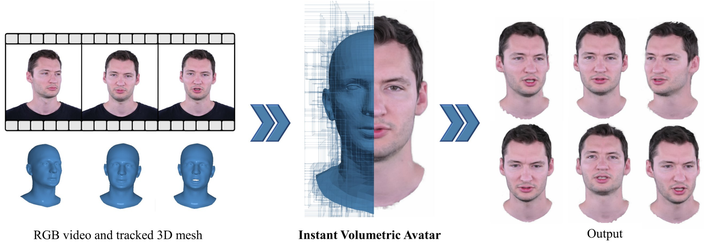
Instant Volumetric Head Avatars (INSTA)

Wojciech Zielonka Timo Bolkart Justus Thies
Max Planck Institute for Intelligent Systems, Tübingen, Germany

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

10 min

Problem



Goal: Given an RGB video and tracked mesh create a controllable metrical avatar.
Problems:

- Current methods need days to create an avatar, thus, they are able only to present a prerecorded image which does not necessarily reflect our current look.
- Methods that are able to quickly optimize and render an avatar are needed for efficient telepresence.

Evaluation

Method	L2 ↓	PSNR ↑	SSIM ↑	LPIPS ↓	Time ↓
NHA	0.0022	27.71	**0.95**	**0.04**	0.63
IMAvatar	0.0023	27.62	0.94	0.06	12.34
NeRFace	**0.0018**	**29.28**	**0.95**	0.07	9.68
Ours	**0.0018**	28.97	**0.95**	0.05	**0.05**

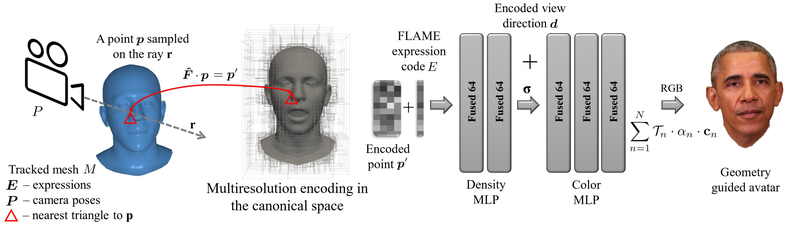
time needed to render a single frame

INSTA needs only 1 minute to optimize an avatar and 10 minutes to converge. Despite the order of magnitude shorter training time, it achieves on-par SOTA results compared to the methods which need from one to several days. Additionally, by leveraging Instant-NGP rendering heuristics we are able to have an interactive rendering frame rate.

Method

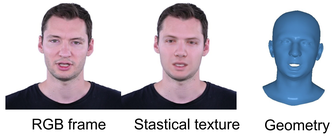
INSTA (Instant face)

By utilizing Instant-NGP [Müller at al. 2022] multi-hashing grid encoder and canonical space anchor equipped with fast mapping INSTA can optimize avatar in several minutes. Given a tracked mesh and the corresponding canonical one with the same topology, we can utilize BVH for a fast nearest triangle search for each point **p** sampped on a ray in the deformed space and use the triangle local coordinate systems (**TBN**) for efficient projection to the canonical space where NeRF [Mildenhall et al. 2020] is embeded.

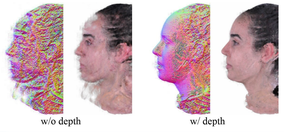


Tracker

To get the precise geometry needed for accurate projection we used a tracker from MICA [Zielonka et al. 2022].



Having tracked geometry we could additionally impose prior to improving NeRF densities.



Applications

Given a source sequence and target actors we can easily perform expressions retargettning.

Results

INSTA produces high-quality photorealistic images which are on par with other SOTA methods. Moreover, after only 60s we get a reasonable avatar that is ready to use.