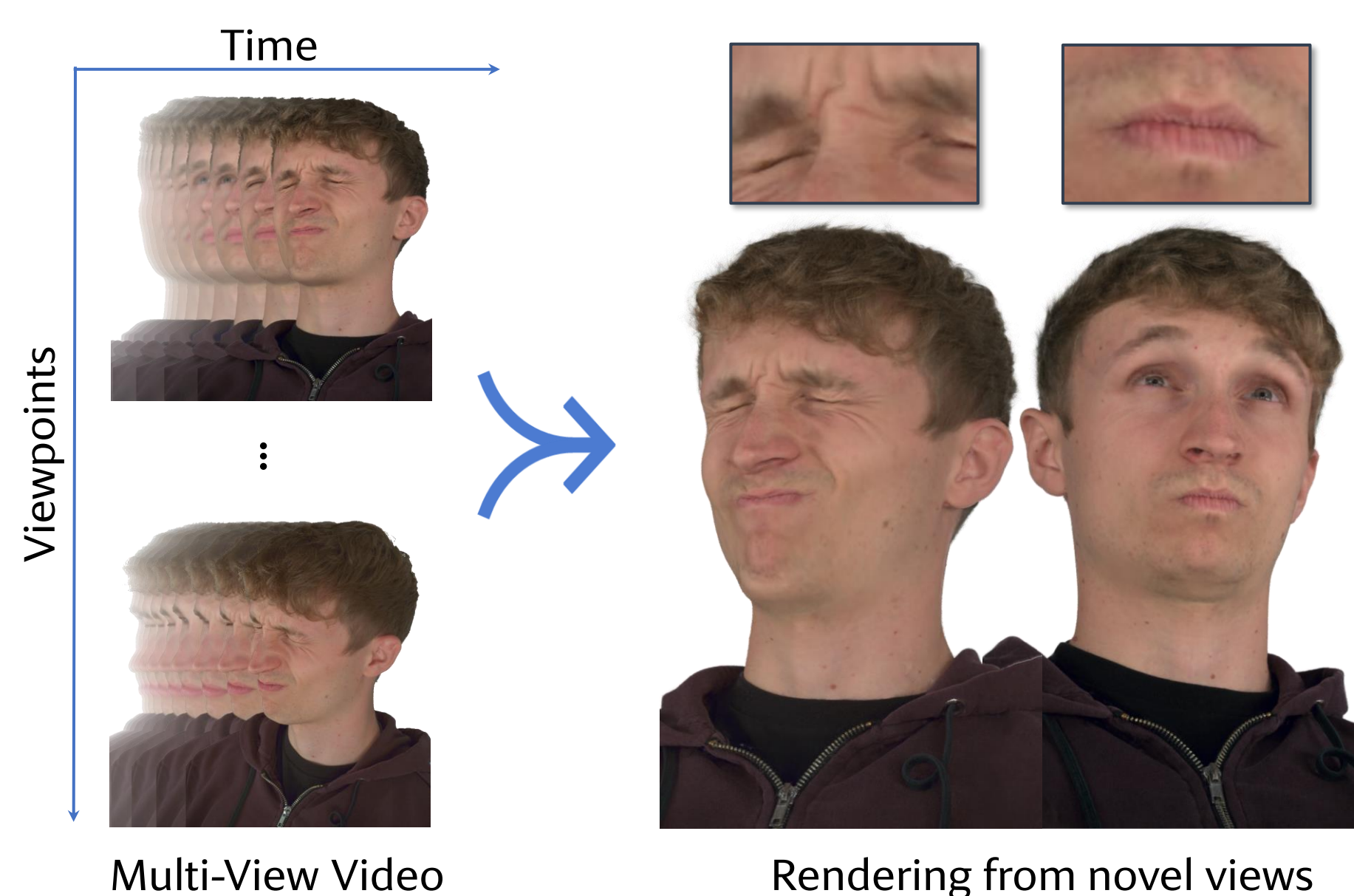


## Task

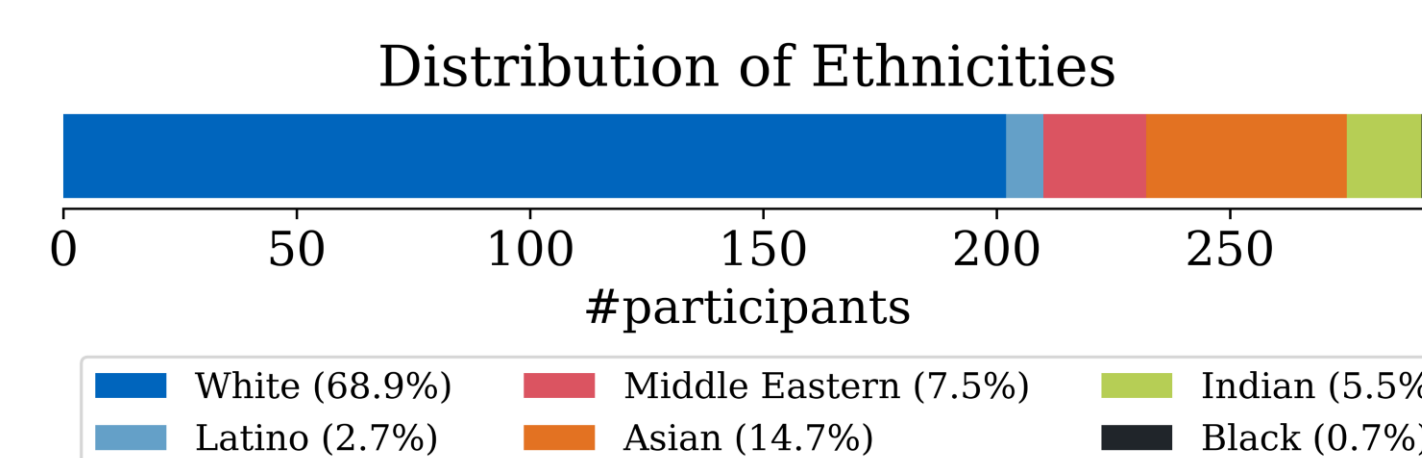
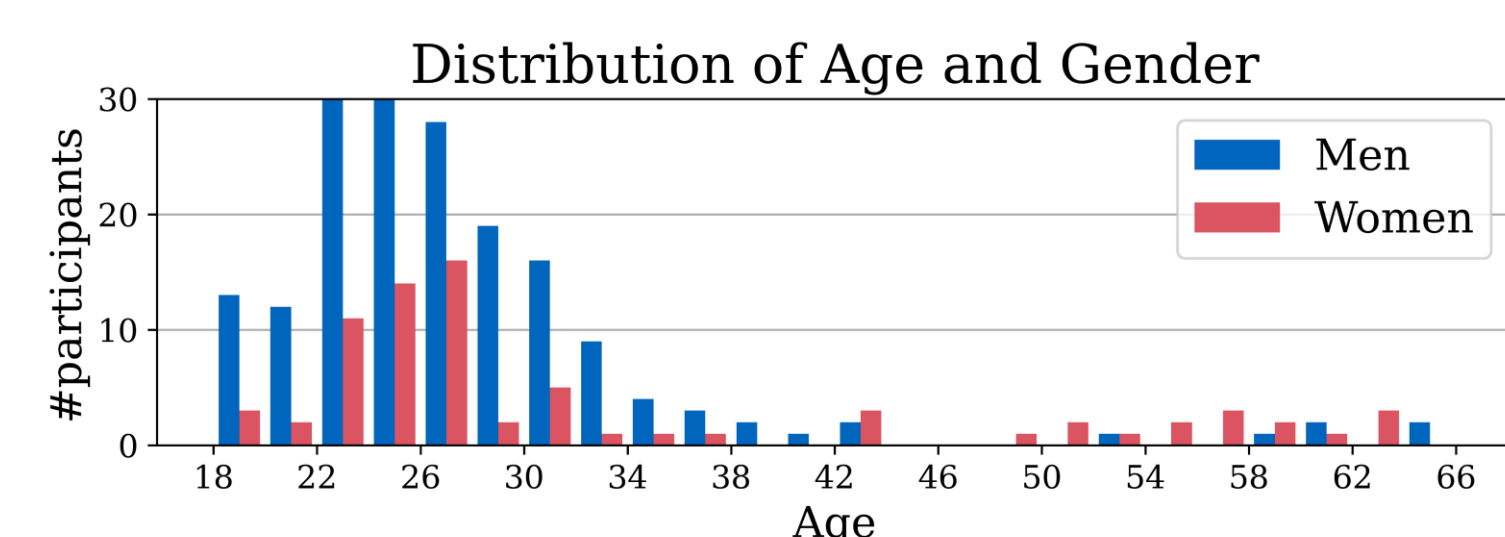


**Input:** 12 time-synchronized videos  
**Output:** Rendering from arbitrary viewpoints and timesteps

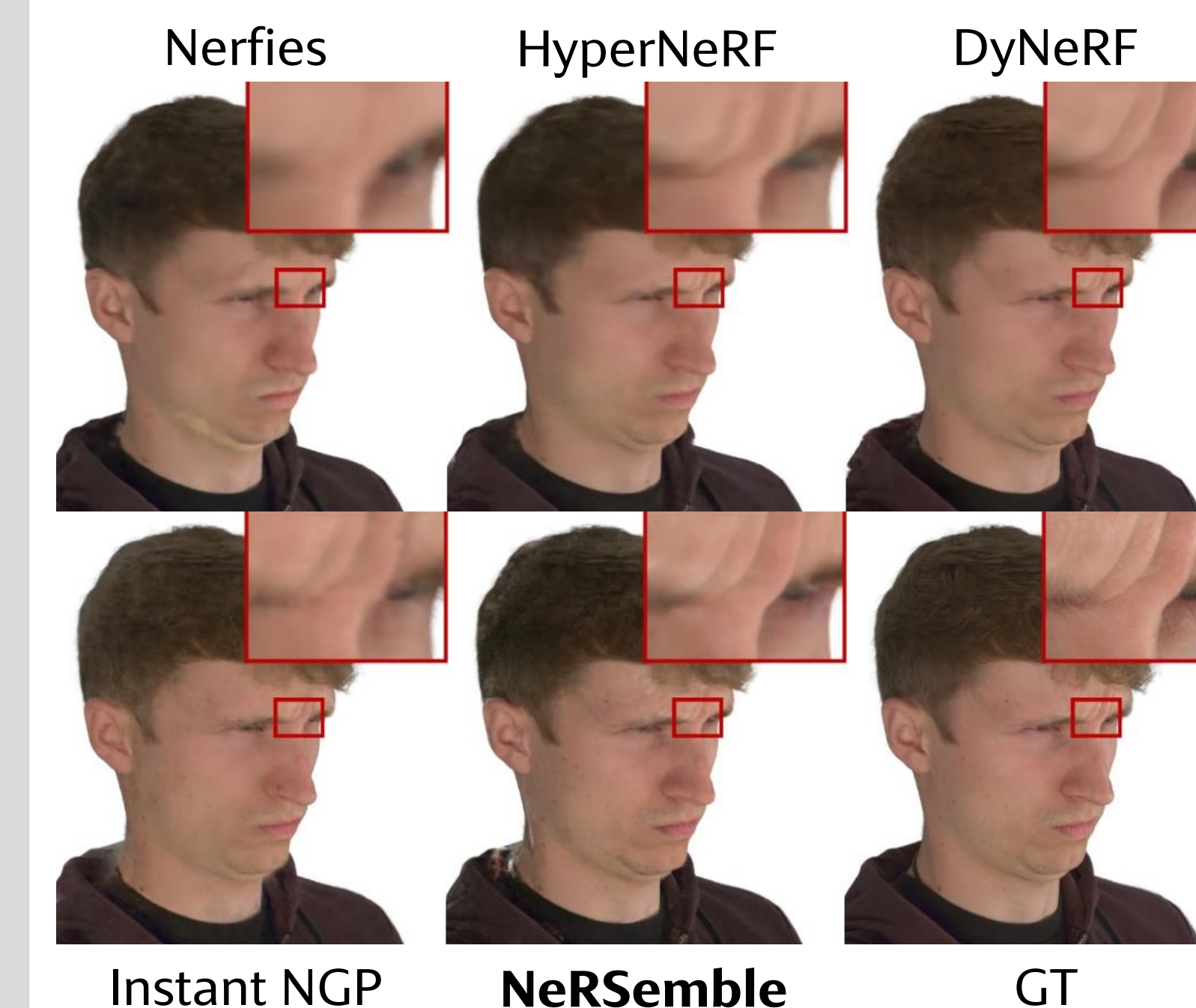
## Multi-view Capture Setup



Microphone Instructions 16 Cameras: 7.1 MP @ 73 fps, Global Shutter, < 1μs Time Sync

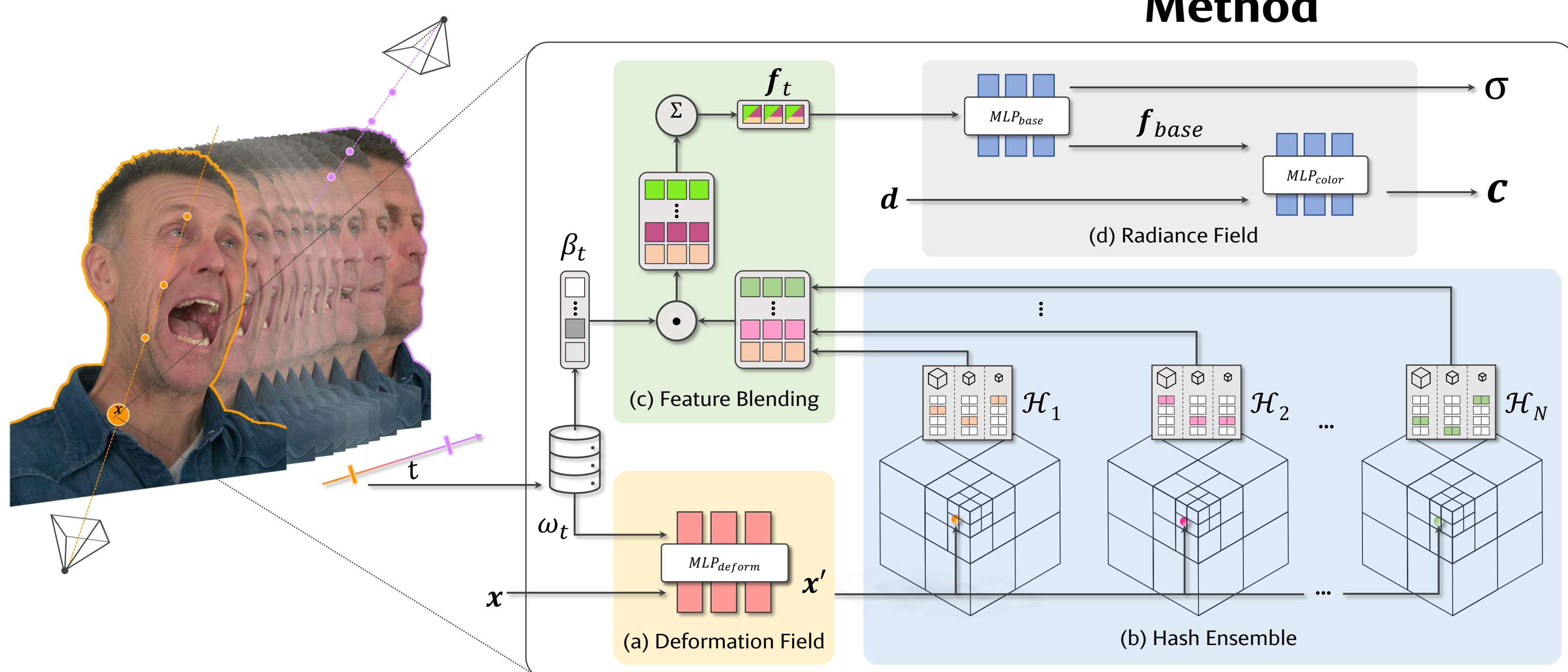


## Results



	Method	PSNR ↑	SSIM ↑	LPIPS ↓
Static	PSR	12.5	0.774	0.341
	Instant NGP	28.8	0.864	0.254
Dynamic	Nerfies	29.5	0.849	0.299
	HyperNeRF	29.6	0.848	0.304
	DyNeRF	30.6	0.860	0.254
	<b>NeRsemble</b>	<b>31.8</b>	<b>0.875</b>	<b>0.212</b>
Parts	NGP + Def.	30.8	0.864	0.231
	Hash Ensemble	30.5	0.857	0.257
Ablations	w/o Depth	31.5	0.873	0.217
	w/o Warmup	31.0	0.866	0.234
	only 16 tables	31.5	0.871	0.218

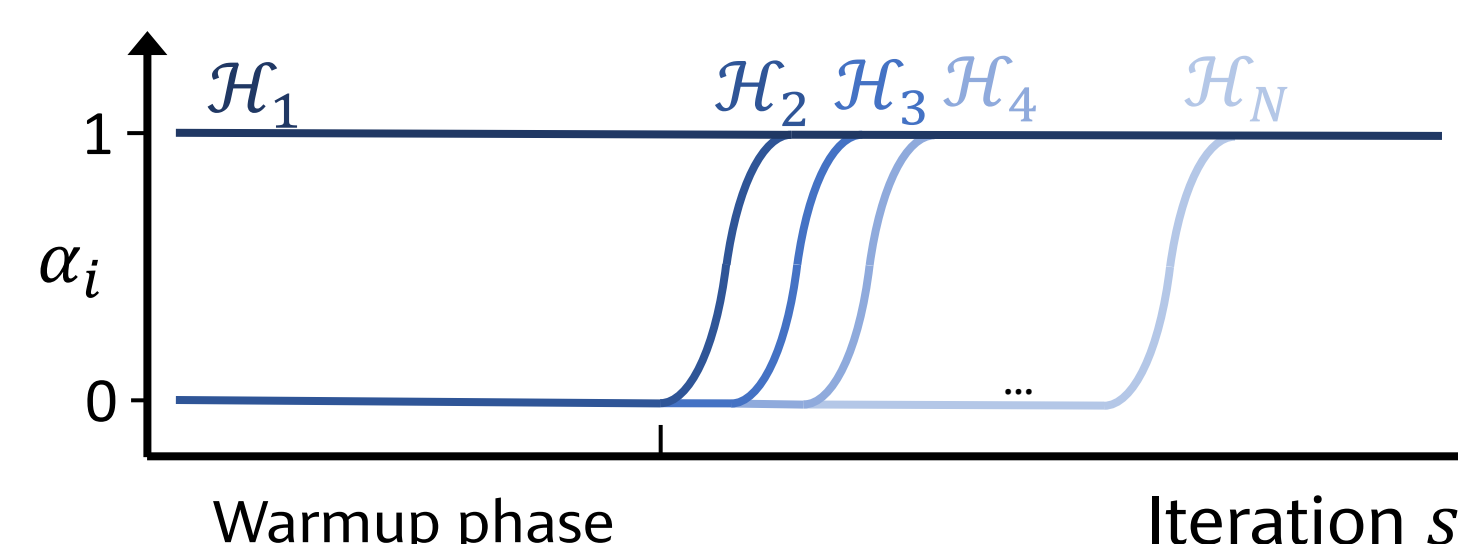
## Method



$$f_t = \sum_{i=1}^N \beta_{t,i}^* \mathcal{H}_i(\mathcal{D}(x, \omega_t))$$

## Hash grid schedule

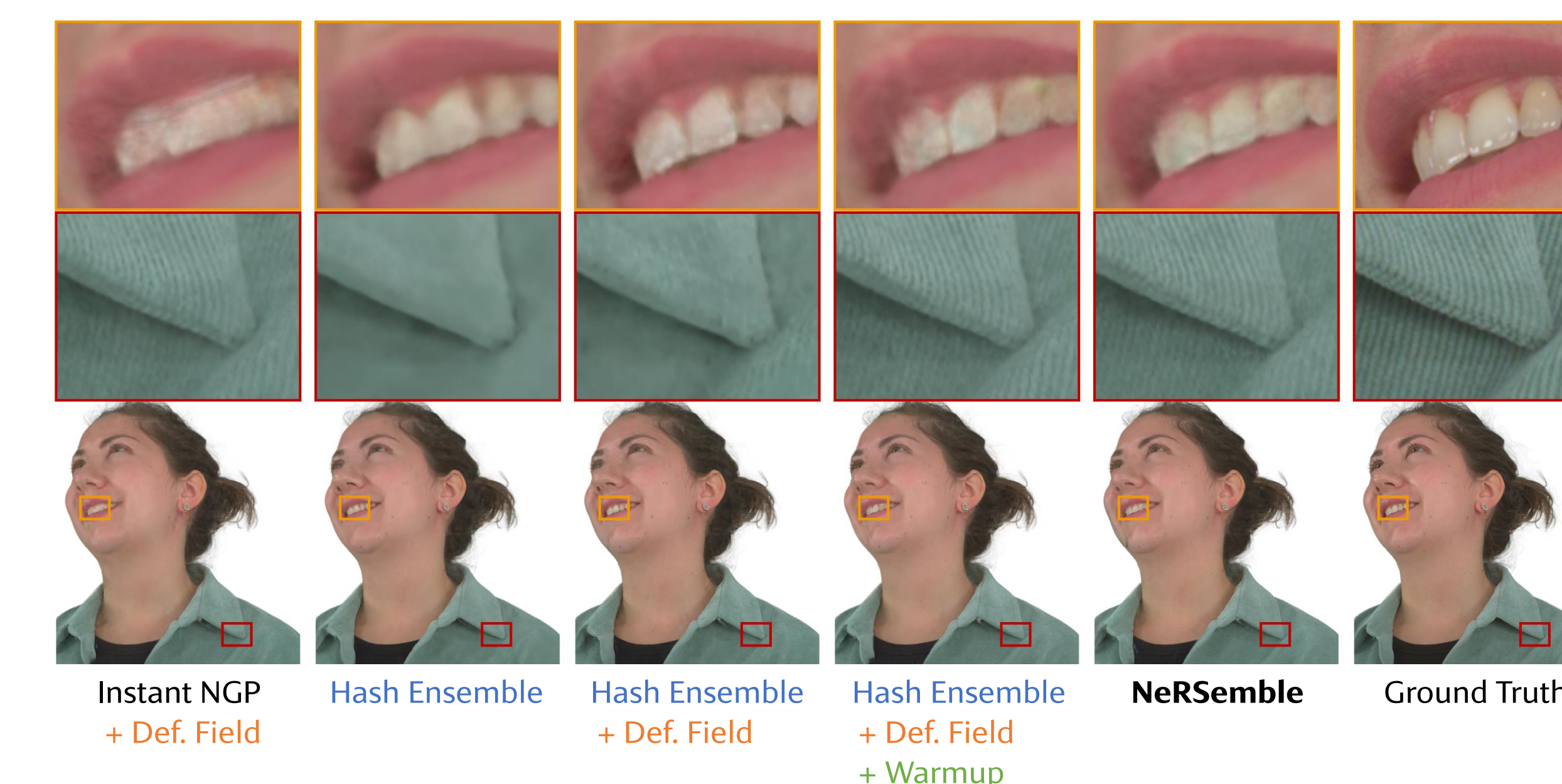
$$\beta_{t,i}^* = \alpha_i(s) \beta_{t,i}$$



## Project Page



## Ablation



## Conclusion

NeRsemble combines a **Deformation Field** (rough motion) with an **Ensemble of Hash Encodings** (fine motion)

High quality dynamic novel view synthesis

Released a large dataset of multi-view video recordings of faces