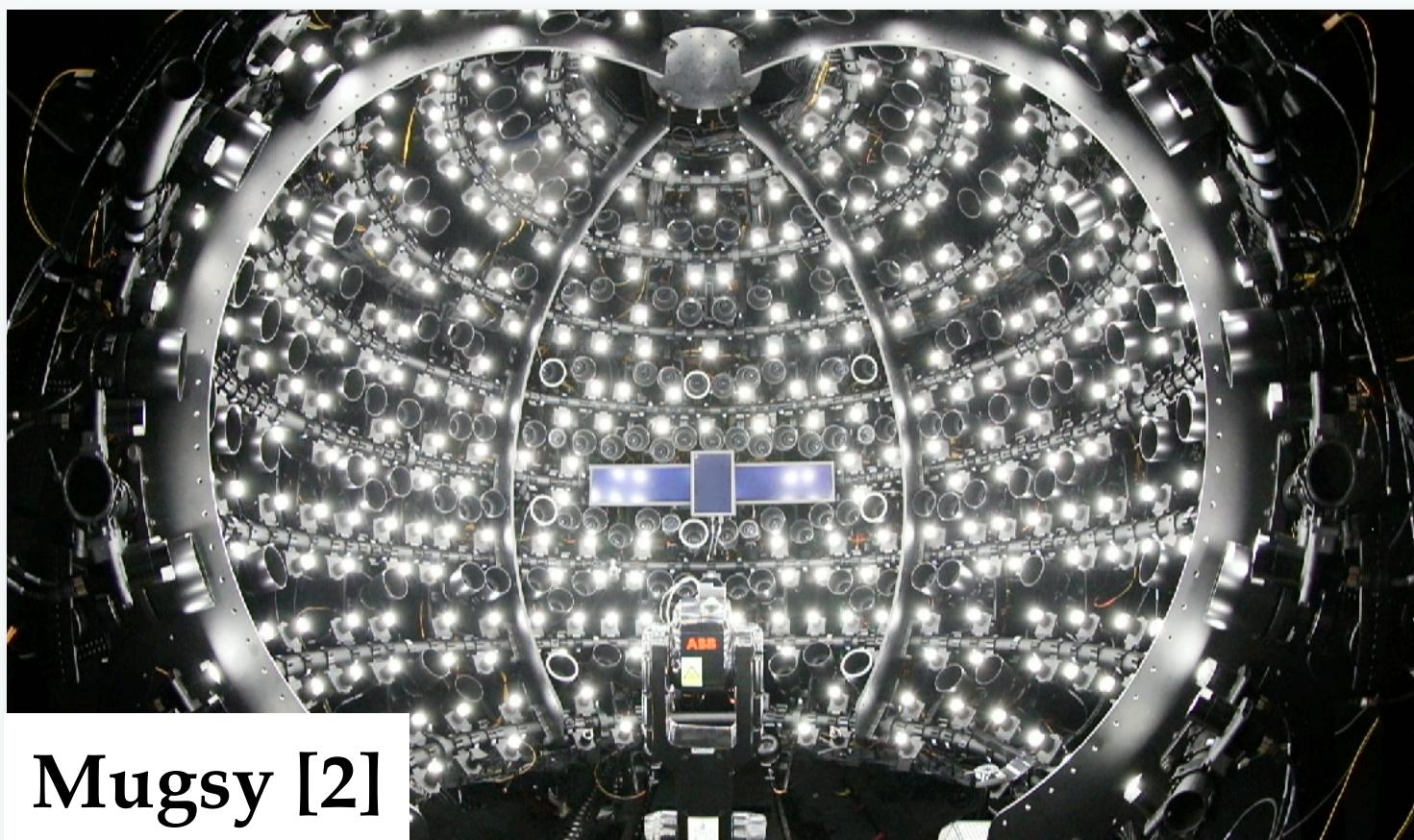


Multiface

A Dataset for Neural Face Rendering

Cheng-hsin Wu*, Ningyuan Zheng*, Scott Ardisson, Rohan Bali, Danielle Belko, Eric Brockmeyer, Lucas Evans, Timothy Godisart, Hyowon Ha, Alexander Hypes, Xuhua Huang, Taylor Koska, Steven Krenn, Stephen Lombardi, Xiaomin Luo, Kevyn McPhail, Laura Millerschoen, Michal Perdoch, Mark Pitts, Alexander Richard, Jason Saragih, Junko Saragih, Takaaki Shiratori, Tomas Simon, Matt Stewart, Autumn Trimble, Xinshuo Weng, David Whitewolf, Chenglei Wu, Shou-I Yu, Yaser Sheikh



Mugsy [2]

150
 # Cameras (v2)

2048 ×
1334
 Resolution

30
 Frame Rate

13
 Identities

216
 Expressions,
 audio & gaze
 segments (v2)



[1]

This dataset consists of recordings of the faces of 13 identities, each captured in a multi-view [1] capture stage with **Mugsy** [2], while performing various facial expressions by following through scripts that cover peak expressions, range of motions, gaze fixations, and phonetically balanced sentences. A **VAE model** [4] that allows for nonlinear interpolation over continuous view angles and expressions is provided.

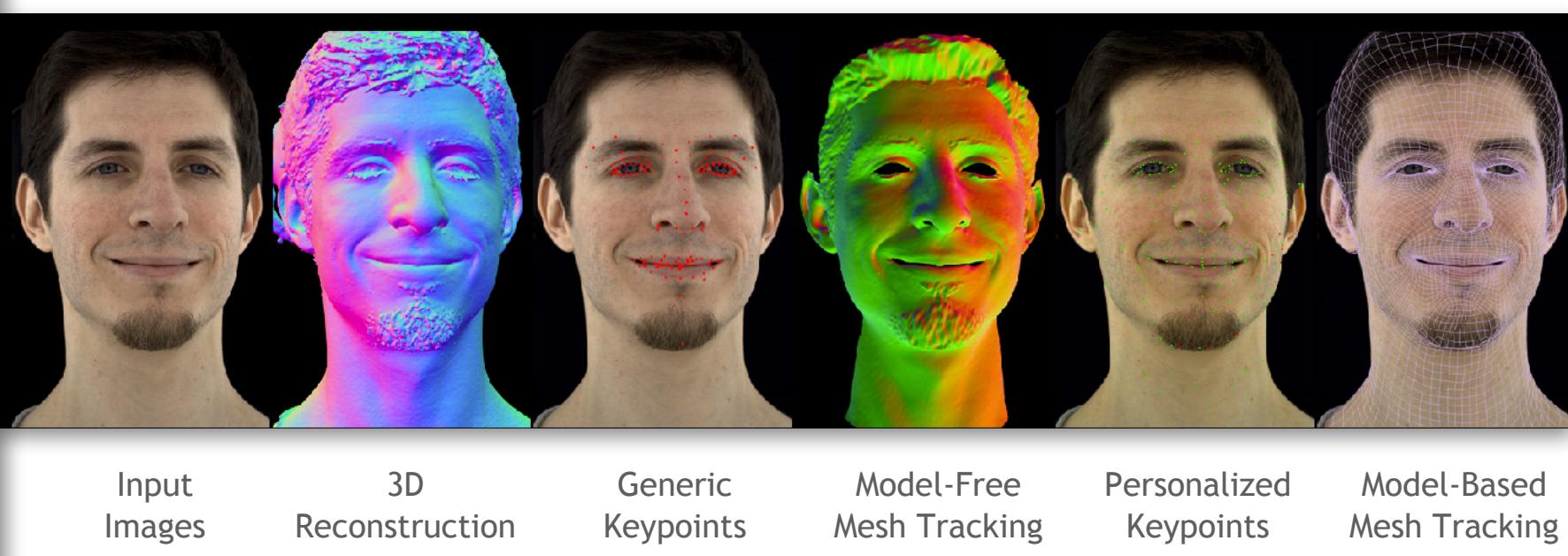
Assets

For each identity, we provide captured **images** from each camera view at a resolution of 2048×1334 pixels, metadata including intrinsic and extrinsic **camera calibrations**, **audio**, and processed artifacts such as **tracked meshes**, **headposes**, **unwrapped textures** at 1024×1024 pixels which come from **Genesis** [3] process.

Genesis [3]

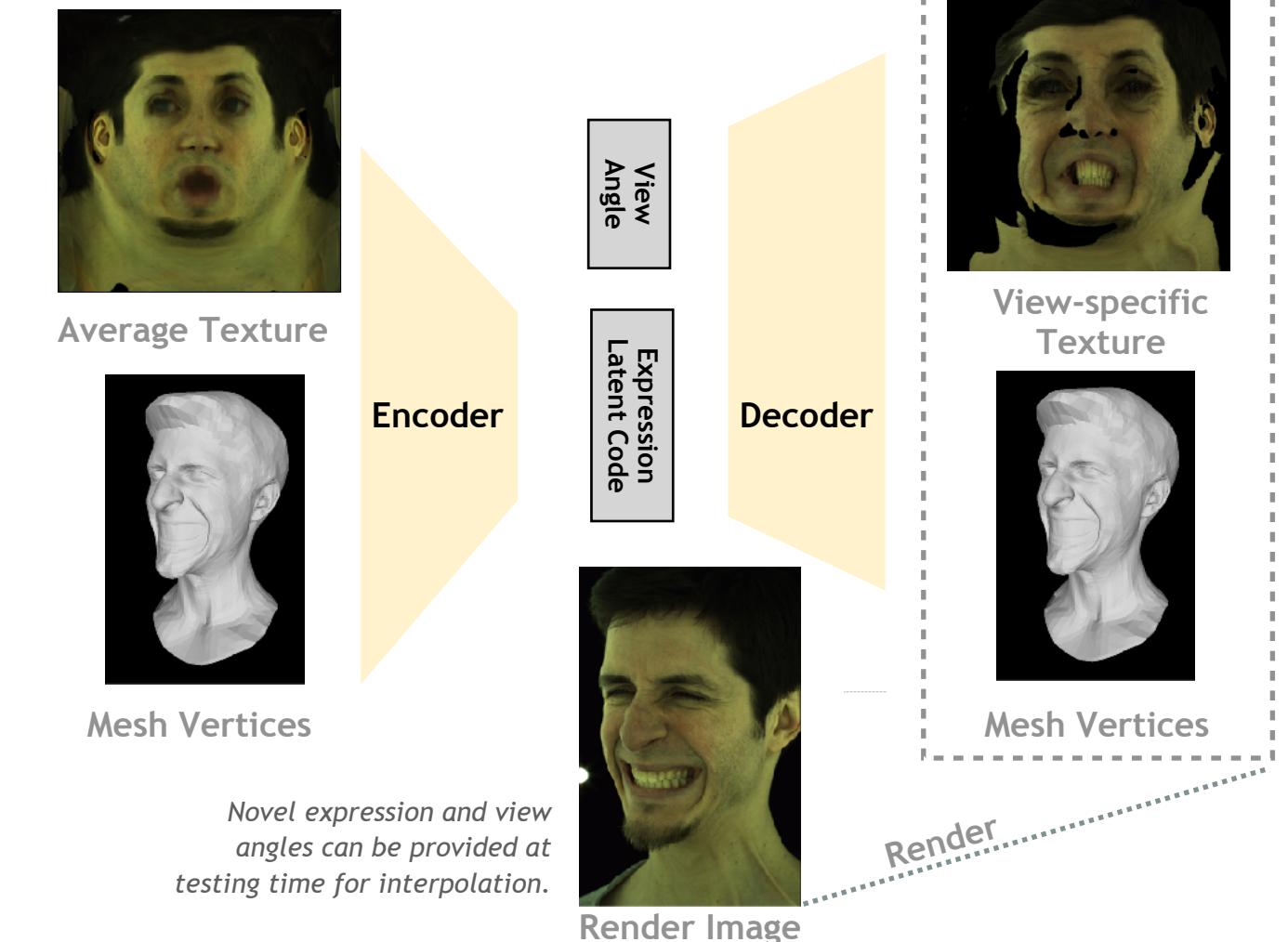
Input: Calibrated sets of 2D images.

Output: Unwrapped texture and 3D mesh for each frame.

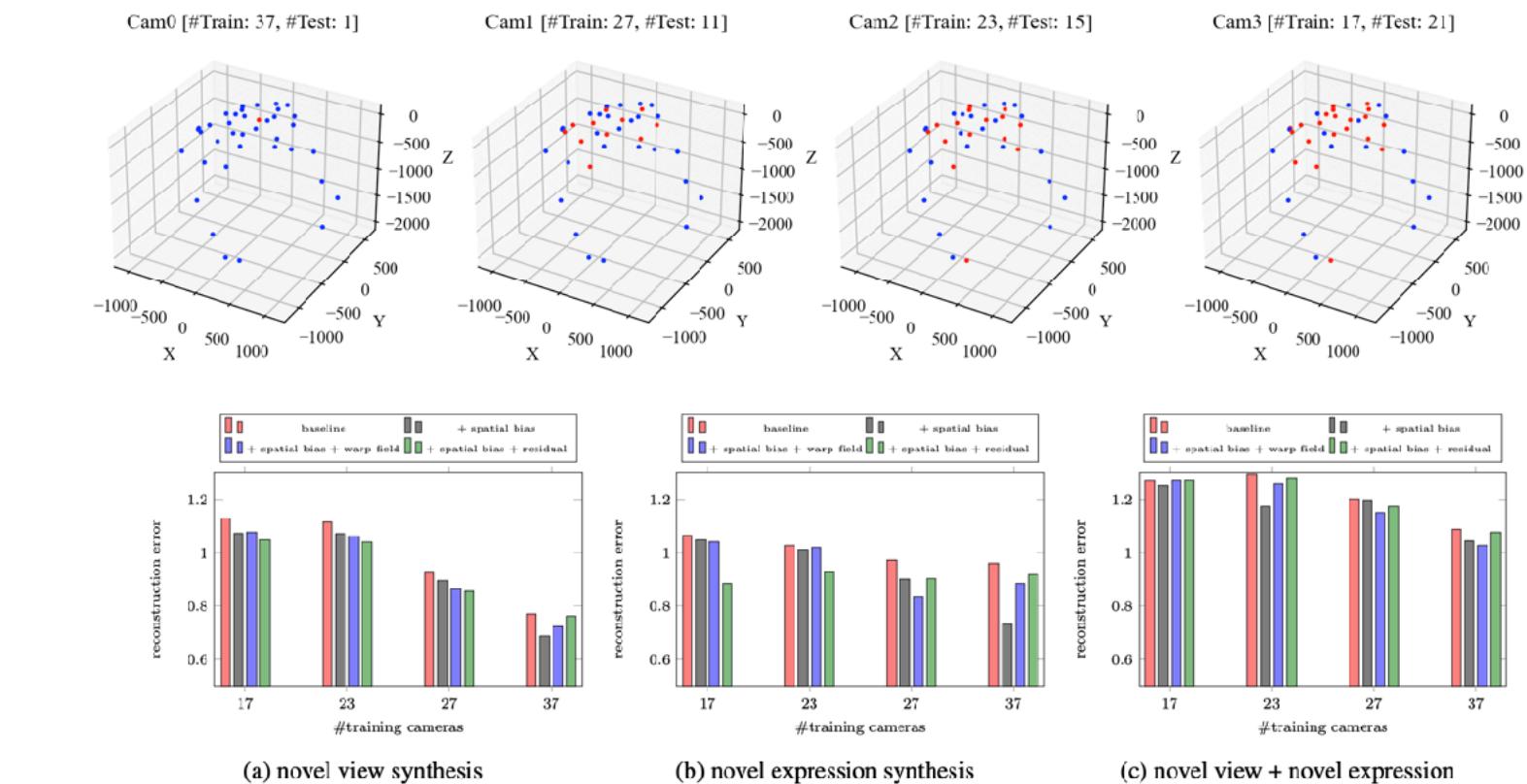


Codec Avatars

VAE Baseline [4]



Architecture Ablation



Related Work/Research

Multiface



Pixel Codec Avatar
 (Mobile SOTA)



MVP
 (PC SOTA)

