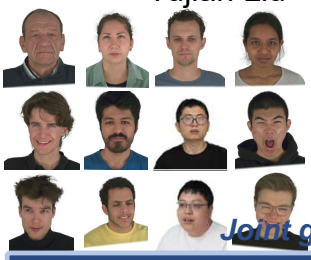
PRCV 2026 中国 · 哈尔滨
The 9th Chinese Conference on Pattern Recognition and Computer Vision
22-25, August, 2026, Harbin

与爱为舞 YuAiWeiWu

MoGaFace: Momentum-Guided and Texture-Aware Gaussian Avatars for Consistent Facial Geometry

Yujian Liu[1,2], Dongxu Shen[3], Chuang Chen[1], Zairan Wang[1], Linlang Cao[1], Fanyu Geng[1], Peng Cao[4], Shidang Xu[2*], Xiaoli Liu[1*]

[1]AiShiWeiLai AI Research,
[2]South China University of Technology
[3]Hong Kong University of Science and Technology (Guangzhou)
[4]Northeastern University



Joint geometry and texture refinement for high-fidelity 3D Gaussian head avatars.

1. Motivation

Existing FLAME-based Gaussian avatar methods suffer from:

(i) Geometry-image misalignment across facial regions

(ii) Cross-view Inconsistency across synchronized observations

(iii) Weak texture refinement for fine-grained facial details

2. Contribution

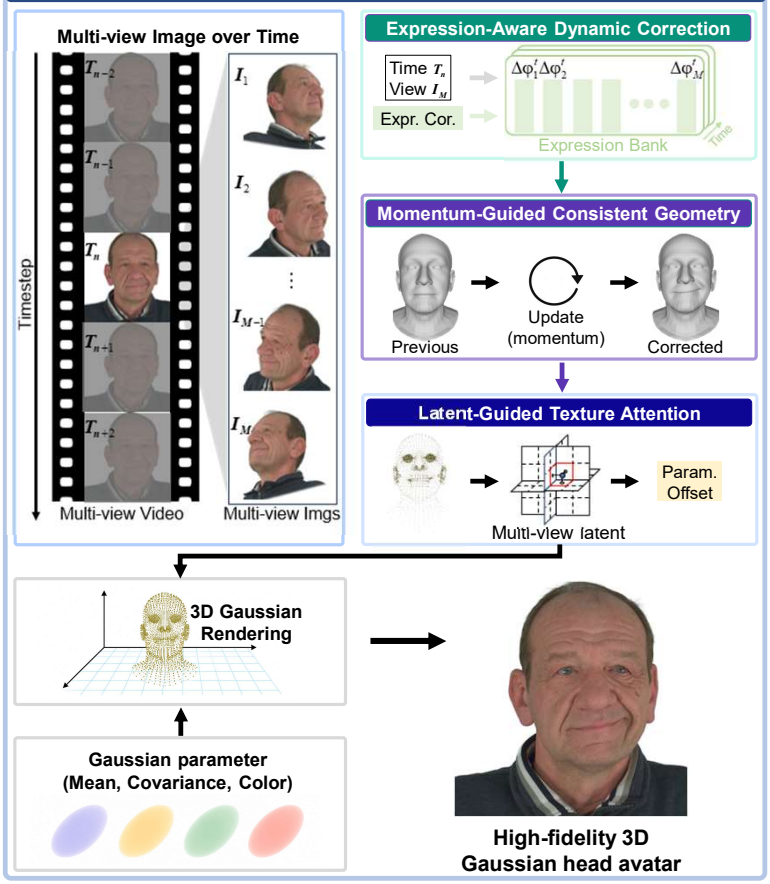
1 Expression-Aware Dynamic Correction
Refines FLAME expressions using multi-view image cues.

2 Momentum-Guided Consistent Geometry
Uses an expression bank to enforce multi-view consistency.

3 Latent-Guided Texture Attention
Injects multi-view latent cues for fine-grained texture refinement.

Works in both camera-aware and camera-free settings.

3. Method Overview



4. Experimental Results

Quantitative results (camera-aware)

Novel-view synthesis (NVS)

Method	PSNR	SSIM	LIPIS
GA	29.71	0.939	0.079
SurFHead	30.02	0.941	0.084
MoGaFace	**30.98**	**0.946**	**0.068**

Self-reenactment

Method	PSNR	SSIM	LIPIS
GA	23.58	0.891	0.093
SurFHead	23.71	0.892	0.098
MoGaFace	**24.70**	**0.900**	**0.084**

Self-reenactment (NVS)

Method	PSNR	SSIM	LIPIS
GA	23.01	0.900	0.090
SurFHead	23.55	0.901	0.098
MoGaFace	**24.75**	**0.902**	**0.090**

+0.96 dB PSNR over the best baseline in NVS

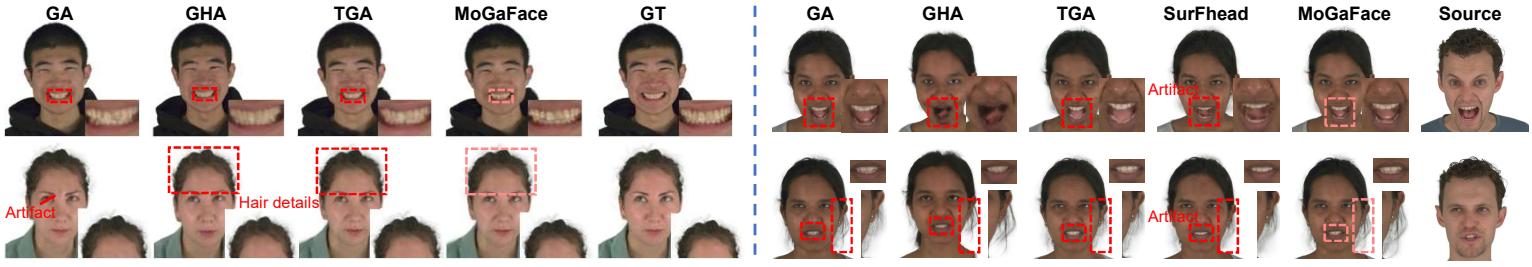
Camera-free results

Self-reenactment

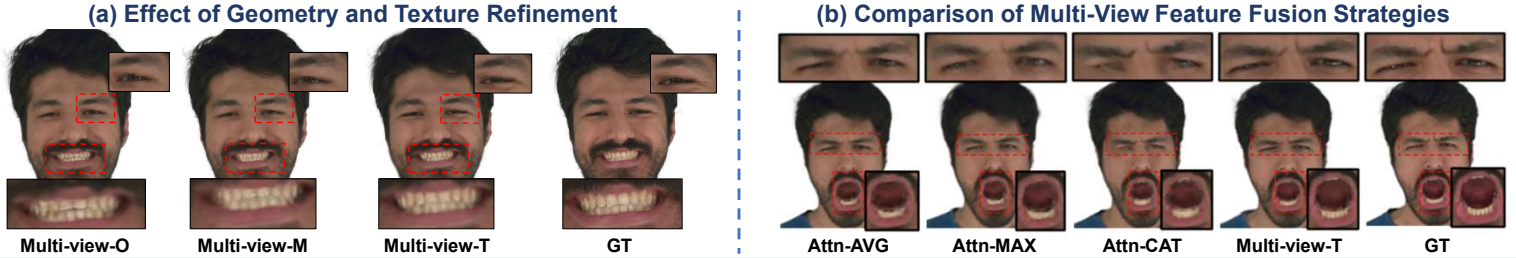
Method	PSNR	SSIM	LIPIS
GA	26.83	0.938	0.034
SurFHead	27.17	0.940	0.034
MoGaFace	**28.47**	**0.943**	**0.031**

+1.30 dB PSNR over the best baseline

5. Qualitative comparison



6. Ablation study



Conclusion

Jointly refines geometry and texture during rendering for realistic and controllable 3D head avatars.

Robust in camera-free real-world scenarios with improved cross-view consistency and rendering stability.

Training converges in *~6 hours*. Inference runs at *24.3 FPS* and reaches *28.6 FPS on RTX 4090*.