LentiAvatar: Pseudo-Multiview Reconstruction and Subpixel Prism Rendering for Real-Time Stereoscopic Communication

Chufeng Fang
fangchf3@mail2.sysu.edu.cn
Sun Yat-sen University
Guangzhou, China

Dongdong Teng
tengdd@mail.sysu.edu.cn
Sun Yat-sen University
Guangzhou, China

Lilin Liu*
liullin@mail.sysu.edu.cn
Sun Yat-sen University
Guangzhou, China

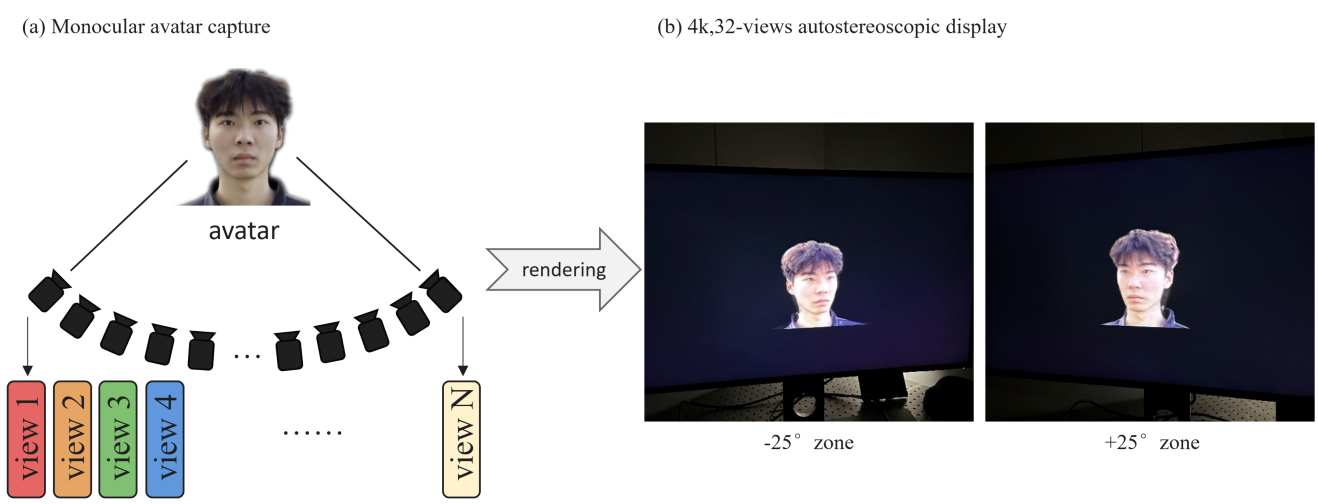

Figure 1: LentiAvatar overview. (a) A monocular portrait sequence is used to reconstruct a Gaussian head avatar, which is rendered from multiple virtual display viewpoints. (b) The rendered views are encoded into a 4K, 32-view glasses-free display raster; captured viewing zones at −25° and +25° show the view-dependent avatar appearance delivered by the lenticular panel.

Abstract

Real-time stereoscopic video communication has long been a goal of immersive telepresence, yet practical systems still require specialized capture rigs or reduce remote users to a single portrait view. We present LentiAvatar, a Gaussian head-avatar system that connects monocular avatar capture with subpixel-encoded glasses-free lenticular display for real-time autostereoscopic communication. From a monocular portrait video, LentiAvatar reconstructs a controllable head avatar and optimizes it for the lateral viewing zones induced by the display. The method uses natural head turns as pseudo-multiview (PMV) supervision to constrain regions that are otherwise weakly observed in monocular training, including hair, ears, jaw contours, and neck boundaries. Reliable side frames are yaw-binned, aligned to virtual cameras, and supervised within a strict head-and-hair domain; contour-aware losses and staged regularization further suppress ghosting, alpha leakage, and depth instability while preserving lateral detail. At runtime, LentiAvatar renders 32 virtual views and encodes them into a 4K lenticular raster with calibrated subpixel-routing masks. The live-tracker prototype sustains 10.65 FPS, and a subject-specific distilled driver raises the same display pipeline to 38.49 FPS.

Keywords

Gaussian avatars, pseudo-multiview reconstruction, glasses-free 3D communication, autostereoscopic displays

1 Introduction

Real-time portrait communication on glasses-free 3D displays requires the avatar pipeline to reconstruct, render, and route multiple views consistently. Earlier telepresence and portrait-reenactment systems connected real-time capture, tracking, and communication, but typically rely on specialized capture infrastructure or remain single-view video renderers [Kim et al. 2018; Orts-Escolano et al. 2016; Thies et al. 2016, 2018]. Neural radiance fields and Gaussian head avatars have made personalized facial geometry and appearance practical from monocular video [Feng et al. 2025; Gafni et al. 2021; Li et al. 2025a; Mildenhall et al. 2020; Qian et al. 2024a; Xiang et al. 2024; Zheng et al. 2022]. INSTA demonstrates minute-level monocular neural-head reconstruction [Zielonka et al. 2023], FlashAvatar and RGBAvatar study efficient monocular Gaussian avatars [Li et al. 2025a; Xiang et al. 2024], and GaussianAvatars and GPAvatar demonstrate high-quality rigged or projection-efficient Gaussian heads [Feng et al. 2025; Qian et al. 2024a]. On a lenticular display, however, the avatar is not judged from one frontal rendering. Multiple synthesized views are routed to calibrated raster subpixels and separated by optics into viewing zones, so side-view

*Corresponding author.

opacity, boundary, or color errors can appear as depth shimmer, ghosting, or color fringing.

This setting highlights a limitation in existing monocular avatar reconstruction. Front-dominant training video strongly constrains the face seen by the input camera, but it gives sparse evidence for the ears, hair silhouette, jawline, posterior head, and neck boundary. Recent Gaussian avatar systems render efficiently and with high frontal fidelity [Feng et al. 2025; Li et al. 2025a; Qian et al. 2024a; Xiang et al. 2024], yet their usual monocular objectives do not explicitly train the lateral views that a multiview display reveals. Directly adding side supervision can be unreliable because the head and hair approximately follow head rotation, while the neck, collar, and background do not; treating them alike can introduce neck-rear ghosting, alpha shells, skin-colored collar contamination, or blurred lateral contours.

Our key observation is that ordinary head turns already contain the missing side-view evidence when they are treated as view-specific supervision rather than as generic temporal frames. A turned frame observes the face, hair, ear, and jaw contour at a particular yaw, and can supervise the virtual camera with the same yaw after alignment refinement and above-neck masking. This converts monocular head motion into a practical pseudo-multiview (PMV) signal for the lateral views exposed by the display.

LentiAvatar addresses stereoscopic video communication with a monocular Gaussian head avatar. During training, it selects natural head turns as PMV observations, ranks and snaps them to yaw bins, refines virtual-camera alignment, gates unreliable matches, and applies supervision inside a strict above-neck matte. Contour-preserving alpha and edge terms retain hair, ear, and jaw boundaries, while shell and collar controls suppress off-surface opacity and color leakage. During display-time stereoscopic rendering, the driven Gaussian avatar is rendered from virtual display views, encoded through calibrated subpixel masks, rasterized as a 4K panel frame, and presented as a glasses-free 3D avatar rather than a conventional single-view portrait. Figure 2 summarizes this reconstruction-to-display pipeline.

We evaluate LentiAvatar with public benchmark comparisons, component analysis, and 4K display-side profiling. On Marcel, LentiAvatar obtains the lowest outside-mesh alpha among the compared methods and ranks second on both neck-rear ghost measures, neck-rear smear, and alpha translucency smear. The live-tracking stereoscopic video communication system reaches 10.65 FPS in the 4K, 32-view configuration after initialization; a subject-specific student driver raises the same display pipeline to 38.49 FPS, confirming that live tracking is the dominant frame-time cost rather than subpixel composition.

Our contributions are:

- A reconstruction-to-display Gaussian head-avatar pipeline for glasses-free 3D communication that reconstructs a controllable avatar from monocular video and encodes 32 rendered views into a calibrated 4K autostereoscopic raster.
- PMV supervision that turns natural monocular head rotations into yaw-matched side observations for lateral display views.
- A reliability-aware side-view objective with strict head-and-hair masking, side-frame ranking, alignment gates, contour losses, and shell/collar regularization.

2 Related Work

2.1 Monocular and Gaussian avatars

Parametric face models provide the tracking and semantic support used by many avatar systems, from 3D morphable models and FLAME to robust in-the-wild fitting [Blanz and Vetter 1999; Feng et al. 2021; Li et al. 2017]. Neural rendering and NeRF-based avatars learned dynamic face appearance from controlled or monocular capture [Athar et al. 2022; Barron et al. 2022; Duan et al. 2023; Gafni et al. 2021; Grassal et al. 2022; Lombardi et al. 2018, 2019; Ma et al. 2021; Mildenhall et al. 2020; Müller et al. 2022; Park et al. 2021a,b; Thies et al. 2019; Zheng et al. 2022, 2023; Zielonka et al. 2023], while real-time portrait reenactment and telepresence systems established the communication setting [Kim et al. 2018; Orts-Escolano et al. 2016; Thies et al. 2016, 2018]. 3D Gaussian Splatting [Kerbl et al. 2023; Zwicker et al. 2001] shifted avatar rendering toward explicit primitives, enabling deformable humans and head avatars with mesh embeddings, rigged Gaussians, efficient embeddings, parametric control, blendshape compression, relighting, tensorial appearance, or hybrid mesh-Gaussian editing [Chen et al. 2024; Dhamo et al. 2024; Feng et al. 2025; Giebenhain et al. 2024; Kocabas et al. 2024; Li et al. 2025a; Ma et al. 2024; Moon et al. 2025; Qian et al. 2024a,b; Saito et al. 2024; Shao et al. 2024; Wang et al. 2025a,d,b,c; Xiang et al. 2024; Xu et al. 2024; Zhang et al. 2025a]. LentiAvatar builds on this explicit-avatar line, but targets the display-coupled failure case where lateral viewing zones expose weakly constrained side contours and neck regions.

2.2 Missing-view and pseudo-view priors

Dense view coverage improves head reconstruction, as shown by multi-view datasets and systems such as NeRSemble [Kirschstein et al. 2023]. Monocular and single-image methods instead infer unobserved regions with learned or generated priors, including diffusion-based Gaussian avatars, pseudo multi-view head synthesis, single-image Gaussian avatars, and related human or upper-body reconstruction methods [Deng et al. 2024; Li et al. 2025b; Qiu et al. 2025; Tang et al. 2025; Zhang et al. 2025b; Zhuang et al. 2025]. These priors improve missing-view plausibility but are not tied to the physical viewing zones of a glasses-free panel. LentiAvatar instead extracts real side evidence from natural head turns, then restricts it with yaw-binning, alignment gates, and a strict head-and-hair domain so that collar, lower-neck, and background pixels do not become side-view supervision.

2.3 Glasses-free and subpixel display

Glasses-free displays are rooted in light-field and image-based rendering, where a scene is sampled as directional views and routed to observer zones [Dodgson 2005; Gortler et al. 1996; Levoy and Hanrahan 1996; Zwicker et al. 2006]. Prior autostereoscopic and computational display systems studied dynamic-scene acquisition, multiview rendering, parallax barriers, multilayer/tensor displays,

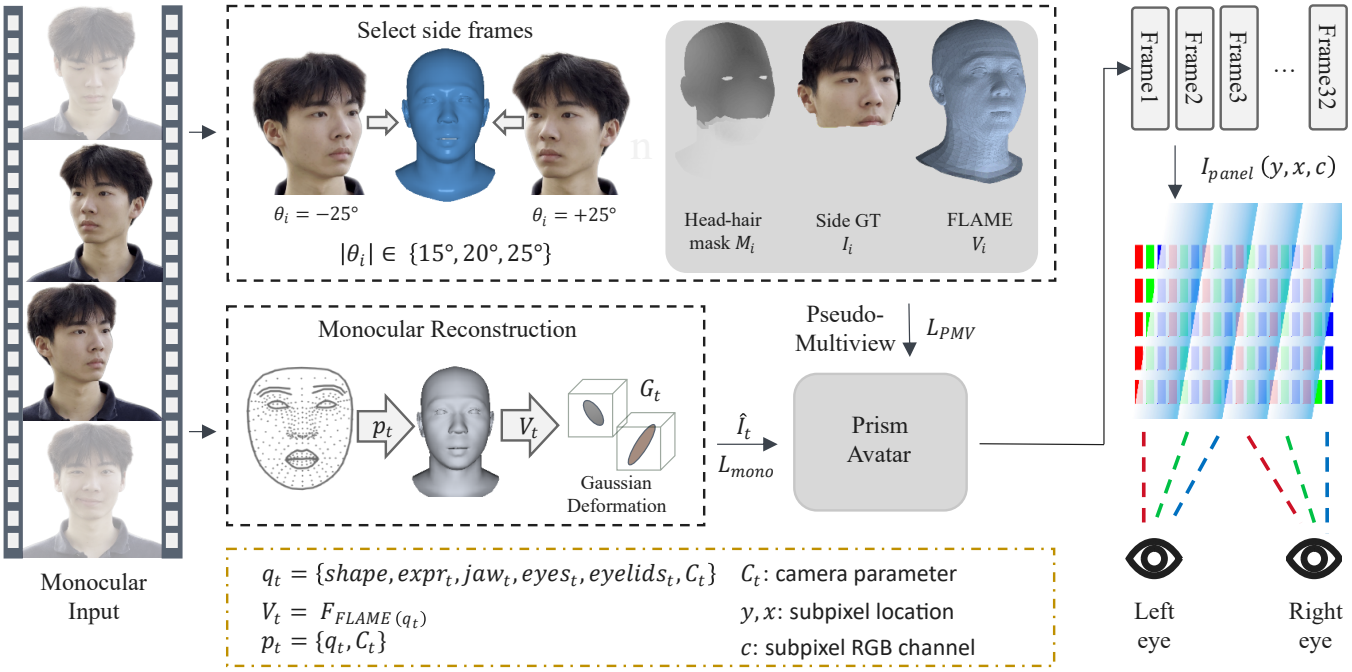

Figure 2: LentiAvatar pipeline. Given a monocular portrait sequence, LentiAvatar first reconstructs a controllable Gaussian head avatar from FLAME-based controls and monocular supervision. Natural head turns are then selected as yaw-binned pseudo side observations, aligned to the corresponding virtual cameras, and restricted to an above-neck matte before applying the PMV loss. At runtime, the trained avatar is driven by the current controls, rendered into multiple display views, and routed through calibrated subpixel masks to form the panel raster $I_{\text{panel}}(y, x, c)$ for glasses-free stereoscopic presentation.

and tiled multi-view VR [Jones et al. 2007; Kooima et al. 2010; Lanman et al. 2010; Matusik and Pfister 2004; Wetzstein et al. 2012]. For lenticular panels, GPU interleaving, chroma or subpixel multiplexing, directional subpixel rendering, and recent light-field display rendering show that RGB subpixel layout, lens slant, and calibrated view routing are part of the rendering model [Lee et al. 2018; Marson and Stern 2015; Pei et al. 2016; Ruijters 2008; Yang et al. 2024]. Adjacent avatar-communication systems use dense capture or XR devices [Chen et al. 2025; Orts-Escolano et al. 2016]; our setting instead couples monocular avatar reconstruction, multiview Gaussian rendering, and calibrated subpixel routing for a live glasses-free panel.

3 Method

The training input is a monocular portrait sequence $\{I_t, A_t, \mathbf{z}_t, y_t\}_{t=1}^{T}$, where I_t is the RGB frame, A_t is the foreground alpha or matte, $\mathbf{z}_t$ contains the tracked expression, jaw, eye, eyelid, and head-pose controls, and y_t is the estimated camera-relative view yaw. We assume the monocular camera calibration is available from the base avatar pipeline and the glasses-free panel provides calibrated raster-routing constants. LentiAvatar has two coupled goals: reconstruct side-stable avatar geometry and appearance from natural head turns, and encode the driven avatar into a calibrated subpixel raster for glasses-free binocular disparity.

The geometric assumption is conservative: the head and hair approximately co-rotate in selected side frames, while the lower neck, collar, and background are unreliable supervision domains. LentiAvatar therefore uses temporal side observations only after yaw matching, alignment refinement, and strict above-neck masking.

3.1 Gaussian avatar representation

Gaussian splatting avatar. We use an explicit Gaussian avatar because it supports fast multi-view rendering for the display path. Following surface splatting and 3D Gaussian Splatting [Kerbl et al. 2023; Zwicker et al. 2001], the avatar $\mathcal{G} = \{g_j\}_{j=1}^{M}$ stores anisotropic primitives with position $\boldsymbol{\mu}_j$, covariance Σ_j, opacity o_j, and color features $\mathbf{c}_j$. As in 3DGS, the covariance is parameterized by a rotation R_j and diagonal scale matrix S_j,

$$\Sigma_j = R_j S_j S_j^{\top} R_j^{\top}, \tag{1}$$

which ensures a positive semidefinite anisotropic support for each primitive. The control code $\mathbf{z}_t$ deforms the canonical avatar into view-conditioned frame parameters,

$$\mathcal{G}_{t,\theta} = \mathcal{D}_{\Theta}(\mathcal{G}, \mathbf{z}_t, \theta), \qquad (\hat{I}_{t,\theta}, \hat{\alpha}_{t,\theta}) = \mathcal{R}(\mathcal{G}_{t,\theta}, \Pi_{\theta}), \tag{2}$$

where θ is the requested virtual-view yaw and Π_{θ} is the corresponding camera. We use the standard front-to-back alpha compositing of 3DGS and recent Gaussian avatar systems [Feng et al. 2025; Kerbl et al. 2023; Li et al. 2025a; Qian et al. 2024a; Xiang et al. 2024]; the primitive representation is not the main novelty of LentiAvatar.

Our contribution is the display-coupled side-view supervision and routing. For side-view supervision, we project FLAME-derived semantic weights for lateral face, ears, chin, jawline, neck, and neck contours into pseudo-views, yielding reliable head support and high-risk neck/collar regions.

3.2 Side-frame selection and pseudo-view cameras

Yaw-binned selection. LentiAvatar constructs pseudo-view observations from natural head rotations rather than using every turned frame. Let $\mathcal{B} = \{15^\circ, 20^\circ, 25^\circ\}$ be the lateral bins used by our display probes. For each frame, we compute

$$\begin{aligned} b_t &= \arg\min_{b\in\mathcal{B}} \big||y_t| - b\big|, \\ a_t &= \mathbf{1}\big[14^\circ \le |y_t| \le 27^\circ\big]\ \mathbf{1}\big[\big||y_t| - b_t\big| \le 2.5^\circ\big], \end{aligned} \tag{3}$$

and accept frame t only when $a_t = 1$. The supervised virtual-view yaw and context rewrite are

$$\hat{y}_t = \mathrm{sign}(y_t) b_t, \qquad \mathbf{z}_t^{\mathrm{pose}} \leftarrow (0, p_t), \quad \mathbf{z}_t^{\mathrm{view}} \leftarrow (\hat{y}_t, 0), \tag{4}$$

where p_t is the tracked pitch. Thus the frame provides a side-view observation for a yaw-matched virtual camera, while the avatar pose context remains near frontal. This keeps PMV supervision one-to-one and avoids treating a physical head turn as an ordinary monocular training view.

Side-frame ranking. Yaw alone is not sufficient because large turns may contain motion, expression changes, poor matte boundaries, or collar contamination. We assign each yaw-valid candidate a score

$$q_t = \sum_m w_m\, Q_m(t) - \sum_n \lambda_n\, R_n(t), \tag{5}$$

where the positive terms score yaw-bin agreement, temporal and expression stability, matte/edge cleanliness, and visible lateral face or hair support, while the risk terms penalize boundary instability and collar contamination. We keep only high-scoring frames per yaw bin and side direction, producing a compact, balanced PMV set with left and right coverage rather than trusting all side-looking frames equally.

Virtual cameras. For both PMV supervision and runtime display, LentiAvatar renders horizontal virtual views on a camera arc around the avatar. Let $\mathbf{c}_0$ be the avatar bounding-box center, h the head-support height, and $\mathbf{e}_y$ the vertical axis. We use

$$\begin{aligned} \mathbf{c} &= \mathbf{c}_0 + 0.08h\,\mathbf{e}_y, \\ \mathbf{o}(\theta) &= \mathbf{c} + r[\sin\theta, 0, \cos\theta]^\top, \\ \Pi_\theta &= \mathrm{LookAt}(\mathbf{o}(\theta), \mathbf{c}, K), \end{aligned} \tag{6}$$

where r is the calibrated avatar-camera radius and K is the perspective intrinsic matrix. During training, $\theta = \hat{y}_t$; during display, the same construction samples 32 virtual views and routes them to the autostereoscopic raster in Sec. 3.4.

3.3 PMV reconstruction

Camera and target alignment. Even after yaw binning, real side frames and virtual cameras are not perfectly aligned because monocular tracking, cropping, and hair silhouettes are imperfect. LentiAvatar therefore applies a bounded local search over camera yaw, radius, screen scale, and translation, followed by a small target-image similarity warp. Both steps maximize support overlap, boundary agreement, and centroid consistency while penalizing large corrections. The resulting side target is used only when the alignment gate passes:

$$g_t^{\mathrm{align}} = \mathbf{1}[\mathrm{IoU} \ge 0.32]\ \mathbf{1}[\mathrm{IoU}_e \ge 0.12]\ \mathbf{1}\big[d_\mu \le 28\ \mathrm{px}\big], \tag{7}$$

with additional implementation bounds on shift, scale, and rotation. Failed matches are downweighted rather than trusted as full supervision.

Strict head-and-hair supervision domain. Let $\hat{I}_t$ and $\hat{\alpha}_t$ be the RGB image and alpha map rendered from the aligned pseudo-view camera, and let I_t^* and A_t^* be the aligned side-frame target. To exclude unreliable non-head regions, we first define the side-view support as a positive-minus-risk semantic mask:

$$P_t(u) = \mathbf{1}[S_t(u) > \tau_s]\ \mathbf{1}[R_t(u) < \tau_r], \tag{8}$$

where S_t rasterizes positive side-head weights and R_t rasterizes neck and jaw-neck risk weights. The positive support covers the visible lateral face, ears, hair-adjacent head support, chin, and jawline; the risk support marks lower-neck, collar-adjacent, and jaw-neck regions that are unreliable in the monocular side frame. The strict PMV target matte is

$$M_t^{\mathrm{hh}}(u) = A_t^*(u) \odot \mathcal{D}_{25}(P_t)(u) \odot B_t(u), \tag{9}$$

where $\mathcal{D}_{25}$ denotes a 25-pixel dilation and B_t is a per-column bottom gate computed from the projected support boundary using a high support quantile, a small pixel extension, and a canonical FLAME height cutoff. This strict domain is the key difference from naive PMV: the RGB target may contain neck, collar, or background pixels, but those pixels do not supervise the side-view avatar.

Reliability-gated side objective. Direct RGB side supervision is applied only in stable interior pixels. We erode the target matte, suppress boundary bands, partial-alpha pixels, and high-chroma regions, and remove RGB supervision from boundary pixels where small alignment errors could imprint color fringes on the Gaussian appearance. The scheduled training objective is

$$\begin{aligned} \mathcal{L}(k) &= \mathcal{L}_{\mathrm{base}} + \gamma(k)\mathcal{L}_{\mathrm{pmv}}, \\ \mathcal{L}_{\mathrm{pmv}} &= \lambda_I\mathcal{L}_I + \lambda_\alpha\mathcal{L}_\alpha + \lambda_{\mathrm{bg}}\mathcal{L}_{\mathrm{bg}} + \lambda_{\mathrm{mesh}}\mathcal{L}_{\mathrm{mesh}} \\ &\quad + \lambda_e\mathcal{L}_e + \lambda_s\mathcal{L}_s. \end{aligned} \tag{10}$$

$\mathcal{L}_{\mathrm{base}}$ includes photometric, SSIM, foreground-alpha, background-alpha, and regularization terms. In $\mathcal{L}_{\mathrm{pmv}}$, $\mathcal{L}_I$ and $\mathcal{L}_\alpha$ supervise safe RGB and head-domain alpha pixels; $\mathcal{L}_{\mathrm{bg}}$ and $\mathcal{L}_{\mathrm{mesh}}$ suppress opacity outside the target support and dilated FLAME mesh; $\mathcal{L}_e$ preserves alpha gradients on ears, hair, and jaw boundaries; and $\mathcal{L}_s$ penalizes semi-transparent smear through the standard $4\hat{\alpha}(1-\hat{\alpha})$ response. For reproducibility, the base weights for L1, SSIM, foreground alpha, and background alpha are $(1.0, 0.08, 0.05, 0.35)$. The PMV weights $(\lambda_I, \lambda_\alpha, \lambda_{\mathrm{bg}}, \lambda_{\mathrm{mesh}}, \lambda_e, \lambda_s)$ are $(0.016, 0.012, 0.025, 0.012, 0.035, 0.004)$ for side learning and $(0.016, 0.004, 0.012, 0.006, 0.026, 0.010)$ for color

fine-tuning. PMV starts at iteration 320, ramps linearly for 700 iterations, and is capped by $\gamma(k) \leq 0.65$; the later no-PMV stages disable $\mathcal{L}_{\text{pmv}}$. Side-aware shell/background cleanup is applied only on mesh and edge supports, using alpha strengths 0.05/0.07 for the two PMV stages, color strength 0.012, and boundary-chroma strength 0.85; the collar anti-skin loss is disabled. This keeps the PMV objective localized to lateral artifacts rather than imposing broad penalties that would remove valid face, hair, ear, or jaw detail.

Staged stabilization. The final model uses a four-stage stabilization schedule: PMV side learning with auto-alignment, strict head-and-hair masking, shell cleanup, and contour preservation; PMV color fine-tuning with milder alpha/background weights; no-PMV low-weight monocular color correction with blend features and the weight module frozen; and a short no-PMV cleanup stage with base and blend color features frozen. This lets side-view evidence shape the avatar first, then removes residual color leakage without globally pruning the face, hair, ear, or jaw details learned from PMV supervision.

3.4 Real-time driving and subpixel display encoding

Subject-specific distilled driver. The display prototype is limited by the per-frame metrical tracker, so we optionally distill the tracker into a subject-specific feed-forward RGB driver. The teacher targets are the 129-D grouped deformation controls already produced for the training sequence; a MobileNetV3-Small backbone with a two-layer MLP head predicts the normalized control vector from a cropped 224×224 portrait frame using a group-weighted smooth-L_1 distillation loss over expression, neck, jaw, eye, eyelid, and translation groups. At inference, the prediction is denormalized with subject statistics and passed to the same Gaussian deformation module as tracker output. The avatar, 32-view renderer, and subpixel compositor are unchanged, so this module only reduces the tracking-stage cost reported in Sec. 4.

Subpixel prism encoding. We encode the multiview avatar output at the level of RGB subpixels.

Classic glasses-free displays create viewpoint-dependent images by optically directing different panel samples to lateral viewing zones [Dodgson 2005; Matusik and Pfister 2004; Wetzstein et al. 2012]. For lenticular, barrier, or grating-based panels, rasterization therefore becomes a view assignment problem at the subpixel level rather than conventional image compositing [Lee et al. 2018; Pei et al. 2016; Yang et al. 2024]. Figure 3 illustrates the physical mapping: the grating couples subpixel position with outgoing angle, so neighboring RGB subpixels can contribute to different observer viewpoints. LentiAvatar uses the calibrated panel constants to instantiate this mapping as binary routing masks for the virtual avatar views.

Let $E \in [0,1]^{H\times W\times 3}$ be the encoded output raster and let $V_i \in [0,1]^{H_v\times W_v\times 3}$ be the rendered image for virtual view $i \in \{1,\dots,N\}$. A subpixel is denoted by $p = (x, y, c)$, where (x, y) is the output pixel location and c is the color channel after the configured RGB/BGR order. Let $\rho(c) \in \{0, 1, 2\}$ map the color channel to its subpixel offset inside an output pixel. The display constants are the raster slant coefficient C_r, the subpixel routing period Δ_r, the reference offset

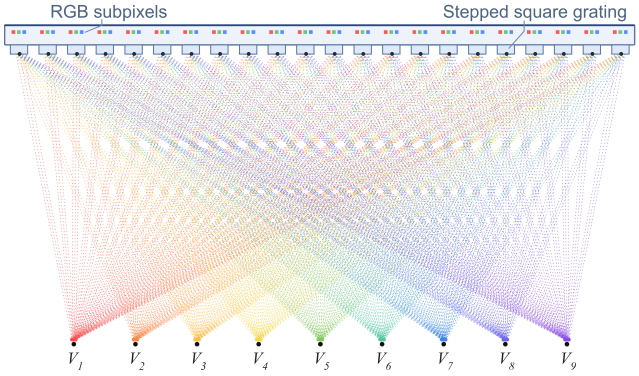

Figure 3: Subpixel-to-view routing in the glasses-free display. RGB subpixels lie underneath a calibrated grating layer, which redirects repeated subpixel positions toward different lateral viewing zones. Colored ray bundles indicate the angular distribution of the routed subpixels. LentiAvatar uses this calibrated routing in the reverse direction: each rendered virtual avatar view is written only to the subpixels whose outgoing rays reach the corresponding viewing zone.

s_{ref}, and the reference view index i_{ref}. For view i, the row-dependent offset is

$$s_i = s_{\text{ref}} - \frac{\Delta_r}{N}(i - i_{\text{ref}}). \tag{11}$$

For row y and integer period index m, the candidate subpixel coordinate is

$$b_i(y, m) = 3C_r y + s_i + m\Delta_r. \tag{12}$$

View i contributes to subpixel p when the subpixel index $x_s = 3x + \rho(c)$ equals $\lfloor b_i(y, m) \rfloor$ for some m, and the fractional part of $b_i(y, m)$ is smaller than Δ_r/N. This defines a binary mask $M_i(p)$. With u_p the bilinear sample coordinate in the rendered view corresponding to output subpixel p, the encoded raster is

$$E(p) = \sum_{i=1}^{N} M_i(p)\, V_i(u_p). \tag{13}$$

Runtime implementation. The display implementation renders $N = 32$ virtual views at 960×540, uniformly sampled over the avatar-viewing range $[-25°, +25°]$, and composes them into a 3840×2160 panel raster. The calibrated masks are precomputed once into sparse GPU sampling tables with primary and secondary view selectors; configurations with more than two active views at a subpixel are rejected. At each frame, the current controls deform the Gaussian avatar, the cached view renderers produce the virtual views, optional mesh-visibility gates suppress unreliable lateral background leakage, and alpha-boundary chroma attenuation reduces residual boundary color artifacts. Primary assignments use indexed overwrite into a half-precision output buffer, while secondary overlaps are accumulated by indexed addition. The final tensor is converted to the calibrated panel format for display.

4 Experiments

4.1 Setup

The experiments test two claims: PMV reconstruction improves side-view robustness for monocular avatars, and the reconstructed avatar can be routed to a 4K glasses-free display at interactive rates. We compare qualitative behavior on public **INSTA** benchmark sequences [Zielonka et al. 2023] with stable horizontal head turns against RGBAvatar [Li et al. 2025a], HRAvatar [Zhang et al. 2025a], FlashAvatar [Xiang et al. 2024], and INSTA [Zielonka et al. 2023]. These public-sequence comparisons are used to validate the reconstruction side of the method: if PMV supervision improves the side views reconstructed from monocular video, the same gains can be carried into the lateral viewing zones of the glasses-free display. Quantitative diagnostics use Marcel, rendered at -25° and $+25^\circ$ from frame 0 with a common aligned crop and 9-pixel FLAME-mask dilation.

For the component study, the avatar is trained with the four-stage schedule in Sec. 3.3: PMV side learning, PMV color fine-tuning, no-PMV low-weight color correction, and conservative no-PMV cleanup. The probe protocol follows the glasses-free display setting: 0° checks the central viewing zone, while $\pm 25^\circ$ checks the lateral zones that are most sensitive to hair, ear, jaw, and neck artifacts.

Figure 4 uses a fixed -25° lateral probe on smooth-turning INSTA sequences. The comparison shows that monocular baselines can preserve plausible reference-view appearance while producing floating opacity, weak side contours, or neck/collar contamination under large yaw. In our setting, this qualitative benchmark is not an endpoint by itself: it verifies that PMV reconstruction improves exactly the side-angle content that will later be replicated across the routed display views. These errors directly affect glasses-free presentation because neighboring synthesized views are routed to different viewing zones, where boundary leakage and side-view instability can appear as depth shimmer, ghosting, or color fringing during stereoscopic communication.

4.2 Evaluation metrics

Standard image metrics do not isolate the artifacts exposed by side viewing zones, so we report six lower-is-better diagnostics matched to Table 1. For each side view, let C be RGB color, A be alpha, M be the FLAME mesh mask, $D = \mathrm{dilate}(M, 9)$, and $N = \{0.46 \le y/H \le 0.90,\ x/W \le 0.42 \text{ or } x/W \ge 0.58\}$. We measure outside support opacity, neck-rear ghosting, and translucent smear as

$$
\begin{aligned}
O_\alpha &= \frac{\sum_p A(p)(1 - D(p))}{\sum_p A(p)}, & G_\mathrm{m} &= \frac{\sum_{p \in N} A(p)(1 - D(p))}{\sum_p A(p)}, \\
G_\mathrm{d} &= \frac{\sum_{p \in N} A(p)(1 - D(p))}{|N|}, & S_\mathrm{n} &= \frac{\sum_{p \in N} 4A(p)(1 - A(p))}{|N|}, \\
S_\alpha &= \frac{\sum_p 4A(p)(1 - A(p))D(p)}{\sum_p D(p)}.
\end{aligned}
\tag{14}
$$

O_α measures outside-support alpha; G_m and G_d measure neck-rear ghost mass and density; S_n and S_α measure translucency. Color heat uses saturation s, alpha boundary B from $P(p) = \mathbf{1}[A(p) > 0.03]$, and weighted mean $\mathrm{WM}(\cdot, \cdot)$:

$$
\begin{aligned}
F_\mathrm{c} = {} & 0.45\ \mathrm{WM}(s, BA) + 0.35\ \mathrm{WM}(s, A(1 - D)) \\
& + 0.20\ \mathrm{WM}(s, 4A(1 - A)D).
\end{aligned}
\tag{15}
$$

For component analysis we additionally report side-contour edge energy $E_\mathrm{c} = \mathrm{WM}(\|\nabla C\|_1, BA)$, where higher values indicate sharper lateral boundaries. Table 1 reports the Marcel mean over -25° and $+25^\circ$ renderings.

Table 1: Marcel side-view artifact diagnostics. Values are averaged over -25° and $+25^\circ$ using a common crop and 9-pixel FLAME-mask dilation. Best and second-best results are highlighted.

Method	$O_\alpha \downarrow$	$G_\mathrm{m} \downarrow$	$G_\mathrm{d} \downarrow$	$S_\mathrm{n} \downarrow$	$S_\alpha \downarrow$	$F_\mathrm{c} \downarrow$
LentiAvatar	0.0691	0.0237	0.0297	0.0304	0.0149	0.2145
RGBAvatar	0.0768	0.0303	0.0384	0.0357	0.0182	0.2117
HRAvatar	0.0889	0.0314	0.0397	0.0177	0.0085	0.2017
FlashAvatar	0.0746	0.0368	0.0435	0.0851	0.0901	0.1685
INSTA	0.0779	0.0227	0.0280	0.0330	0.0270	0.2315

Table 1 compares side-view reconstructions on Marcel. The most direct display-facing signal is outside-support opacity: floating alpha outside the mesh is routed to neighboring viewing zones and can appear as a halo or ghost layer on the lenticular panel. LentiAvatar gives the lowest O_α and remains competitive on both neck-rear ghost measures, which is consistent with the strict head-and-hair supervision domain. INSTA attains slightly lower neck-rear ghost mass and density, and HRAvatar or FlashAvatar are lower on selected translucency or color-fringe scores, but these gains do not uniformly translate to cleaner lateral support in Fig. 4. The result is therefore best read as a display-specific trade-off rather than a generic image-metric leaderboard: LentiAvatar suppresses off-surface side opacity while preserving enough lateral structure for the 32-view routing pipeline. This is the aspect that matters most to stereoscopic video chat, because each conversational frame is expanded into many laterally routed views; artifacts that remain near the neck boundary, collar, or outer silhouette are therefore repeatedly exposed across viewing zones instead of staying hidden in a single frontal portrait.

4.3 Component-wise construction of PMV reconstruction

Figure 5 and Table 2 isolate the contribution of each reconstruction component. The no-PMV baseline lacks lateral evidence and therefore under-constrains side contours at $\pm 25^\circ$. Naive PMV introduces yaw-matched side supervision and already increases side-contour energy from 0.1201 to 0.1234, but it also raises outside-support alpha from 0.0625 to 0.0649 and color-fringe heat from 0.2412 to 0.2474, showing that side views alone are not yet safe supervision. Adding the strict head-and-hair matte yields the cleanest intermediate neck-rear ghost scores by removing lower-neck and collar regions from the pseudo-view target, while frame ranking and yaw-bin selection further regularize which natural turns are allowed to supervise each lateral view. Alignment refinement then reduces residual real-to-virtual mismatch, especially in the lateral boundary region where small offsets would otherwise imprint edge ghosts or color pull.

Figure 4: Qualitative comparison on public benchmark sequences. Rows show representative subjects from the public INSTA benchmark sequences. The leftmost column shows the sequence reference image, and the remaining columns compare LentiAvatar with recent monocular head-avatar baselines rendered at -25°. This lateral probe emphasizes the regime targeted by LentiAvatar: preserving identity, hair, ear, and jaw structure while reducing neck leakage and boundary artifacts that become prominent in multiview autostereoscopic presentation.

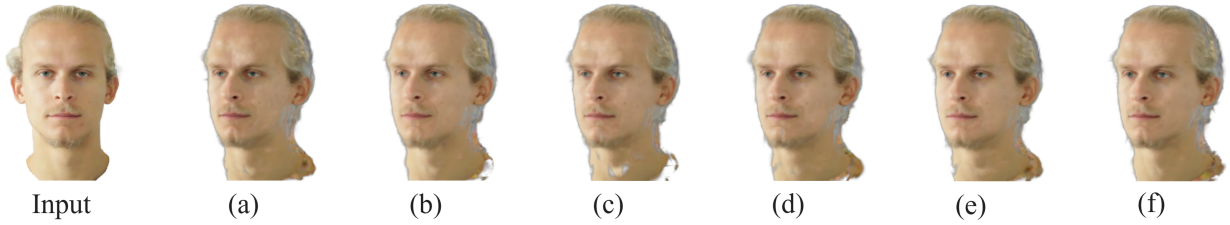

Figure 5: Progressive PMV ablation. The input is followed by lateral-view renderings from (a) no PMV, (b) naive PMV, (c) strict head-and-hair matte supervision, (d) side-frame ranking with yaw-bin selection, (e) alignment and camera refinement, and (f) the final LentiAvatar model with staged stabilization.

The final staged model combines these components with the later no-PMV cleanup schedule to achieve the best overall artifact-detail balance. Relative to No PMV, LentiAvatar reduces outside-support alpha by 2.9%, alpha translucency smear by 3.1%, and color-fringe heat by 4.6%, while increasing side-contour energy by 5.6%. This progression is consistent with the qualitative trend in Fig. 5: the early PMV stages recover missing side structure, and the later stabilization stages retain that structure while removing residual opacity leakage and chromatic boundary artifacts.

Table 2: Component ablation on Marcel at ±25°. Metrics follow Sec. 4.2; all are lower-is-better except E_c.

Variant	$O_\alpha \downarrow$	$G_m \downarrow$	$G_d \downarrow$	$S_n \downarrow$	$S_\alpha \downarrow$	$F_c \downarrow$	$E_c \uparrow$
No PMV	0.0625	0.0170	0.0223	0.0276	0.0187	0.2412	0.1201
Naive PMV	0.0649	0.0167	0.0215	0.0255	0.0239	0.2474	0.1234
+HH matte	0.0631	0.0161	0.0204	0.0364	0.0409	0.2403	0.1242
+Rank/bin	0.0646	0.0172	0.0225	0.0263	0.0202	0.2477	0.1233
+Align	0.0630	0.0186	0.0242	0.0269	0.0195	0.2374	0.1233
LentiAvatar	0.0607	0.0164	0.0214	0.0291	0.0181	0.2300	0.1268

4.4 Autostereoscopic display profiling

For each control frame, the runtime renderer generates 32 avatar views at 960 × 540 and subpixel-composes them into a 3840 × 2160 raster. On an RTX 4090 after initialization, Table 3 compares live metrical tracking with the subject-specific RGB student driver from Sec. 3.4. The comparison uses the same trained avatar, view set, renderer, and subpixel compositor, isolating the cost of estimating the driving controls from the cost of the display pipeline.

Table 3: Runtime profile. Timings are in milliseconds except FPS; stage timings are independently measured counters rather than an additive budget.

Runtime component or metric	Live tracker	Student driver
End-to-end FPS ↑	10.65	38.49
Tracking stage ↓	109.15	4.68
Render 32 views ↓	18.59	16.09
Subpixel composition ↓	1.67	1.67
Display presentation ↓	1.23	0.79
Total loop ↓	109.15	26.15

The profile shows that the live prototype is not bottlenecked by subpixel routing. Composing the 4K lenticular raster takes only 1.67 ms, and rendering all 32 views remains below 19 ms in both settings. Instead, the live metrical tracker dominates the frame time: its 109.15 ms stage cost limits the full system to 10.65 FPS even after initialization. Replacing this tracker with the subject-specific student driver reduces the driving stage to 4.68 ms and raises the same 4K, 32-view display pipeline to 38.49 FPS. For stereoscopic video chat, this difference is practically important because conversational quality depends on how smoothly the avatar controls can be refreshed, not only on how quickly the panel image can be composed. The student-driver setting therefore shows that once the driving cost is reduced, the same avatar renderer and calibrated subpixel encoder can support a much smoother live communication prototype; future system work should primarily reduce or generalize the control-estimation stage.

5 Limitations and Future Work

LentiAvatar relies on informative horizontal head turns in the monocular training sequence. If side frames are sparse, misaligned, or affected by expression changes, hair motion, or poor mattes, PMV supervision weakens. The strict above-neck domain reduces collar and lower-neck contamination, but also limits lower-neck and clothing reconstruction; side-view quality remains below synchronized multi-view capture, especially for challenging hair, ears, and rear-neck regions.

The real-time prototype is also constrained by avatar driving. Live metrical tracking dominates frame time, while the distilled driver improves throughput but still requires subject-specific samples and a separate distillation stage. Future work will reduce tracking and multiview rendering latency, improve perceived realism on the glasses-free panel, jointly tune virtual-camera baselines with panel routing, and extend the head-only setting toward full-body stereoscopic communication.

6 Conclusion

We presented LentiAvatar, a monocular Gaussian head-avatar system for subpixel-routed glasses-free stereoscopic communication. It converts natural head turns into strict above-neck PMV supervision, reducing lateral ghosting and alpha leakage while preserving side detail. At runtime, the trained avatar is encoded into a 4K autostereoscopic raster, and a subject-specific distilled driver supports real-time 32-view display above 30 FPS.

References

ShahRukh Athar, Zexiang Xu, Kalyan Sunkavalli, Eli Shechtman, and Zhixin Shu. 2022. RigNeRF: Fully Controllable Neural 3D Portraits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 20364–20373.

Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 5470–5479. doi:10.1109/CVPR52688.2022.00539

Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 187–194. doi:10.1145/311535.311556

Jianchuan Chen, Jingchuan Hu, Gaige Wang, Zhonghua Jiang, Tiansong Zhou, Zhiwen Chen, and Chengfei Lv. 2025. TaoAvatar: Real-Time Lifelike Full-Body Talking Avatars for Augmented Reality via 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 10723–10734.

Yufan Chen, Lizhen Wang, Qijing Li, Hongjiang Xiao, Shengping Zhang, Hongxun Yao, and Yebin Liu. 2024. MonoGaussianAvatar: Monocular Gaussian Point-Based Head Avatar. In *ACM SIGGRAPH 2024 Conference Papers*. ACM, New York, NY, USA, Article 58, 9 pages. doi:10.1145/3641519.3657499

Yu Deng, Duomin Wang, and Baoyuan Wang. 2024. Portrait4D-v2: Pseudo Multi-View Data Creates Better 4D Head Synthesizer. In *Computer Vision – ECCV 2024*. Springer, Cham, Switzerland, 316–333. doi:10.1007/978-3-031-72643-9_19

Helisa Dhamo, Yinyu Nie, Arthur Moreau, Jifei Song, Richard Shaw, Yiren Zhou, and Eduardo Pérez-Pellitero. 2024. HeadGaS: Real-Time Animatable Head Avatars via 3D Gaussian Splatting. In *Computer Vision – ECCV 2024*. Springer, Cham, Switzerland, 459–476. doi:10.1007/978-3-031-72627-9_26

Neil A. Dodgson. 2005. Autostereoscopic 3D Displays. *Computer* 38, 8 (2005), 31–36. doi:10.1109/MC.2005.252

Hao-Bin Duan, Miao Wang, Jin-Chuan Shi, Xu-Chuan Chen, and Yan-Pei Cao. 2023. BakedAvatar: Baking Neural Fields for Real-Time Head Avatar Synthesis. *ACM Transactions on Graphics* 42, 6, Article 225 (2023), 14 pages. doi:10.1145/3618399

Wei-Qi Feng, Dong Han, Ze-Kang Zhou, Shunkai Li, Xiaoqiang Liu, Pengfei Wan, Di Zhang, and Miao Wang. 2025. GPAvatar: High-fidelity Head Avatars by Learning Efficient Gaussian Projections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 250–259.

Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. 2021. Learning an Animatable Detailed 3D Face Model from In-the-Wild Images. *ACM Transactions on Graphics* 40, 4, Article 88 (2021), 13 pages. doi:10.1145/3450626.3459936

Guy Gafni, Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2021. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 8649–8658.

Simon Giebenhain, Tobias Kirschstein, Martin Rünz, Lourdes Agapito, and Matthias Nießner. 2024. NPGA: Neural Parametric Gaussian Avatars. In *SIGGRAPH Asia*

2024 Conference Papers. ACM, New York, NY, USA, Article 127, 11 pages. doi:10.1145/3680528.3687689

Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. 1996. The Lumigraph. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, NY, USA, 43–54. doi:10.1145/237170.237200

Philip Grassal, Malte Prinzler, Titus Leistner, Carsten Rother, Matthias Nießner, and Justus Thies. 2022. Neural Head Avatars from Monocular RGB Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 18653–18664.

Andrew Jones, Ian McDowall, Hideshi Yamada, Mark Bolas, and Paul Debevec. 2007. Rendering for an Interactive 360 Degree Light Field Display. *ACM Transactions on Graphics* 26, 3, Article 40 (2007), 10 pages. doi:10.1145/1276377.1276427

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4, Article 139 (2023), 14 pages.

Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Nießner, Patrick Pérez, Christian Richardt, Michael Zollhöfer, and Christian Theobalt. 2018. Deep Video Portraits. *ACM Transactions on Graphics* 37, 4, Article 163 (2018), 14 pages. doi:10.1145/3197517.3201283

Tobias Kirschstein, Shenhan Qian, Simon Giebenhain, Tim Walter, and Matthias Niessner. 2023. NeRSemble: Multi-view Radiance Field Reconstruction of Human Heads. *ACM Transactions on Graphics* 42, 4, Article 161 (2023), 14 pages.

Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. 2024. HuGS: Human Gaussian Splats. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 505–515.

Robert Kooima, Andrew Prudhomme, Jürgen Schulze, Daniel Sandin, and Thomas DeFanti. 2010. A Multi-Viewer Tiled Autostereoscopic Virtual Reality Display. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*. ACM, New York, NY, USA, 171–174. doi:10.1145/1889863.1889899

Douglas Lanman, Matthew Hirsch, Yunhee Kim, and Ramesh Raskar. 2010. Content-Adaptive Parallax Barriers: Optimizing Dual-Layer 3D Displays Using Low-Rank Light Field Factorization. *ACM Transactions on Graphics* 29, 6, Article 163 (2010), 10 pages. doi:10.1145/1882261.1866164

Seok Lee, Juyoung Park, Jingu Heo, Byongmin Kang, Dongwoo Kang, Hyoseok Hwang, Jinho Lee, Yoonsun Choi, Kyuhwan Choi, and Dongkyung Nam. 2018. Autostereoscopic 3D Display Using Directional Subpixel Rendering. *Optics Express* 26, 16 (2018), 20233–20247. doi:10.1364/OE.26.020233

Marc Levoy and Pat Hanrahan. 1996. Light Field Rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, NY, USA, 31–42. doi:10.1145/237170.237199

Linzhou Li, Yumeng Li, Yanlin Weng, Youyi Zheng, and Kun Zhou. 2025a. RGBAvatar: Reduced Gaussian Blendshapes for Online Modeling of Head Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 10747–10757.

Peng Li, Wangguandong Zheng, Yuan Liu, Tao Yu, Yangguang Li, Xingqun Qi, Xiaowei Chi, Siyu Xia, Yan-Pei Cao, Wei Xue, Wenhan Luo, and Yike Guo. 2025b. PSHuman: Photorealistic Single-image 3D Human Reconstruction using Cross-Scale Multiview Diffusion and Explicit Remeshing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 16008–16018.

Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Transactions on Graphics* 36, 6, Article 194 (2017), 17 pages. doi:10.1145/3130800.3130813

Stephen Lombardi, Jason Saragih, Tomas Simon, and Yaser Sheikh. 2018. Deep Appearance Models for Face Rendering. *ACM Transactions on Graphics* 37, 4, Article 68 (2018), 13 pages. doi:10.1145/3197517.3201401

Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural Volumes: Learning Dynamic Renderable Volumes from Images. *ACM Transactions on Graphics* 38, 4, Article 65 (2019), 14 pages. doi:10.1145/3306346.3323020

Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. 2021. Pixel Codec Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 64–73.

Shengjie Ma, Yanlin Weng, Tianjia Shao, and Kun Zhou. 2024. 3D Gaussian Blendshapes for Head Avatar Animation. In *ACM SIGGRAPH 2024 Conference Papers*. ACM, New York, NY, USA, 1–10. doi:10.1145/3641519.3657462

Avishai M. Marson and Adrian Stern. 2015. Horizontal Resolution Enhancement of Autostereoscopy Three-Dimensional Displayed Image by Chroma Subpixel Downsampling. *Journal of Display Technology* 11, 10 (2015), 800–806. doi:10.1109/JDT.2014.2382712

Wojciech Matusik and Hanspeter Pfister. 2004. 3D TV: A Scalable System for Real-Time Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes. *ACM Transactions on Graphics* 23, 3 (2004), 814–824. doi:10.1145/1015706.1015805

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision – ECCV 2020*. Springer, Cham, Switzerland, 405–421.

SeungJun Moon, Hah Min Lew, Seungeun Lee, Ji-Su Kang, and Gyeong-Moon Park. 2025. GeoAvatar: Adaptive Geometrical Gaussian Splatting for 3D Head Avatar. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Piscataway, NJ, USA, 12811–12821.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics* 41, 4, Article 102 (2022), 15 pages. doi:10.1145/3528223.3530127

Sergio Orts-Escolano, Christoph Rhemann, Sean Ryan Fanello, Wayne Chang, Adarsh Kowdle, Yury Degtyarev, David Kim, Philip L. Davidson, Sameh Khamis, Mingsong Dou, Vladimir Tankovich, Charles T. Loop, Qin Cai, Philip A. Chou, Sarah Mennicken, Julien P. C. Valentin, Vivek Pradeep, Shenlong Wang, Sing Bing Kang, Pushmeet Kohli, Yuliya Lutchyn, Cem Keskin, and Shahram Izadi. 2016. Holoportation: Virtual 3D Teleportation in Real-Time. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 741–754. doi:10.1145/2984511.2984517

Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. 2021a. Nerfies: Deformable Neural Radiance Fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Piscataway, NJ, USA, 5865–5874.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. 2021b. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. *ACM Transactions on Graphics* 40, 6, Article 238 (2021), 12 pages.

Renjing Pei, Zheng Geng, and ZhaoXing Zhang. 2016. Subpixel Multiplexing Method for 3D Lenticular Display. *Journal of Display Technology* 12, 10 (2016), 1197–1204. doi:10.1109/JDT.2016.2576471

Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Niessner. 2024a. GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 20299–20309.

Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, and Siyu Tang. 2024b. 3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 5020–5030.

Lingteng Qiu, Shenhao Zhu, Qi Zuo, Xiaodong Gu, Yuan Dong, Junfei Zhang, Chao Xu, Zhe Li, Weihao Yuan, Liefeng Bo, Guanying Chen, and Zilong Dong. 2025. AniGS: Animatable Gaussian Avatar from a Single Image with Inconsistent Gaussian Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 21148–21158.

Daniel Ruijters. 2008. Dynamic Resolution in GPU-Accelerated Volume Rendering to Autostereoscopic Multiview Lenticular Displays. *EURASIP Journal on Advances in Signal Processing* 2009, Article 843753 (2008), 8 pages. doi:10.1155/2009/843753

Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 130–141.

Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, and Zeyu Wang. 2024. SplattingAvatar: Realistic Real-Time Human Avatars with Mesh-Embedded Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 1606–1616.

Jiapeng Tang, Davide Davoli, Tobias Kirschstein, Liam Schoneveld, and Matthias Niessner. 2025. GAF: Gaussian Avatar Reconstruction from Monocular Videos via Multi-view Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 5546–5558.

Justus Thies, Michael Zollhöfer, and Matthias Nießner. 2019. Deferred Neural Rendering: Image Synthesis Using Neural Textures. *ACM Transactions on Graphics* 38, 4, Article 66 (2019), 12 pages. doi:10.1145/3306346.3323035

Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. 2016. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 2387–2395. doi:10.1109/CVPR.2016.262

Justus Thies, Michael Zollhöfer, Christian Theobalt, Marc Stamminger, and Matthias Nießner. 2018. HeadOn: Real-Time Reenactment of Human Portrait Videos. *ACM Transactions on Graphics* 37, 4, Article 164 (2018), 13 pages. doi:10.1145/3197517.3201350

Cong Wang, Di Kang, Heyi Sun, Shenhan Qian, Zixuan Wang, Linchao Bao, and Song-Hai Zhang. 2025a. MeGA: Hybrid Mesh-Gaussian Head Avatar for High-Fidelity Rendering and Head Editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 26274–26284.

Jie Wang, Jiucheng Xie, Xianyan Li, Feng Xu, Chi-Man Pun, and Hao Gao. 2025d. GaussianHead: High-Fidelity Head Avatars With Learnable Gaussian Derivation. *IEEE Transactions on Visualization and Computer Graphics* 31, 7 (2025), 4141–4154. doi:10.1109/TVCG.2025.3561794

Shaofei Wang, Tomas Simon, Igor Santesteban, Timur Bagautdinov, Junxuan Li, Vasu Agrawal, Fabian Prada, Shoou-I Yu, Pace Nalbone, Matt Gramlich, Roman Lubachersky, Chenglei Wu, Javier Romero, Jason Saragih, Michael Zollhoefer, Andreas Geiger, Siyu Tang, and Shunsuke Saito. 2025b. Relightable Full-Body Gaussian

Codec Avatars. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers* (Vancouver, BC, Canada) *(SIGGRAPH Conference Papers '25)*. ACM, New York, NY, USA, 12 pages. doi:10.1145/3721238.3730739

Yating Wang, Xuan Wang, Ran Yi, Yanbo Fan, Jichen Hu, Jingcheng Zhu, and Lizhuang Ma. 2025c. 3D Gaussian Head Avatars with Expressive Dynamic Appearances by Compact Tensorial Representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 21117–21126.

Gordon Wetzstein, Douglas Lanman, Matthew Hirsch, and Ramesh Raskar. 2012. Tensor Displays: Compressive Light Field Synthesis Using Multilayer Displays with Directional Backlighting. *ACM Transactions on Graphics* 31, 4, Article 80 (2012), 11 pages. doi:10.1145/2185520.2185576

Jun Xiang, Xuan Gao, Yudong Guo, and Juyong Zhang. 2024. FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 1802–1812.

Yuelang Xu, Benwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, and Yebin Liu. 2024. Gaussian Head Avatar: Ultra High-Fidelity Head Avatar via Dynamic Gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 1931–1941.

Zongyuan Yang, Baolin Liu, Yingde Song, Lan Yi, Yongping Xiong, Zhaohe Zhang, and Xunbo Yu. 2024. DirectL: Efficient Radiance Fields Rendering for 3D Light Field Displays. *ACM Transactions on Graphics* 43, 6, Article 234 (2024), 19 pages. doi:10.1145/3687897

Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Kangjie Chen, Minghan Qin, Yu Li, and Haoqian Wang. 2025a. HRAvatar: High-Quality and Relightable Gaussian Head Avatar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 26285–26296.

Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Yang Li, Minghan Qin, Yu Li, and Haoqian Wang. 2025b. GUAVA: Generalizable Upper Body 3D Gaussian Avatar. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. IEEE, Piscataway, NJ, USA, 14205–14217.

Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C. Bühler, Xu Chen, Michael J. Black, and Otmar Hilliges. 2022. I M Avatar: Implicit Morphable Head Avatars from Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 13545–13555.

Yufeng Zheng, Yifan Wang, Gordon Wetzstein, Michael J. Black, and Otmar Hilliges. 2023. PointAvatar: Deformable Point-Based Head Avatars from Videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 21057–21067.

Yiyu Zhuang, Jiaxi Lv, Hao Wen, Qing Shuai, Ailing Zeng, Hao Zhu, Shifeng Chen, Yujiu Yang, Xun Cao, and Wei Liu. 2025. IDOL: Instant Photorealistic 3D Human Creation from a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 26308–26319.

Wojciech Zielonka, Timo Bolkart, and Justus Thies. 2023. Instant Volumetric Head Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 4574–4584.

Matthias Zwicker, Wojciech Matusik, Frédo Durand, and Hanspeter Pfister. 2006. Antialiasing for Automultiscopic 3D Displays. In *Proceedings of the Eurographics Symposium on Rendering*. The Eurographics Association, Aire-la-Ville, Switzerland, 73–82. doi:10.2312/EGWR/EGSR06/073-082

Matthias Zwicker, Hanspeter Pfister, Jeroen van Baar, and Markus Gross. 2001. Surface Splatting. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, New York, NY, USA, 371–378. doi:10.1145/383259.383300