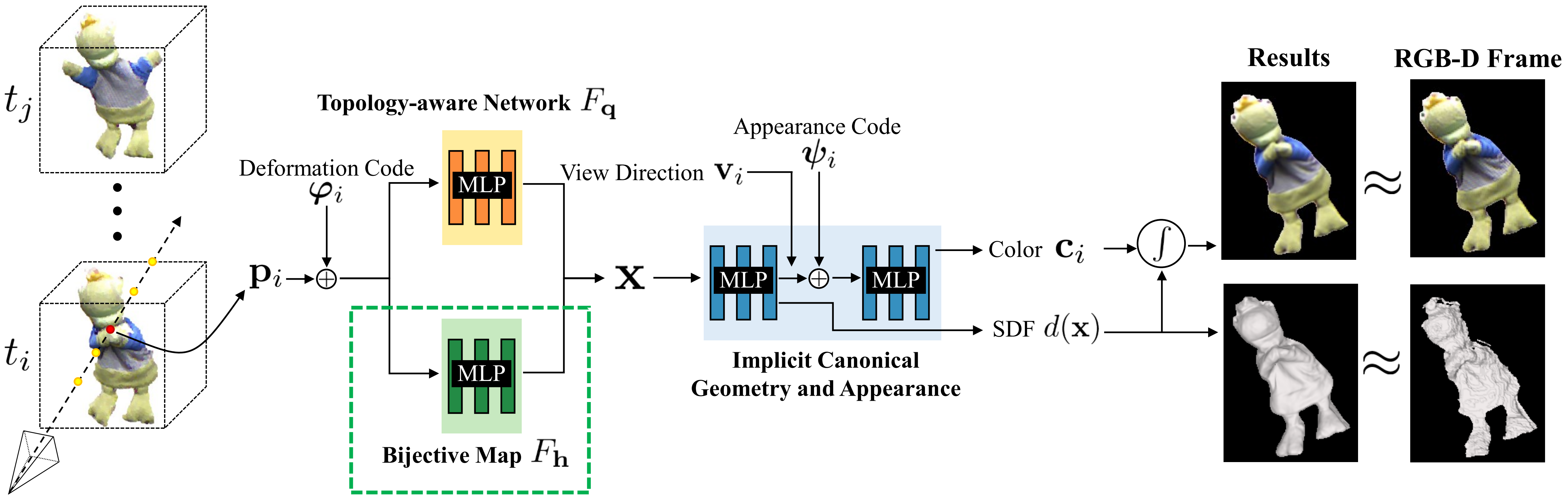
How to achieve these ideas?

Pipeline of NDR



Ben Mildenhall, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." ECCV, 2020.
Keunhong Park, et al. "Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields." TOG, 2021.

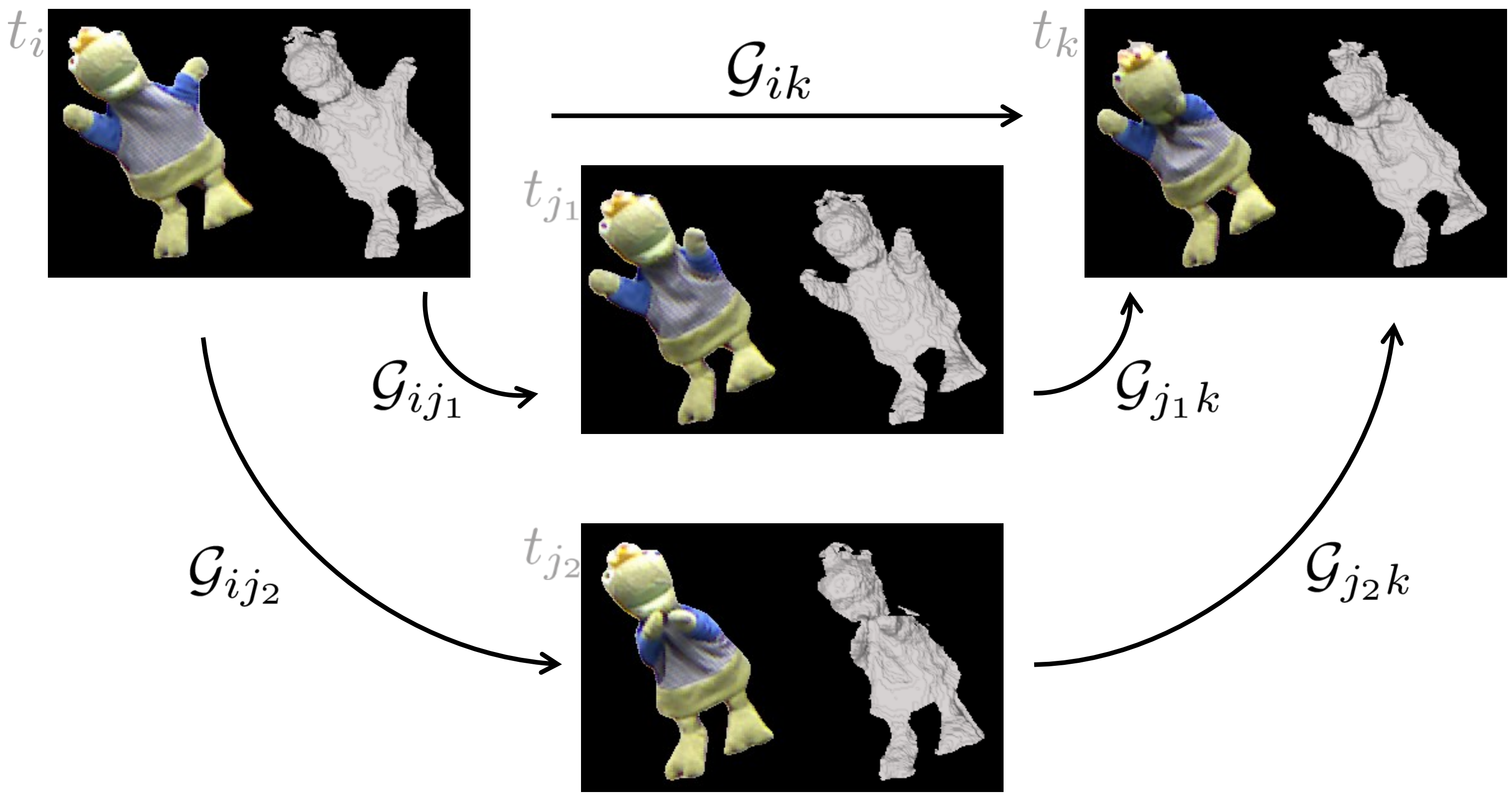
Pipeline of NDR



Maintain cycle consistency

Ben Mildenhall, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." ECCV, 2020.
Keunhong Park, et al. "Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields." TOG, 2021.
Peng Wang, et al. "Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction." NeurIPS, 2021.

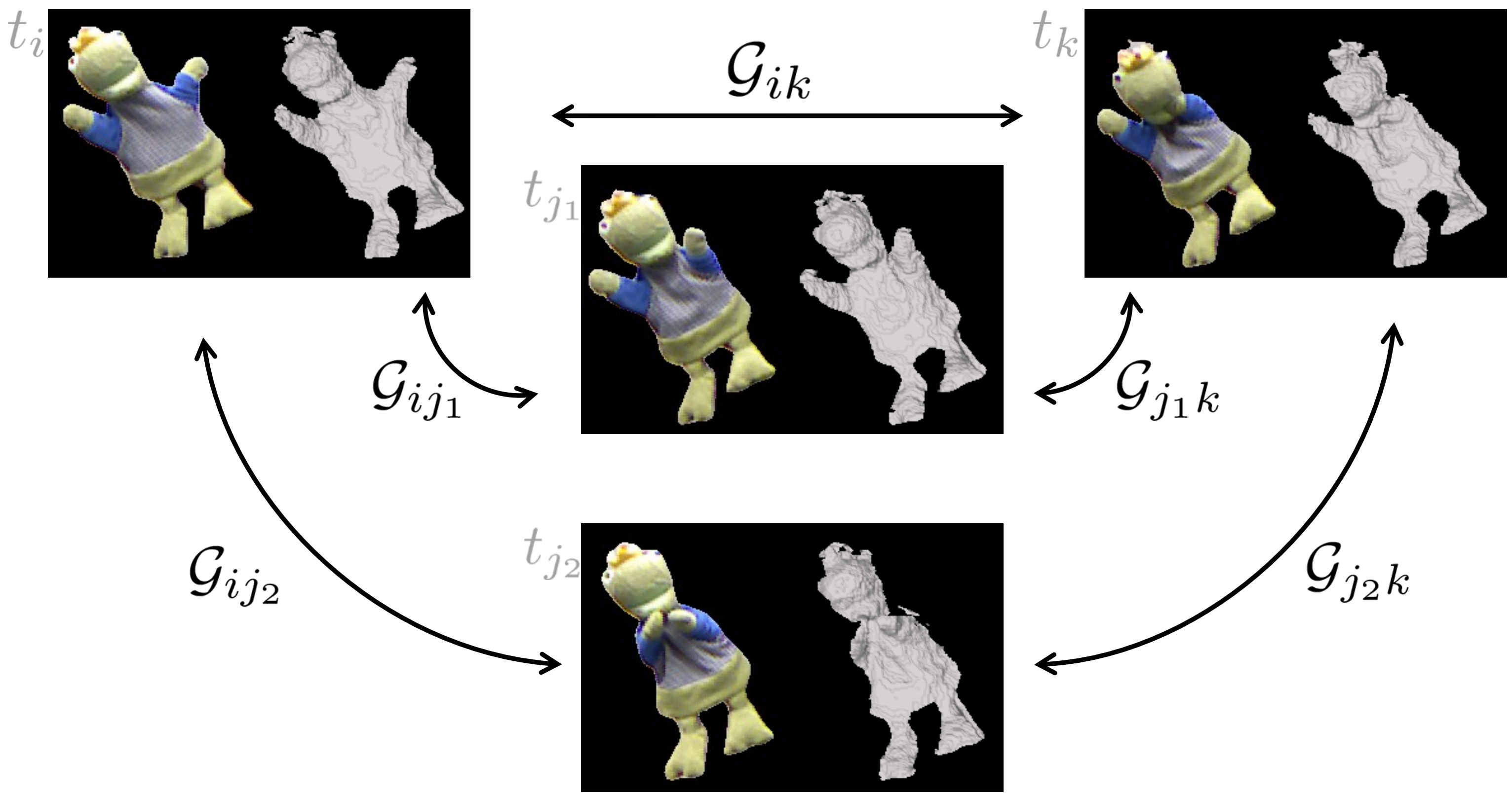
Observation: Deformation between different frames



Does cycle consistency property (invariant on deforming path) exist?

$$\mathcal{G}_{ik} \overset{?}{=} \mathcal{G}_{j_1k} \circ \mathcal{G}_{ij_1} \overset{?}{=} \mathcal{G}_{j_2k} \circ \mathcal{G}_{ij_2} \overset{?}{=} \mathcal{G}_{j_2k} \circ \mathcal{G}_{j_1j_2} \circ \mathcal{G}_{ij_1}$$

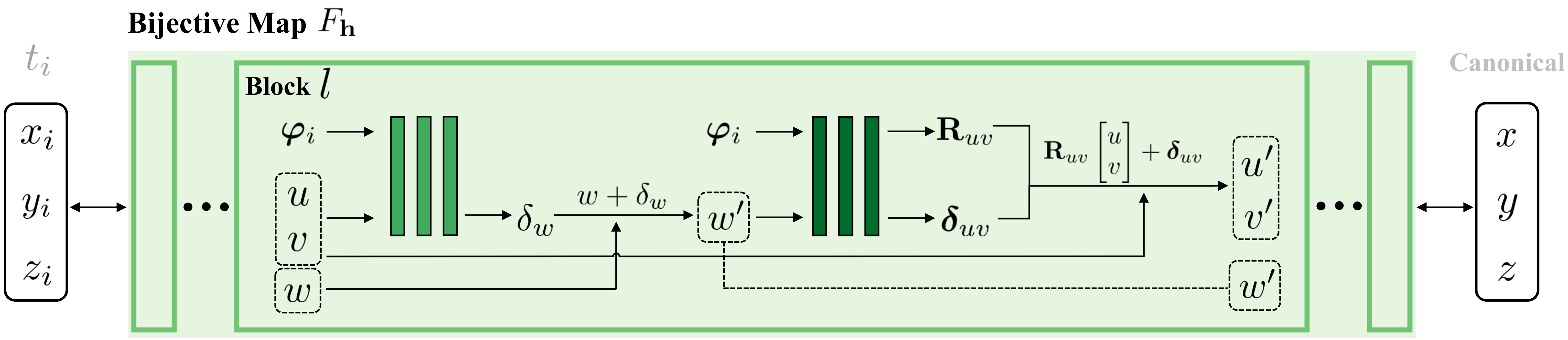
Observation: Deformation between different frames



One of the sufficient conditions: Deformation $\mathcal{G}$ is a bijective map.

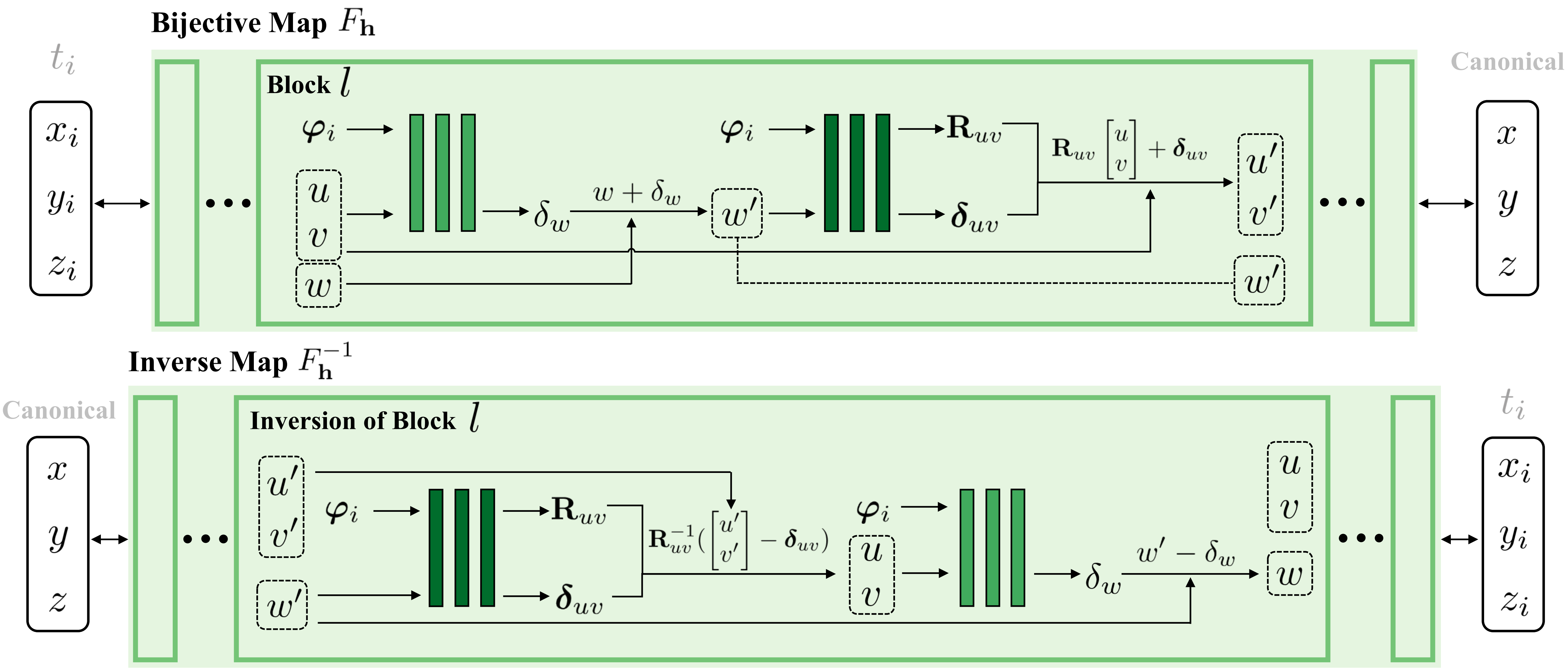
$$\mathcal{G}_{ik} = \mathcal{G}_{j_1k} \circ \mathcal{G}_{ij_1} = \mathcal{G}_{j_2k} \circ \mathcal{G}_{ij_2} = \mathcal{G}_{j_2k} \circ \mathcal{G}_{j_1j_2} \circ \mathcal{G}_{ij_1}$$

Key insight: Bijective map in the deformation field



Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. “Density estimation using real NVP.” ICLR, 2017.

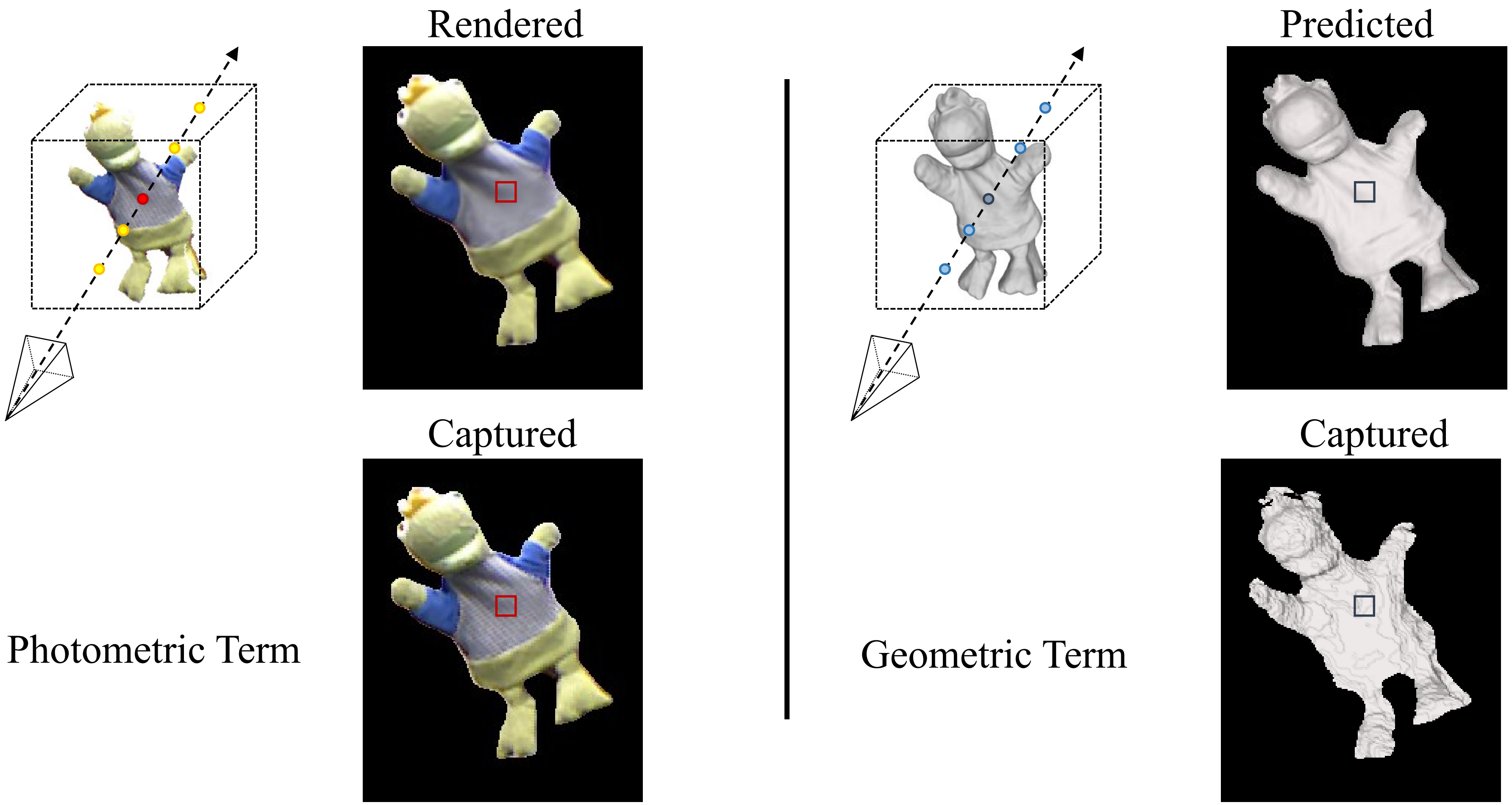
Key insight: Bijective map in the deformation field



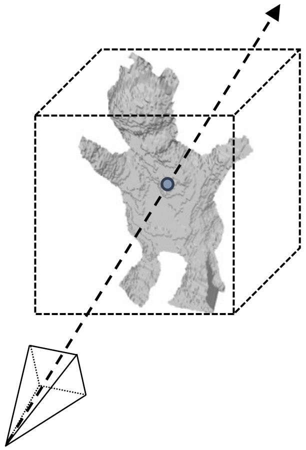
Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. "Density estimation using real NVP." ICLR, 2017.

How to optimize the representation?

Loss Terms: RGB-D supervision

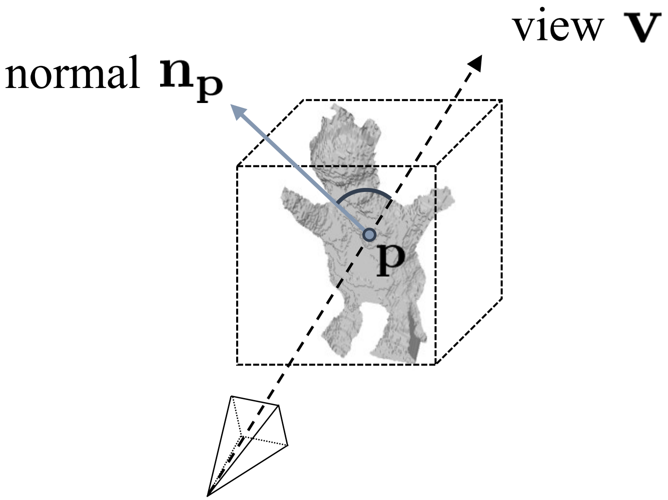


Loss Terms: Constraints on depth map



$$\mathrm{SDF} = 0$$

Zero-level Set Term



$$\angle(\mathbf{n_p}, \mathbf{v}) > 90^\circ$$

Visible Term

Some Representative Results

(please see webpage for more)

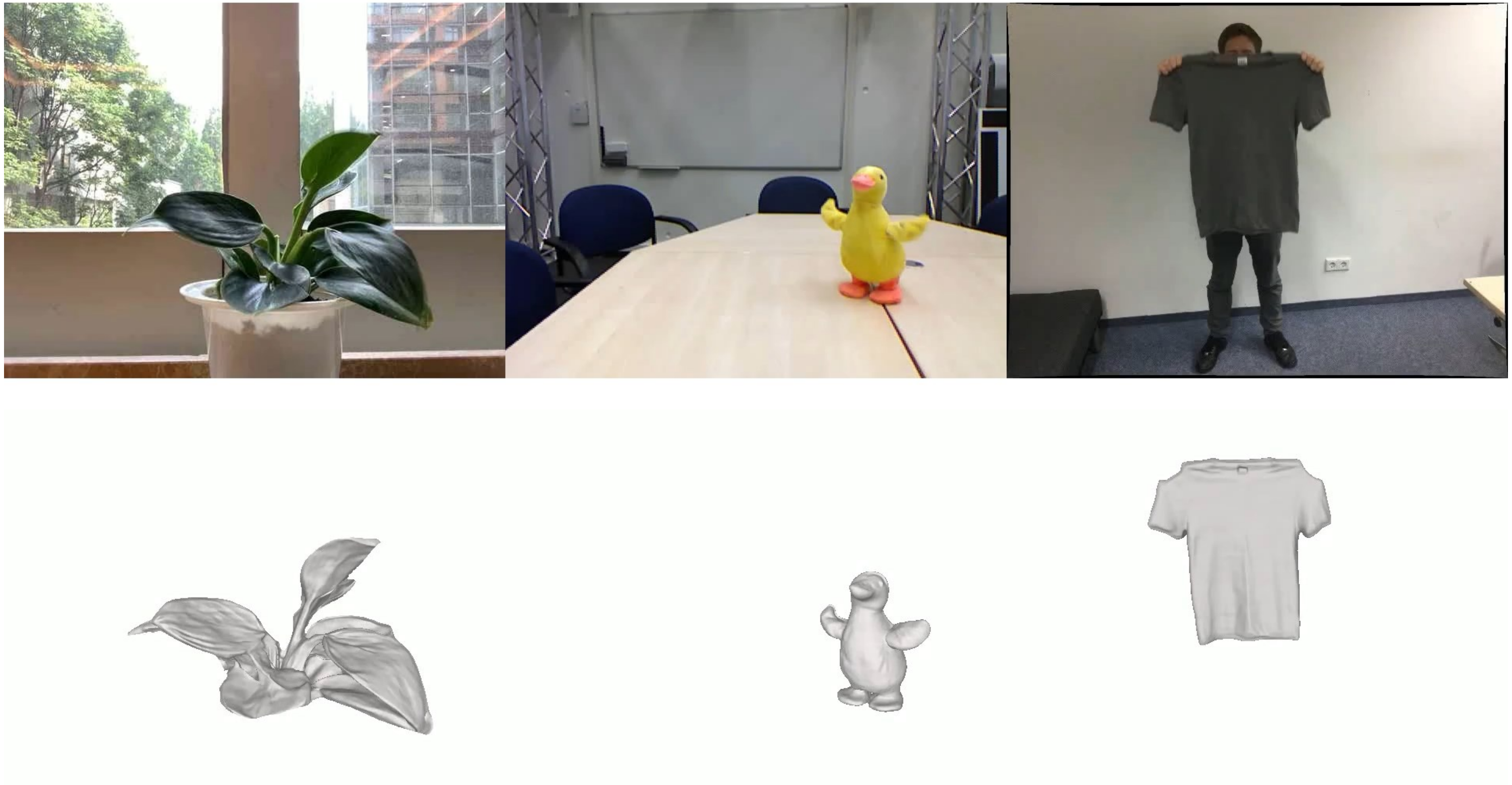
Results: Human body / head

Input RGB

Rendered

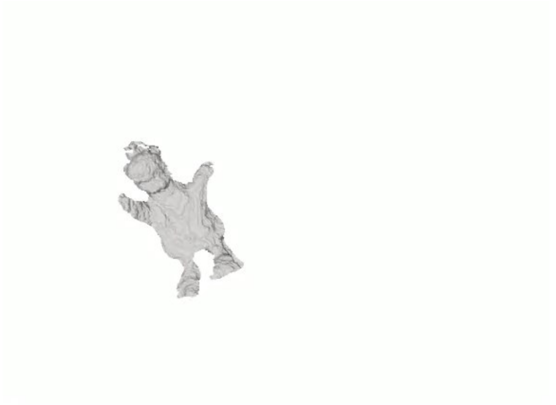
Reconstructed

Results: General objects



Comparisons: RGB-D reconstruction methods

Input RGB



Input depth

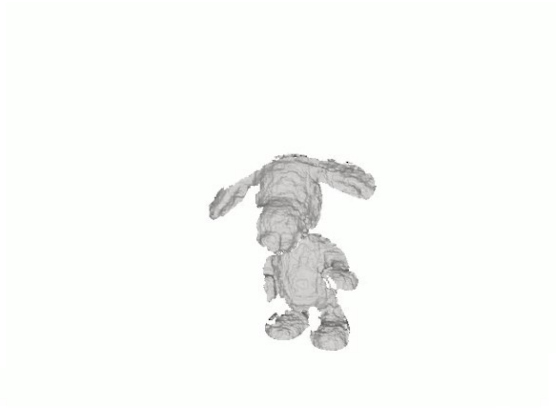
DynamicFusion
[Newcombe et al. 2015]

OcclusionFusion
[Lin et al. 2022]

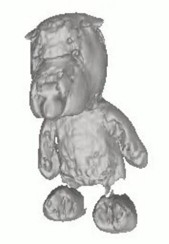
Ours

Comparisons: RGB reconstruction methods

Input RGB



Input depth



BANMo
[Yang et al. 2022]

Ours (only RGB)

Ours (full)

NDR: Summary

- Integrating observed RGB-D frames to a high-fidelity textured shape
- Maintaining cycle consistency between arbitrary two frames
- Being able to handle changes in topology-varying regions

Limitation: The optimization time is long

Neural Surface Reconstruction of Dynamic Scenes with Monocular RGB-D Camera

Thanks for watching

Project Page： https://ustc3dv.github.io/ndr/

Code： https://github.com/USTC3DV/NDR-code