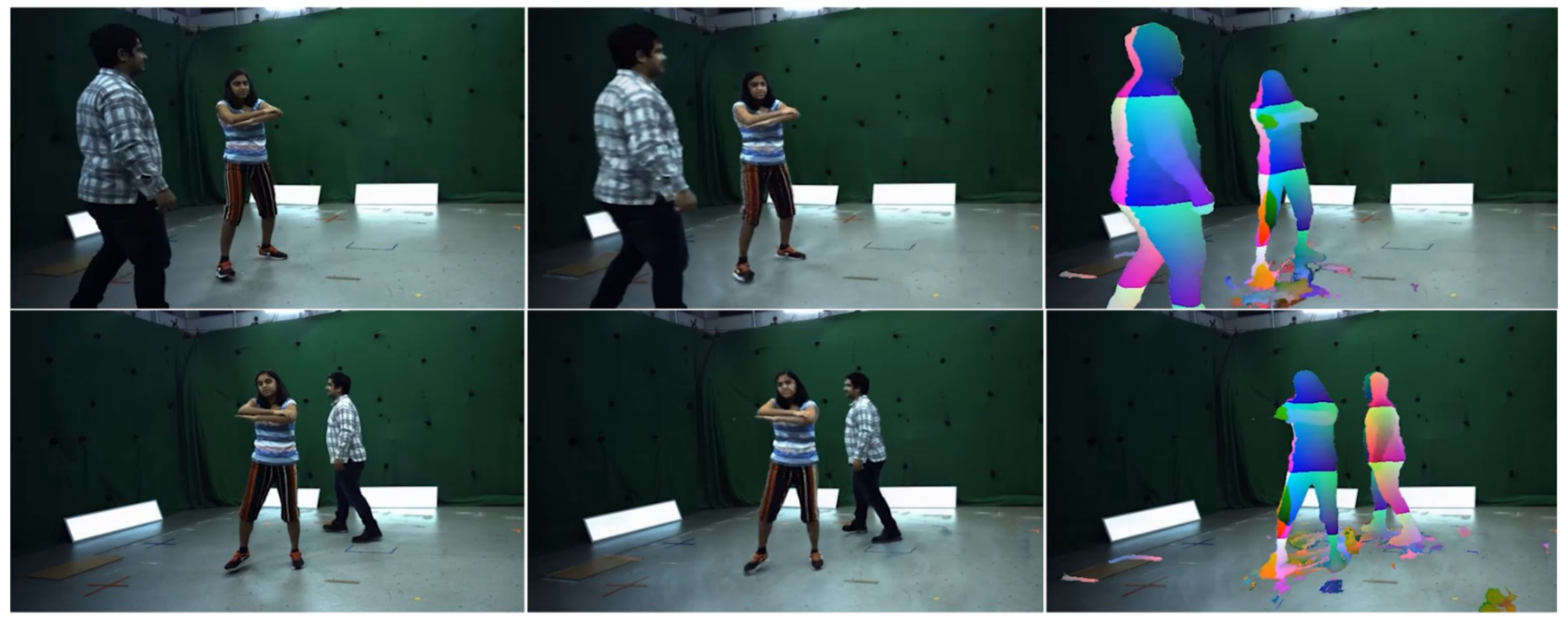
SceNeRFlow: Time-Consistent Reconstruction of General Dynamic Scenes

Edith Tretschk   Vladislav Golyanik   Michael Zollhöfer   Aljaž Božič   Christoph Lassner   Christian Theobalt

max planck institut informatik
SIC Saarland Informatics Campus
REALITY LABS

Goal

General dynamic NeRF with time consistency/correspondences even for large motion



Ground Truth   Reconstruction   Correspondences

Problem Setting and Context

Input
General non-rigid scene captured with multi-view RGB videos (with known camera parameters and background images)

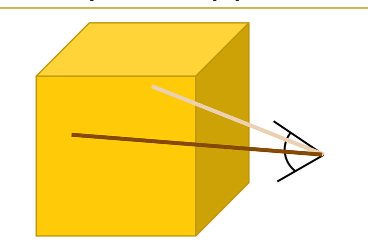
→

Output
Time-consistent reconstruction of geometry, appearance, and deformations

Prior Work: Either category-specific (*e.g.* humans) or only handles small motion (*e.g.* only consistent over short time windows)
→ Ours is first method to get correspondences for large general motion!

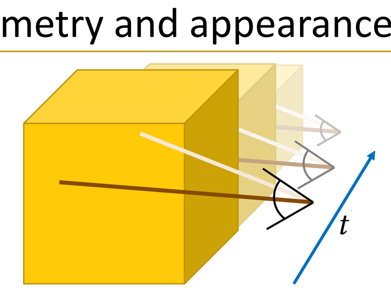
High-Level Method Idea:

Static NeRF
No deformation, only geometry and appearance



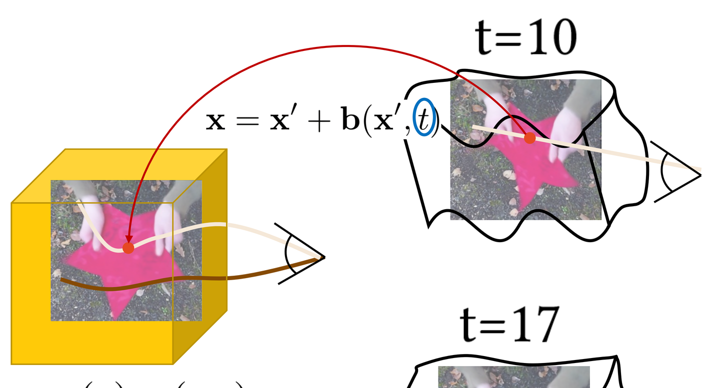
$\mathbf{v}(\mathbf{x}) = (o, c)$

Volumetric Video
Entangle deformation with geometry and appearance



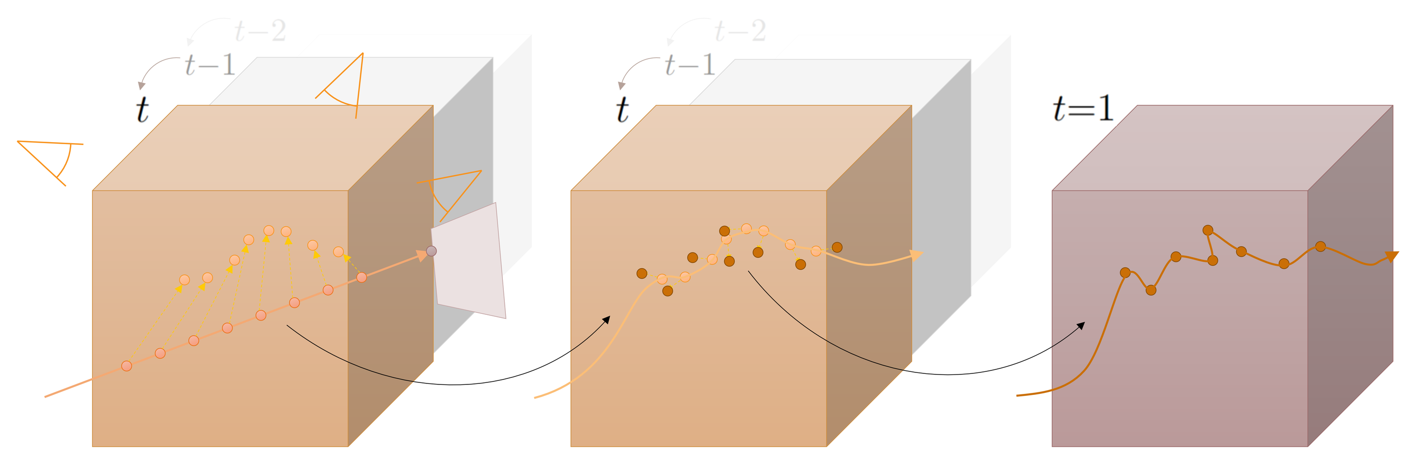
$\mathbf{v}(\mathbf{x}, t) = (o, c)$

Deformable NeRF
Disentangle deformation from geometry and appearance



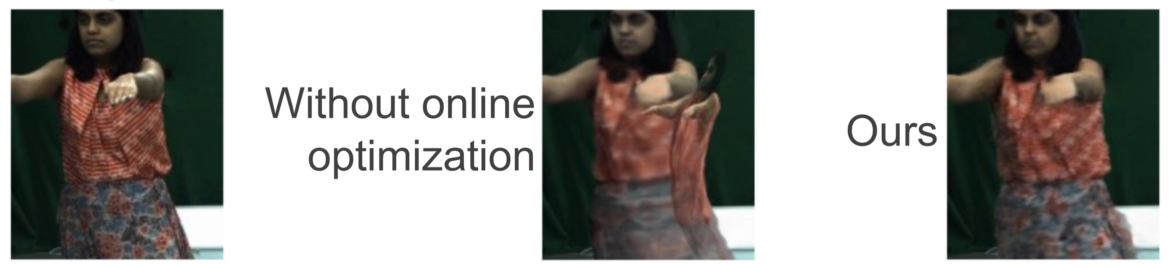
$\mathbf{v}(\mathbf{x}) = (o, c)$

Method



Coarse Deformations   Fine Deformations   Canonical Model

- Build static canonical model (*i.e.* geometry & appearance) at *t=1*
- Online optimization of deformations at *t>1*, regularized by as-rigid-as-possible deformation smoothness loss



Input   Without online optimization   Ours

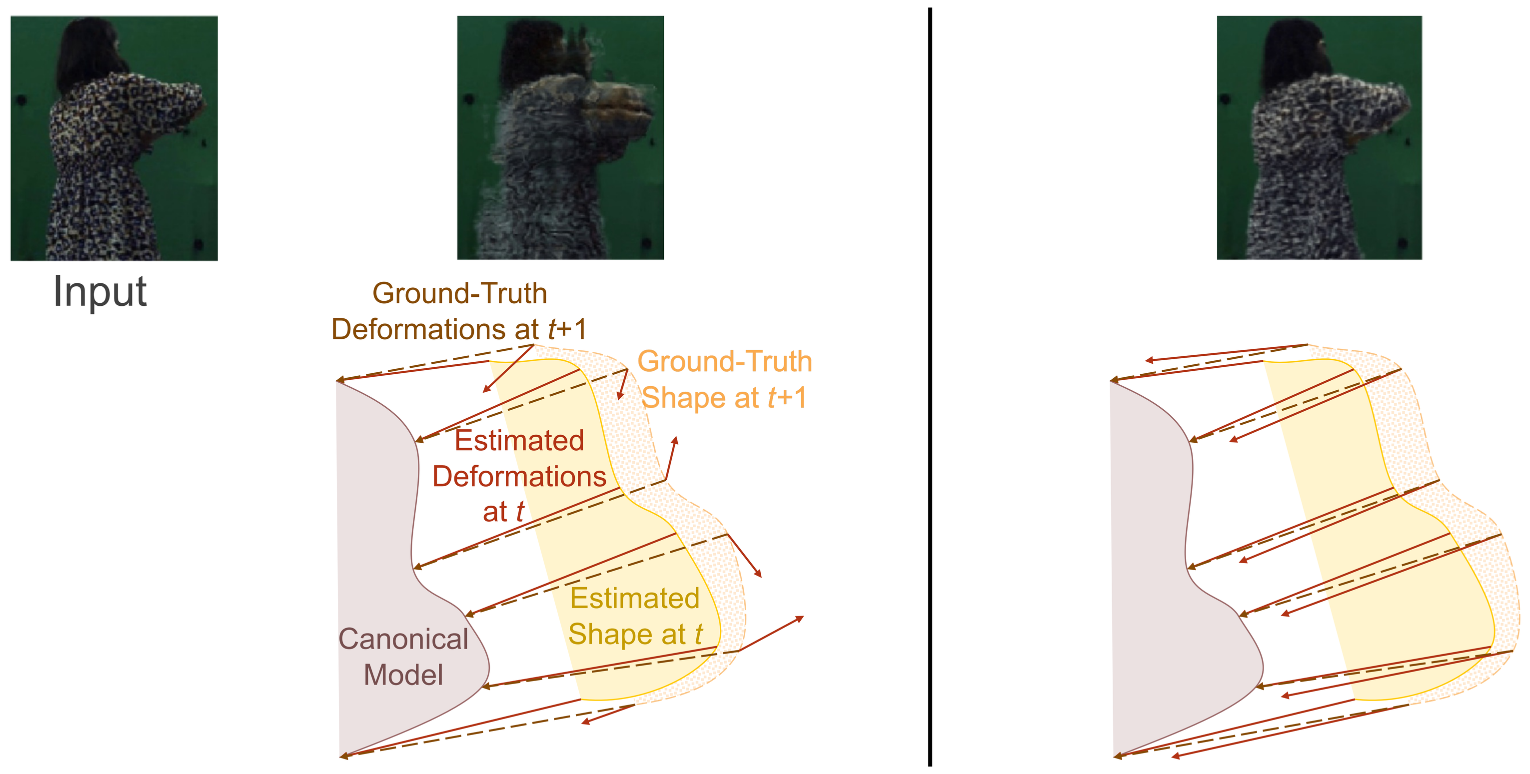
- Decompose into coarse and fine deformations

Input   Without coarse deformations   Without fine deformations   Ours

Unexpected Challenge: Doing all this yields *very* strong artifacts!

Why? Backwards deformation models have bad initialization for large motion!

Solution: Initialize *surrounding* space via deformation smoothness loss



Input

Bonus: *Fast* As-Rigid-As-Possible Deformation Smoothness

Issue
Nerfies [1] is slow because its elastic loss requires for each point on the ray (1) three backward passes and (2) a 3x3 SVD

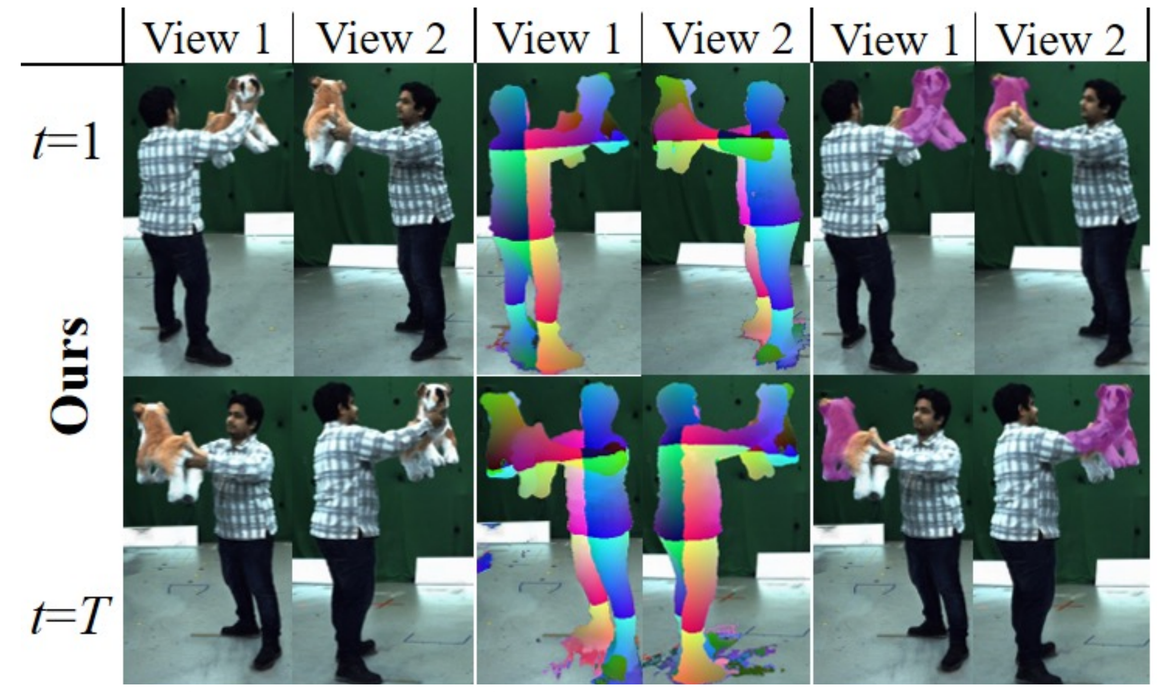
Insight 1
Automatic differentiation is great at Jacobian-vector products

Insight 2
Can relax ARAP's rotation constraint from *SO(3)* to *O(3)*, *i.e.* allow for reflections

Combining Both Insights
O(3) is equivalent to norm preservation
→ Norm preservation loss via Jacobian-vector product *in single backward pass without SVD*:

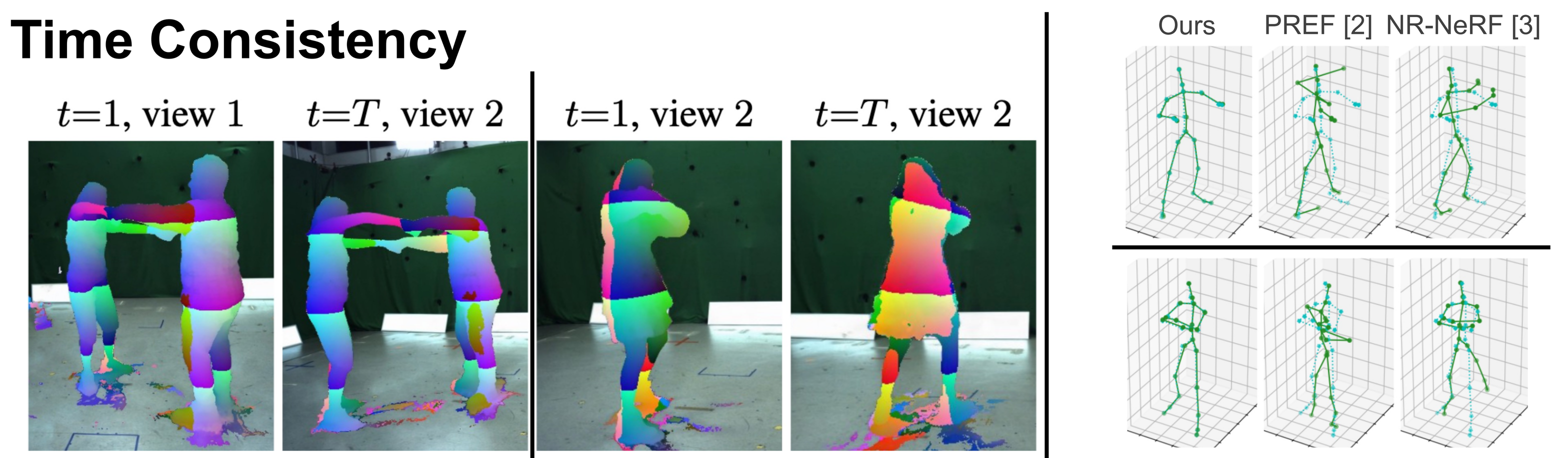
$$\mathcal{L}_{\text{norm}} = \frac{1}{RS} \sum_r \sum_i \mathbb{E}_{\mathbf{e}} \left[\left| \| \mathbf{J}^\top_{\mathbf{r}_r(s_i)} \mathbf{e} \|_2 - 1 \right| \right]$$

Results



Ground Truth   Correspondences   Editing

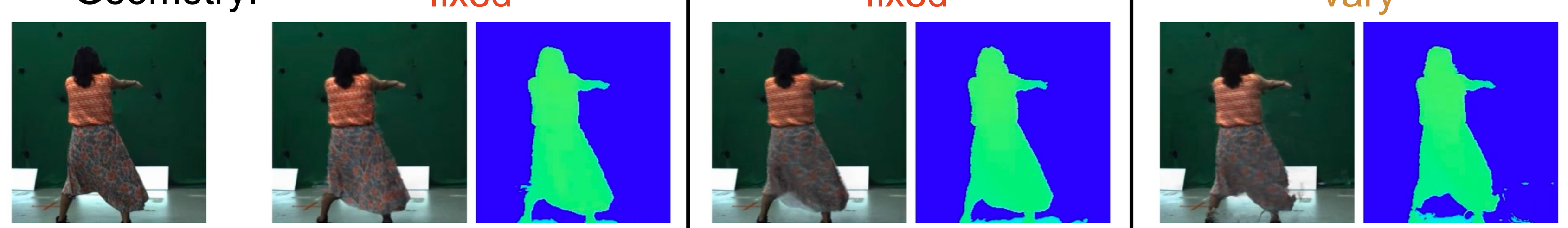
Time Consistency



Use reconstructions (blue) to track joints until final frame (ground truth in green)

Ablation: Letting the Canonical Model Vary Over Time

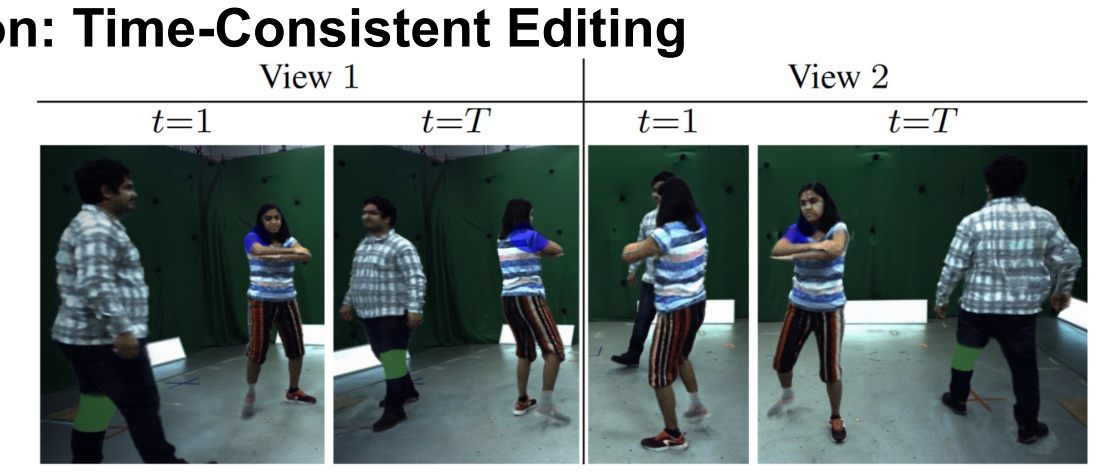
Appearance:	fixed	vary	vary
Geometry:	fixed	fixed	vary



Ground Truth   Ours

→Varying the canonical model gives better reconstruction *but loosens correspondences!*
→Trade-off between novel-view synthesis quality and temporal consistency

Application: Time-Consistent Editing



References:
[1] Park *et al.*: Nerfies: Deformable Neural Radiance Fields. ICCV 2021.
[2] Song *et al.*: PREF: Predictability Regularized Neural Motion Fields. ECCV 2022.
[3] Tretschk *et al.*: Non-Rigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video. ICCV 2021.

Acknowledgments: All data capture and evaluation was done at MPII. Research conducted by Vladislav Golyanik and Christian Theobalt at MPII was supported in part by the ERC Consolidator Grant 4DReply (770784). This work was also supported by a Meta Reality Labs research grant.

Code is available!
github.com/facebookresearch/SceNeRFlow

Video results:
vcai.mpi-inf.mpg.de/projects/scenerflow