High-fidelity Multi-view Hair Reconstruction via Spherical Orientation Distribution Functions

ZHONGSI TANG, State Key Lab of CAD&CG, Zhejiang University, China
JIAHAO GENG, Faceunity Inc., China
XINYUAN LIU, State Key Lab of CAD&CG, Zhejiang University, China
ZHONG REN, State Key Lab of CAD&CG, Zhejiang University, China
YOUYI ZHENG, State Key Lab of CAD&CG, Zhejiang University, China
YANLIN WENG, State Key Lab of CAD&CG, Zhejiang University, China
KUN ZHOU*, State Key Lab of CAD&CG, Zhejiang University, China



Fig. 1. High-fidelity strand-level hair reconstruction from casual handheld captures. Our method infers explicit, scalp-connected 3D strands from monocular orbit videos. Each pair illustrates a reference frame alongside the reconstructed hairstyle. By leveraging a probabilistic volumetric field to represent local orientation distributions, our framework overcomes the structural ambiguities in traditional unimodal representations. The approach robustly reconstructs complex, subject-specific structures—including layered waves and densely tangled curls—without compromising high-frequency visible details.

High-fidelity multi-view 3D hair reconstruction from casual handheld captures remains a formidable challenge due to severe self-occlusions and the intricate topology of individual strands. A critical bottleneck in existing 3D hair representations—both volumetric grids and continuous implicit fields—is the reliance on deterministic spatial queries that assign a single local orientation. This unimodal assumption inevitably collapses and averages crossing or layered strands into erroneous directions, leading to a loss of internal structural details. To overcome this, we introduce SODF-Hair, a novel framework that represents local hair structures using Spherical Orientation Distribution Functions (SODFs). By encoding a multi-modal distribution within each voxel, our SODF representation effectively resolves the structural ambiguities of tangled and overlapping fibers, preserving complex local variations even under limited grid resolutions. Our pipeline explicitly decomposes the reconstruction into exterior observation and interior completion. To bridge 2D image evidence into 3D, we first reconstruct a visible hair-segment shell using fixed-scale oriented 3D Gaussian proxies. Their elongated covariances yield projections that match the streak-like texture of visible hair in the input images, enabling the recovery of fine visible segment geometry while preventing the proxies from degenerating into free-form blobs. We then formulate the recovery of invisible internal structures as an inpainting task. A geometry-conditioned probabilistic model, leveraging a compact latent SODF codec and rectified flow matching, predicts the missing multi-modal distributions within the closed volume. Finally, rather than

*Corresponding author: Kun Zhou.

Authors' Contact Information: Zhongsi Tang, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, zhongsitang@zju.edu.cn; Jiahao Geng, Faceunity Inc., Hangzhou, China, jiahaogeng@outlook.com; Xinyuan Liu, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, liuxinyuanxy@gmail.com; Zhong Ren, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, renzhong@zju.edu.cn; Youyi Zheng, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, youyizheng@zju.edu.cn; Yanlin Weng, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, weng@cad.zju.edu.cn; Kun Zhou, State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China, kunzhou@acm.org.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 1557-7368/2026/12-ART192
https://doi.org/10.1145/3842588

prematurely committing to a single streamline, we assemble explicit scalp-connected strands via a field-guided beam search and constrained graph optimization. Extensive evaluations suggest that our approach faithfully preserves subject-specific visible details while generating structurally coherent internal strand geometry, demonstrating robust performance across diverse hairstyles.

CCS Concepts: • **Computing methodologies** → **Reconstruction**; **Shape modeling**; *Neural networks*; *Computer vision.*

Additional Key Words and Phrases: 3D reconstruction, orientation field, 3D Gaussians, differentiable rendering

ACM Reference Format:
Zhongsi Tang, Jiahao Geng, Xinyuan Liu, Zhong Ren, Youyi Zheng, Yanlin Weng, and Kun Zhou. 2026. High-fidelity Multi-view Hair Reconstruction via Spherical Orientation Distribution Functions. *ACM Trans. Graph.* 45, 6, Article 192 (December 2026), 16 pages. https://doi.org/10.1145/3842588

1 Introduction

Hair is a crucial carrier of personal identity, and its realistic 3D modeling is pivotal for high-fidelity digital humans in films, games, and interactive avatars. However, reconstructing strand-level hairstyles from casual multi-view captures remains a formidable challenge. The inherent geometric complexity of real hair—characterized by dense, layered, and tangled fibers—coupled with severe self-occlusion, renders the recovery of complete interior structures fundamentally under-constrained. Because optical evidence alone is insufficient to resolve these occluded regions, the reconstruction relies heavily on how the internal hair geometry is mathematically parameterized. Consequently, the choice of 3D representation plays a decisive role in determining whether complex local topologies can be faithfully preserved or are inevitably smoothed out.

Existing volumetric and implicit methods predominantly commit to a deterministic, single orientation per spatial query, a unimodal assumption that inherently struggles with complex fiber geometries. Explicit voxel grids provide direct spatial access but are fundamentally limited by memory and computation; under practical resolutions, multiple strand layers inevitably pass through the same voxel. Forcing a single dominant orientation in such regions collapses crossing or tangled strands into an erroneous average direction [Saito et al. 2018; Zhou et al. 2018]. Conversely, continuous neural implicit fields theoretically offer infinite resolution, avoiding the voxel grid limitation [Kuang et al. 2022; Sklyarova et al. 2023]. However, practical strand extraction still requires dense discrete sampling. In current implicit formulations, the queried spatial output is typically reduced to occupancy paired with a single local orientation. This restriction prematurely discards local orientation uncertainty, failing to capture the multi-modal nature of intersecting fibers. Both paradigms suffer from making a premature hard commitment, erasing high-frequency structural variations and misleading subsequent strand extraction.

To address these representational limitations, we propose SODF-Hair, a novel framework built upon Spherical Orientation Distribution Functions (SODFs) to reconstruct high-fidelity, strand-level hairstyles. Rather than forcing a single geometric direction at the representation level, our SODF parameterizes a probabilistic, multi-modal distribution of orientations within each volumetric cell. By uniformly discretizing the spherical domain to preserve sharp directional modes, this representation accommodates complex local topologies—such as intersecting or layered strands—without demanding prohibitively dense spatial grids. It shifts the formulation from deterministic vectors to probabilistic distributions, effectively delaying discrete connectivity decisions until a global geometric context can be evaluated.

Building upon the SODF representation, we formulate the comprehensive reconstruction of the hairstyle as a geometry-conditioned volumetric inpainting task. Our pipeline explicitly decouples the preservation of visible evidence from the inference of occluded interiors. First, we reconstruct visible hair geometry using fixed-scale oriented 3D Gaussian segment proxies. The elongated covariances of these primitives naturally align with the streak-like texture of visible hair, establishing a high-fidelity subject-specific completion domain while filtering out image noise.

Second, rather than directly optimizing invisible regions from 2D images, we introduce SODF-Flow, a latent SODF inpainting network driven by rectified conditional flow matching. To make the high-dimensional spherical distribution tractable for generative modeling, we compress the SODF into a compact yet expressive latent space using a trained auto-encoder. Conditioned on the reconstructed Gaussian segment shell and a scalp prior, SODF-Flow predicts the missing multi-modal orientation distributions inside the closed volume.

Finally, recognizing that a multi-modal probabilistic field does not define a unique streamline, we replace conventional deterministic ODE-based tracing with a multi-hypothesis strand assembly mechanism. By integrating field-guided beam search with constrained graph-based topology selection, our approach thoroughly explores the preserved local directional ambiguities. This ensures the generation of globally compatible, explicit scalp-connected strands while preserving the observed high-frequency exterior details.

Extensive evaluations show that our approach outperforms state-of-the-art methods, faithfully reconstructing intricate layered strand patterns and compact volumetric structures for diverse hairstyles.

In summary, our main contributions are:

- We propose Spherical Orientation Distribution Functions (SODFs), a novel volumetric representation that captures multi-modal spherical hair orientation distributions, effectively resolving the structural ambiguities and information loss inherent in deterministic unimodal representations.
- We introduce fixed-scale oriented 3D Gaussian segment proxies to accurately reconstruct visible hair boundaries from rom casually captured images, providing robust geometric constraints aligned with the streak-like texture of hair.
- We formulate interior hair generation as an inpainting problem conditioned on geometry, leveraging SODF-Flow via rectified flow matching on a compact latent SODF space to predict plausible interior multi-modal structures.
- We present a field-guided multi-hypothesis strand assembly algorithm, combining beam search and graph optimization to convert probabilistic orientation fields into coherent, scalp-connected hair strands.

2 Related Work

2.1 Hair Reconstruction and Inference

Reconstructing full 3D hairstyles requires balancing the fidelity of visible exterior observations with the structural plausibility of occluded interiors. Early observation-driven methods recover hair volumes or visible geometry, primarily using multi-view stereo and image-space orientation cues [Jakob et al. 2009; Luo et al. 2012, 2013a,b; Maeda et al. 2023; Nam et al. 2019; Paris et al. 2004, 2008; Rosu et al. 2022; Shen et al. 2023; Wei et al. 2005]. Subsequent advancements leverage differentiable rendering with neural or explicit primitives. For instance, GroomCap [Zhou et al. 2024] optimizes a NeRF-based hair volume under dense calibrated capture, while Dr.Hair [Takimoto et al. 2024] differentiably renders line segments. These purely observation-driven approaches excel at preserving high-frequency visible details. However, because occluded interior regions are fundamentally under-constrained by optical evidence, unregularized optimization often diffuses or hallucinates internal structures, leading to incomplete or structurally unnatural strands.

To address the severe ambiguity in unobserved regions, researchers have long relied on data-driven statistics. Early single-view methods pioneered this direction by utilizing databases or manually designed priors to retrieve, deform, and synthesize plausible hairstyles from highly limited observations [Wang et al. 2009; Chai et al. 2015, 2016, 2012; Hu et al. 2015]. Later learning-based methods, such as NeuralHDHair [Wu et al. 2022] and HairStep [Zheng et al. 2023], replace database retrieval with neural intermediate representations to infer dense 3D hair geometry from limited image evidence. For more complex image-based settings, recent optimization frameworks regularize strand fitting with learned strand or hairstyle priors, as exemplified by Neural Haircut [Sklyarova et al. 2023]. Other image-based methods, including GaussianHair [Luo et al. 2024] and Gaussian Haircut [Zakharov et al. 2024], adopt strand-based Gaussian primitives for differentiable hair modeling and reconstruction. Recent single-image methods such as Im2Haircut [Sklyarova et al. 2025] and HairOrbit [Jin et al. 2026] additionally leverage diffusion-based priors or generated multi-view cues. While such learned priors and generative cues significantly improve the global plausibility of unseen regions, they may also bias the reconstruction toward the learned dataset or generative manifold, smoothing out subject-specific fine image evidence such as flyaways and complex curls.

Recognizing the tension between preserving visible evidence and inferring hidden structures, MonoHair [Wu et al. 2024] proposes a decoupled strategy: it first reconstructs a detailed outer line map and subsequently infers the inner structure using a feed-forward network [Kuang et al. 2022]. Our pipeline shares this decoupled philosophy, utilizing fixed-scale Gaussian segments to establish a robust completion domain. However, we fundamentally depart from existing methods in how the internal geometry is mathematically parameterized and generated, shifting from deterministic fields to a geometry-conditioned generative inpainting framework.

2.2 Representations for Hair Geometry

The mathematical representation of hair geometry dictates the upper bound of reconstruction quality. Existing paradigms can be broadly categorized into global strand latents, explicit volumetric grids, and continuous implicit fields, most of which inherently commit to a deterministic geometric assumption.

Global Strand Representations. To make strand-level hair modeling tractable, prior work develops hierarchical, parametric, and procedural representations [Chang et al. 2025; Chen et al. 2024; He et al. 2025], while generative methods encode individual strands or complete hairstyles into compact latent spaces using VAEs [Rosu et al. 2025; Sklyarova et al. 2023, 2024; Zhou et al. 2023]. Recent strand generative models further extend this direction with stronger conditional or autoregressive priors [Long et al. 2025; Luo et al. 2026]. While these global strand representations provide strong macro-structure constraints for synthesis, they are typically organized by 2D scalp maps or sparse guide strands. Consequently, they struggle to locally align with dense, layered, and mixed multi-view image observations, often sacrificing local geometric accuracy for global coherence.

Explicit Volumetric Representations. Representing hair with occupancy and orientation values on a 3D explicit grid [Saito et al. 2018; Zhang et al. 2018; Zhou et al. 2018] provides direct spatial access, facilitating local completion and strand extraction. However, explicit volumes are inherently bottlenecked by memory and computational constraints. More critically, most existing grids mandate a deterministic formulation, storing only a single dominant orientation per voxel. In complex hairstyles where multiple strand layers intersect or tangle, this unimodal assumption inevitably averages divergent fiber populations into a single invalid intermediate vector, fundamentally failing to model local structural ambiguities.

The need to represent multiple local fiber directions also arises in diffusion MRI (dMRI) tractography, where local fiber orientations are recovered from measurements whose sampling directions and counts vary across devices and protocols, typically from tens to hundreds. To obtain a unified orientation representation from such sparse angular samples and to model crossing fibers, Q-ball [Tuch 2004] imaging and constrained spherical deconvolution (CSD) [Jeurissen et al. 2014; Tournier et al. 2007] commonly use spherical harmonics (SH) for great-circle integration in q-space and response-function deconvolution. Recent SH-based methods, such as Elaldi et al. [2024], learn neural response functions for deconvolution; even high-angular-resolution tractography settings remain on the order of hundreds of samples, e.g., 4 shells with 90 gradients each (360 measurements) in DiSCo [Rafael-Patino et al. 2021], with SH typically truncated to low orders such as $l_{\max} = 10$.

In contrast, layered hair induces sharper angular peaks than low-order SH can preserve. Recent graphics work also organizes direction space using spherical Voronoi partitions or icosahedral spatial–directional encodings [Di Sario et al. 2026; Weier et al. 2026]. We discretize antipodal directions on an icosphere and compress the resulting distribution with a topologically-aware auto-encoder, preserving sharp multi-modal directions while enabling simple and efficient queries.

Continuous Implicit Fields. Neural implicit fields encode occupancy and orientations as continuous functions, theoretically bypassing the resolution limits of explicit grids [Jin et al. 2026; Kuang et al. 2022; Mildenhall et al. 2020; Sklyarova et al. 2023; Wu et al. 2022;

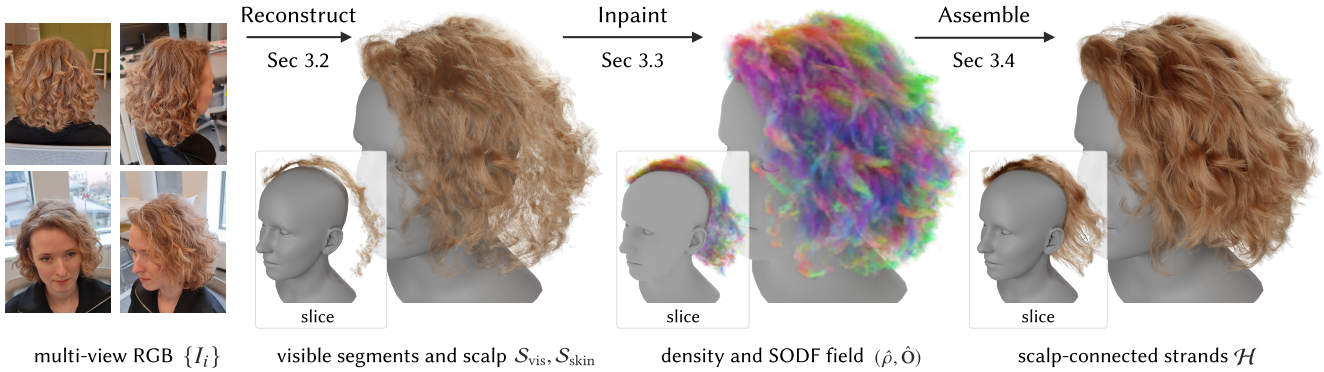


Fig. 2. Overview of our method. Given multi-view RGB images $\{I_i\}_{i=1}^N$, we first reconstruct a visible segment shell $\mathcal{S}_{\text{vis}}$ and fit the scalp surface $\mathcal{S}_{\text{skin}}$ to define a subject-specific completion domain and provide geometric conditions for field prediction. A geometry-conditioned probabilistic hair field predicts density and multi-modal antipodal orientation distributions $(\hat{\rho}, \hat{\mathbf{O}})$ inside this domain. Because the completed field provides probabilistic orientation distributions rather than explicit strand connectivity, we extract visible chained segments, grow SODF-guided candidate paths by beam search, and assemble them into explicit scalp-connected strands $\mathcal{H}$ using constrained graph optimization.

Zhou et al. 2024]. They have become the standard for sparse-view orientation prediction and multi-view optimization. Nevertheless, extracting explicit strands from implicit fields still requires dense discrete sampling. In current formulations, the spatial query output is commonly reduced to a deterministic local orientation. This forces an early geometric commitment during evaluation, prematurely discarding local orientation uncertainty and making it exceedingly difficult to expose the full multi-modal nature of intersecting fibers to the strand-growing stage.

In contrast, our approach introduces SODF as an explicit volumetric representation. Instead of assigning a single deterministic vector, each voxel retains an antipodal spherical orientation distribution. This design explicitly models local multi-modal orientation mixtures, effectively accommodating the structural ambiguities of layered strands. By integrating this probabilistic representation with a multi-hypothesis beam-search assembly, we successfully preserve complex local topologies that are otherwise lost in unimodal deterministic representations.

3 Method

3.1 Overview

Given casually captured multi-view RGB images $\{I_i\}_{i=1}^N$ of a target hairstyle from a handheld monocular device, our objective is to reconstruct explicit, scalp-connected 3D hair strands that are both visually consistent with the input observations and structurally plausible in occluded regions. To address the inherent ambiguity of internal hair geometry, SODF-Hair follows a decoupled philosophy: we deliberately separate the preservation of reliable optical evidence from the generative inference of occluded structures, utilizing Spherical Orientation Distribution Functions (SODFs) as the core volumetric representation. As illustrated in Fig. 2, the pipeline consists of three sequential stages.

Hair Shell Reconstruction (Sec. 3.2). To bridge 2D image observations with the 3D completion domain, we first reconstruct a dense, subject-specific visible segment shell $\mathcal{S}_{\text{vis}}$. Rather than directly extracting strands from noisy multi-view orientations, we parameterize the visible region using fixed-scale oriented 3D Gaussian segment proxies. These proxies explicitly capture high-frequency visible details while filtering image noise. By fusing these optimized segments with a fitted bust $\mathcal{B}$, we establish a watertight spatial domain $\mathcal{E}$ that provides rigorous boundary conditions for downstream volumetric completion.

Latent SODF Inpainting (Sec. 3.3). Within the closed volume $\mathcal{E}$, we formulate the recovery of occluded internal structures as a geometry-conditioned inpainting task. To accommodate the multi-modal nature of layered and tangled fibers, we represent local hair geometry probabilistically using SODFs. To make this high-dimensional spherical representation tractable, we compress it into a compact latent space via an auto-encoder. Conditioned on the reconstructed segment shell and scalp normals, our DiT-based generative module, SODF-Flow, employs rectified conditional flow matching to predict the missing multi-modal density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$ throughout the interior volume.

Field-Guided Strand Assembly (Sec. 3.4). The predicted density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$ provides rich multi-modal directional likelihoods but does not define unique streamlines. To convert this probabilistic field into explicit scalp-connected strands without forcing a premature geometric commitment, we replace conventional deterministic tracing with a multi-hypothesis assembly mechanism. By integrating field-guided beam search for path proposal with constrained graph optimization for topology selection, this stage thoroughly explores local directional ambiguities and ensures global structural compatibility, yielding the final high-fidelity 3D hairstyle.

3.2 Hair Shell Reconstruction

This stage distills the multi-view observations into two geometric outputs: a visible segment shell $\mathcal{S}_{\text{vis}}$ that retains observed strand

evidence, and a closed hair-completion volume $\mathcal{E}$ that specifies where unobserved structure may be inferred.

Visible-Segment Initialization. For each calibrated view I_i, we first extract a hair mask M_i using SAM2 [Ravi et al. 2025] and compute an unsigned 2D orientation map Θ_i with log-Gabor filters. Following the standard multi-view preprocessing of Wu et al. [2024], we fuse these per-view cues into a coarse exterior line map, represented as an oriented point cloud whose samples carry 3D position, local tangent direction, and color. However, this initial geometry often suffers from noise and topological holes due to severe self-occlusion and weak texture. To robustly refine this initial geometry, we parameterize the visible hair region using fixed-scale 3D Gaussian segment proxies. For each segment proxy g_k, let $\mathbf{t}_k$ denote its local hair-axis direction and let $R(\mathbf{t}_k)$ be the rotation that aligns the anisotropic covariance with this axis. Unlike general 3D Gaussian splatting designed for unconstrained volumetric radiance fields, we explicitly enforce a strongly anisotropic covariance [Luo et al. 2024; Zakharov et al. 2024; Zhou et al. 2024] for each segment g_k:

$$\Sigma_k = R(\mathbf{t}_k)\,\mathrm{diag}(\sigma_{\parallel}^2, \sigma_{\perp}^2, \sigma_{\perp}^2)R(\mathbf{t}_k)^{\top}, \qquad \sigma_{\parallel} \gg \sigma_{\perp}. \tag{1}$$

This parameterization makes each Gaussian a short, oriented segment that follows the streak-like structure of visible hair in the images, while preventing it from degenerating into a free-form blob.

Bust-Aware Segment Optimization. To refine these proxies from image evidence, the image-space losses should supervise only hair, while the head geometry is still needed to account for occlusion. We therefore keep the two representations separate and model the bust as a mesh $\mathcal{B}$, fitted by optimizing an SMPL-X [Pavlakos et al. 2019] model under multi-view mask constraints. Its projected body and face contours are aligned with their corresponding semantic masks, while the head region is constrained to lie inside the overall foreground mask in each view, preventing the fitted scalp from protruding beyond the observed hair silhouette. During hair-shell optimization, $\mathcal{B}$ remains fixed. For each calibrated view, we render a depth map of $\mathcal{B}$ and use it to exclude hair proxies whose projected depths lie behind the bust; only the remaining view-visible proxies contribute to the color, silhouette, and orientation losses. Thus, $\mathcal{B}$ accounts for head occlusion without contributing hair appearance or direction.

We optimize these segment proxies through differentiable rendering, driven by a joint objective:

$$\mathcal{L}_{\mathrm{G}} = \lambda_{\mathrm{rgb}}\mathcal{L}_{\mathrm{rgb}} + \lambda_{\alpha}\mathcal{L}_{\alpha} + \lambda_{\mathrm{o}}\mathcal{L}_{\mathrm{o}}. \tag{2}$$

Here, $\mathcal{L}_{\mathrm{rgb}}$ combines L_1 and SSIM penalties for photometric fidelity, $\mathcal{L}_{\alpha}$ is an L_1 loss between the rendered opacity and M_i, and $\mathcal{L}_{\mathrm{o}}$ is a sign-invariant cosine loss between the projected segment direction and Θ_i. To avoid invalid direction averaging, an issue also noted by GroomCap [Zhou et al. 2024], we render the orientation using the tangent of the foremost visible proxy rather than alpha-blending directions. In addition to standard 3DGS cloning and pruning, we periodically remove proxies that remain occluded across all views and receive no direct image supervision, keeping the representation concentrated as a thin, surface-aligned shell.

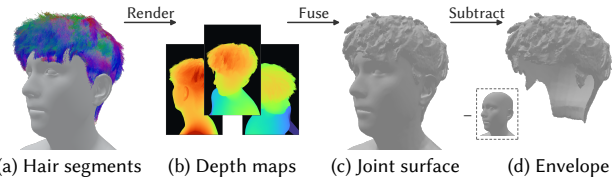


Fig. 3. Hair-envelope construction. The visible segment shell and fitted bust $\mathcal{B}$ are jointly depth-rendered, fused into a closed mesh, and separated by Boolean subtraction to obtain the completion volume $\mathcal{E}$.

Upon convergence, the optimized segments constitute a dense visible hair-segment shell $\mathcal{S}_{\mathrm{vis}}$ that faithfully captures local image-consistent strand details.

Hair-Envelope Construction. The visible segment shell captures observed exterior geometry but does not by itself define an interior volume. To obtain the spatial domain required for volumetric completion, we construct a closed hair envelope jointly with the fitted bust $\mathcal{B}$.

As shown in Fig. 3, we rasterize $\mathcal{S}_{\mathrm{vis}}$ as oriented line segments jointly with $\mathcal{B}$ in each calibrated view to obtain multi-view hair–bust depth maps. TSDF fusion and mesh closure recover a watertight joint mesh, from which $\mathcal{B}$ is removed by Boolean subtraction. The enclosed hair volume is

$$\mathcal{E} = \Omega(\mathcal{S}_{\mathrm{vis}} \cup \mathcal{B}) \setminus \Omega(\mathcal{B}), \tag{3}$$

Here, $\Omega(\mathcal{S}_{\mathrm{vis}} \cup \mathcal{B})$ denotes the joint volume reconstructed by depth fusion, and $\Omega(\mathcal{B})$ denotes the bust volume. We further extract the interface between $\mathcal{E}$ and the bust as the scalp surface $\mathcal{S}_{\mathrm{skin}}$. Then we sample scalp roots $\mathcal{R}$ on $\mathcal{S}_{\mathrm{skin}}$ using Poisson-disk sampling. Thus, $\mathcal{E}$ defines the completion domain, $\mathcal{S}_{\mathrm{skin}}$ provides scalp-normal conditions, and $\mathcal{R}$ supplies root locations for downstream strand assembly.

3.3 Latent SODF Inpainting

After obtaining the closed volume bounded by the visible hair-segment shell and the scalp, we complete the SODFs in the invisible interior. We formulate this step as a geometry-conditioned volumetric inpainting problem: the SODFs converted from visible hair segments and the scalp normals serve as conditions, while the missing SODFs inside $\mathcal{E}$ are generated by the model. This is a conditional generation task because only sparse orientation evidence is available near the observed hair, yet the output must infer a full field of local orientation distributions throughout the interior. Since internal strands are strongly correlated with exterior visible strands but are not a simple smooth extrapolation, the completed SODFs should remain consistent with the exterior hairstyle while preserving sharp local orientation modes.

To this end, we introduce SODF-Flow, a DiT-based generative model that performs SODF completion as inpainting conditioned on visible hair segments and the scalp prior. The visible hair segments provide local orientation constraints observed from images, while the scalp prior provides a geometric prior for plausible strand

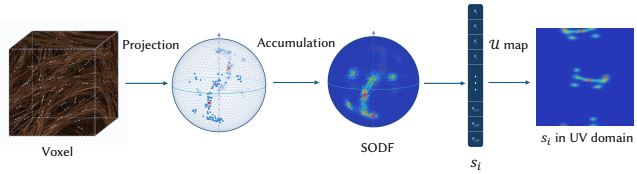


Fig. 4. Voxel-wise SODF representation. From left to right, the figure shows a local voxel containing multiple hair orientations; the assignment of these orientations onto a level-4 icosphere, where each antipodal face pair defines one unsigned orientation bin and the red cross marks the antipodal equivalent of the red assignment; the normalized antipodal orientation distribution visualized as a heatmap on the icosphere; the corresponding 2560-D SODF vector $\mathbf{s}_i$ stored for the voxel; and its UV unfolding $\mathcal{U}(\mathbf{s}_i)$ used as the auto-encoder's internal convolutional layout.

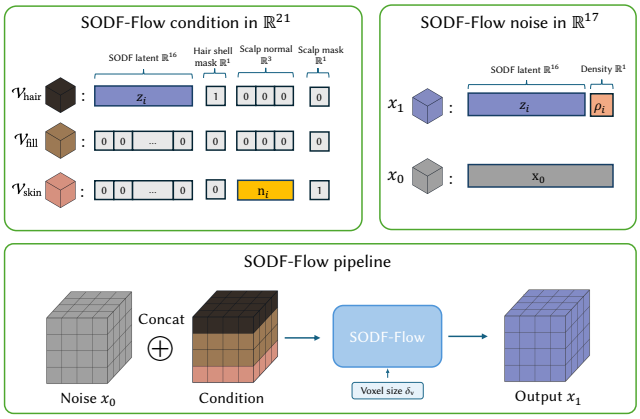


Fig. 5. SODF-Flow. For each voxel, the input feature concatenates a 17-dimensional noisy generation variable and a 21-dimensional condition vector. The upper-left inset shows the condition encoding for $\mathcal{V}_{\text{fill}}$, $\mathcal{V}_{\text{hair}}$, and $\mathcal{V}_{\text{skin}}$: fill voxels use an all-zero condition, visible hair-segment shell voxels provide $\mathbf{z}_i^{\text{vis}}$ and the hair-shell mask, and scalp voxels provide $\mathbf{n}_i$ and the scalp mask. The upper-right inset shows that both the noisy variable and the output are 17-dimensional, with 16 dimensions for the SODF latent code and one for density. The physical voxel size δ_v conditions the network separately.

growth. SODF-Flow is responsible for generating SODF distributions inside the closed volume that are structurally consistent with these sparse conditions.

SODF Representation. We represent local hair orientations with an antipodal spherical distribution. The unsigned spherical domain is discretized by the faces of a level-4 icosphere. Compared with latitude-longitude grids, this near-uniform discretization avoids pole and seam singularities, and its antipodal face pairs can be precomputed directly. By folding each antipodal face pair into one axis bin, we obtain $K = 2560$ orientation bins. Let B_k denote the antipodal face pair of bin k, and let $\mathbf{a}_k$ be the corresponding representative axis. Fig. 4 illustrates this discretization and the subsequent voxel-wise accumulation of segment orientations.

For a voxel v_i, each intersecting hair segment contributes its in-voxel length to the antipodal bin of its direction. Let $m_{i,k}$ denote the accumulated length in bin k. The normalized SODF is

$$s_{i,k} = \frac{m_{i,k}}{\sum_{j=1}^{K} m_{i,j}}, \qquad k = 1, \ldots, K. \tag{4}$$

We additionally store a scalar hair density ρ_i, obtained by converting the total in-voxel strand length into an estimated hair-volume fraction. Unlike a deterministic orientation field, $\mathbf{s}_i$ can retain multiple local orientation modes in crossing, layered, or tangled regions. For compact notation, we denote the collection of voxel-wise SODFs by the field $\mathbf{O}$, where $\mathbf{O}(v_i) = \mathbf{s}_i$.

Latent SODF Auto-Encoder. Each raw SODF is a high-dimensional distribution over spherical axes. Directly training SODF-Flow to generate this distribution is computationally expensive and difficult to optimize. We therefore train a SODF auto-encoder that compresses each SODF into a compact latent code while preserving sharp and multi-modal orientation patterns. The SODF itself is defined on the K icosphere bins, which also serve as the domain for supervision. Since this discretization is irregular for standard CNNs, we use a UV chart only as an internal convolutional layout for the auto-encoder, as shown in Fig. 4: the SODF is mapped to the chart, encoded and decoded by CNNs, and then projected back to the same icosphere bins:

$$\mathbf{z}_i = E_\phi(\mathcal{U}(\mathbf{s}_i)), \tag{5}$$

$$\hat{\mathbf{s}}_i = \Pi\big(D_\psi(\mathbf{z}_i)\big), \qquad \mathbf{z}_i \in \mathbb{R}^{16}. \tag{6}$$

Here, $\mathcal{U}(\cdot)$ maps the axis distribution to the UV-domain chart, E_ϕ and D_ψ denote the CNN encoder and decoder, and $\Pi(\cdot)$ maps the decoded chart-space prediction back to the same K icosphere bins and normalizes it into a valid SODF. The latent code $\mathbf{z}_i$ is then used as the compact SODF feature for SODF-Flow.

Two implementation details are used to make the chart-based auto-encoder compatible with the structure of SODFs. First, because hair orientations are unsigned, antipodal spherical bins are tied when mapping between the icosphere discretization and the UV chart. Second, since the horizontal boundary of the UV chart is a continuous longitude seam, we use cyclic padding along this dimension rather than zero padding. The auto-encoder is trained with a distribution-aware objective:

$$\mathcal{L}_{\text{AE}} = \lambda_{\text{sqrt}}\mathcal{L}_{\text{sqrt}} + \lambda_{\text{js}}\mathcal{L}_{\text{js}} + \lambda_{\text{ten}}\mathcal{L}_{\text{ten}} + \lambda_{\text{lat}}\mathcal{L}_{\text{lat}}. \tag{7}$$

All reconstruction terms are evaluated after projecting the decoded output back to the original K-axis spherical discretization. The term $\mathcal{L}_{\text{sqrt}} = \|\sqrt{\hat{\mathbf{O}}} - \sqrt{\mathbf{O}}\|_2^2$ measures reconstruction error in square-root probability space, and the Jensen–Shannon divergence $\mathcal{L}_{\text{js}}$ further aligns the predicted and target probability distributions. To preserve the dominant axial statistics, $\mathcal{L}_{\text{ten}}$ matches their probability-weighted second-order orientation tensors. Finally, $\mathcal{L}_{\text{lat}}$ regularizes the latent code to keep the compact representation well conditioned.

SODF-Flow. With the compact latent SODF representation, we formulate interior completion as conditional inpainting in the latent space, as illustrated in Fig. 5. We discretize the adaptive bounding

box of $\mathcal{E}$ into an $r \times r \times r$ grid. Because $\mathcal{E}$ is subject-specific, the physical voxel size δ_v varies across hairstyles and is supplied to SODF-Flow as a scale condition. For each active voxel, including voxels in the closed hair-completion volume and scalp voxels, we construct a voxel feature of dimension 38. If the active set contains N voxels, this gives a voxel feature matrix in $\mathbb{R}^{N\times 38}$ before patchification.

Each voxel feature is the concatenation of a noisy generation state and a condition vector, as illustrated in Fig. 5:

$$\mathbf{h}_i(t) = [\mathbf{x}_i(t), \mathbf{c}_i] \in \mathbb{R}^{38}, \qquad \mathbf{x}_i(t) \in \mathbb{R}^{17}, \quad \mathbf{c}_i \in \mathbb{R}^{21}. \tag{8}$$

Collecting all active voxels gives the voxel feature matrix

$$\mathbf{H}_t = [\mathbf{h}_1(t), \ldots, \mathbf{h}_N(t)]^\top \in \mathbb{R}^{N\times 38}. \tag{9}$$

The first part, $\mathbf{x}_i(t)$, is the noisy variable to be denoised by rectified flow. Its clean target contains a 16-dimensional latent SODF code and a one-dimensional density value:

$$\mathbf{y}_i = [\mathbf{z}_i,\ \rho_i] \in \mathbb{R}^{17}. \tag{10}$$

The second part, $\mathbf{c}_i$, is a 21-dimensional condition vector composed of a 16-dimensional latent SODF of the visible hair-segment shell, a one-dimensional hair-shell mask, a three-dimensional scalp normal, and a one-dimensional scalp mask. According to the voxel type, the condition is assigned as

$$\mathbf{c}_i = \begin{cases} \mathbf{0}_{21}, & v_i \in \mathcal{V}_{\text{fill}}, \\ \left[\mathbf{z}_i^{\text{vis}}, 1, \mathbf{0}_3, 0\right], & v_i \in \mathcal{V}_{\text{hair}}, \\ \left[\mathbf{0}_{16}, 0, \mathbf{n}_i, 1\right], & v_i \in \mathcal{V}_{\text{skin}}, \end{cases} \tag{11}$$

where $\mathcal{V}_{\text{fill}}$ denotes voxels to be completed, $\mathcal{V}_{\text{hair}}$ denotes visible hair-segment shell voxels, and $\mathcal{V}_{\text{skin}}$ denotes scalp voxels. Therefore, voxels in the unknown interior have all-zero conditions; visible hair-segment shell voxels provide only the corresponding latent SODF and set the hair-shell mask to one; scalp voxels provide only the scalp normal and set the scalp mask to one. This unified voxel feature format allows the model to jointly observe the unknown region, the visible hair segments, and the scalp voxels.

SODF-Flow uses a DiT backbone on sparse voxel features defined over the active hair volume. We apply a 3D patchify operation, where each $2 \times 2 \times 2$ voxel block is grouped and linearly projected into one DiT token:

$$\mathbf{u}_m(t) = P_{\text{emb}}\left(\text{vec}\{\mathbf{h}_i(t) \mid v_i \in \mathcal{P}_m\}\right), \tag{12}$$

where $\mathcal{P}_m$ is the m-th voxel patch and P_{emb} denotes the patch embedding layer.

In our full-resolution setting, we use $r = 128$. The resulting active sparse volume contains roughly 3×10^5 voxels per sample, making full self-attention over all tokens computationally prohibitive. To make full-resolution attention tractable, we adopt Spatial Sparse Attention (SSA) [Wu et al. 2025] in the DiT blocks. For a token $\mathbf{u}_m^l$ at layer l, SSA updates it using sparse key-value tokens selected from the 3D token layout:

$$\mathbf{u}_m^{l+1} = \text{SSA}\left(\mathbf{u}_m^l, \{\mathbf{u}_n^l, \mathbf{p}_n\}_{n=1}^M\right), \tag{13}$$

where $\mathbf{p}_n$ denotes the 3D coordinate of token $\mathbf{u}_n^l$. SSA constructs sparse interactions from the 3D token layout, combining spatial compression, blockwise selection, and local window attention. This enables spatially coherent reasoning over the full active volume while avoiding the quadratic cost of global attention. The output tokens are then unpatchified back to voxel-level 17-dimensional predictions. The first 16 dimensions give the completed latent SODF code $\hat{\mathbf{z}}_i$, and the last dimension gives the density $\hat{\rho}_i$. The decoder maps $\hat{\mathbf{z}}_i$ back to a completed per-voxel SODF $\hat{\mathbf{s}}_i$, yielding the density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$ for the following maximum-likelihood strand growth stage.

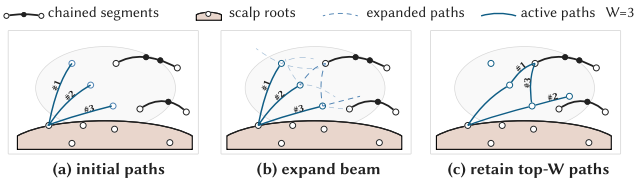


Fig. 6. **Field-guided beam search for candidate path proposal.** Starting from a terminal, beam search expands the current active paths along multiple directions, scores the expanded hypotheses by $\phi(P)$, and retains the top-W paths as the next beam. This expand–score–retain process is repeated until paths reach compatible endpoints or leave the SODF field. Those paths form the candidate set $\mathcal{Q}$.

3.4 Field-Guided Strand Assembly

The completed density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$ provides dense multi-modal directional likelihoods, but it lacks root-to-tip polarity, explicit connectivity, and topological uniqueness. Converting this probabilistic field into explicit strands requires resolving both local structural ambiguity and global geometric compatibility. To avoid the premature hard commitment inherent in conventional ODE-based single-line tracing, we formulate strand extraction as a global maximum-likelihood assembly problem.

Our framework uses three inputs: the fixed visible chained segments C obtained by merging visible segments from $\mathcal{S}_{\text{vis}}$, scalp roots $\mathcal{R}$ sampled on $\mathcal{S}_{\text{skin}}$, and the completed density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$. Each chained segment is treated as fixed visible geometry with two endpoints, and these endpoints are used as terminals together with the scalp roots. Let $\mathcal{P}$ denote the space of all possible generated paths, and let $\mathfrak{I}$ denote the family of topologically valid path sets that satisfy spatial compatibility, scalp connectivity, and endpoint constraints. The objective is to select an optimal forest $\mathcal{F}^*$ that maximizes the joint likelihood under the predicted field:

$$\mathcal{F}^* = \arg\max_{\mathcal{F}\subseteq\mathcal{P},\ \mathcal{F}\in\mathfrak{I}} \prod_{P\in\mathcal{F}} p(P \mid \hat{\rho}, \hat{\mathbf{O}}). \tag{14}$$

For a discrete path $P = \{(\mathbf{x}_t, \mathbf{d}_t)\}_{t=1}^{|P|}$, its individual likelihood is determined by the SODF directional support, the density confidence, and a geometric curvature penalty:

$$p(P \mid \hat{\rho}, \hat{\mathbf{O}}) \propto \left(\prod_{t=1}^{|P|} \hat{O}(\mathbf{x}_t, \mathbf{d}_t)\right)^{1/|P|} \Psi_\rho(P; \hat{\rho})\Psi_\kappa(P). \tag{15}$$

Algorithm 1 Field-guided multi-hypothesis strand assembly.

Require: Completed density and SODF field $(\hat{\rho}, \hat{\mathbf{O}})$, visible hair-segment shell $\mathcal{S}_{\text{vis}}$, scalp roots $\mathcal{R}$, beam width W, max steps $T_{\max}$
Ensure: Scalp-connected strands $\mathcal{H}$

Step 1: Chain construction.
Build fixed chained segments C from $\mathcal{S}_{\text{vis}}$.
Terminal set $\mathcal{T} \leftarrow \mathcal{R} \cup \text{Endpoints}(C)$.
Step 2: Field-guided candidate path proposal.
Global candidate set $\mathcal{Q} \leftarrow \emptyset$.
for $e \in \mathcal{T}$ **do**
  Active paths $\mathcal{A}_e \leftarrow$ initial paths from e.
  Terminated paths $\mathcal{A}_{\text{term}} \leftarrow \emptyset$.
  for $t = 1, \ldots, T_{\max}$ **do**
    $\widetilde{\mathcal{A}}_e \leftarrow$ expand $\mathcal{A}_e$ along possible directions.
    Score $P \in \widetilde{\mathcal{A}}_e$ by $\phi(P)$.
    $\widetilde{\mathcal{A}}_e \leftarrow$ top $W - |\mathcal{A}_{\text{term}}|$ paths in $\widetilde{\mathcal{A}}_e$.
    Move paths that hit another endpoint $e' \in \mathcal{T}$ or leave the SODF field as free-tip paths from $\widetilde{\mathcal{A}}_e$ to $\mathcal{A}_{\text{term}}$.
    $\mathcal{A}_e \leftarrow$ remaining paths in $\widetilde{\mathcal{A}}_e$.
    if $|\mathcal{A}_{\text{term}}| \geq W$ **then**
      break
    end if
  end for
  $\mathcal{Q} \leftarrow \mathcal{Q} \cup \mathcal{A}_{\text{term}}$.
end for
Step 3: Constrained topology selection.
Sort $\mathcal{Q}$ in descending order of $\phi(P)$.
$\mathcal{F} \leftarrow \emptyset$.
for $P \in \mathcal{Q}$ **do**
  if $\mathcal{F} \cup \{P\} \in \mathfrak{I}$ **then**
    $\mathcal{F} \leftarrow \mathcal{F} \cup \{P\}$.
  end if
end for
Step 4: Strand concatenation and refinement.
Assemble $C \cup \mathcal{F}$ into continuous strands.
Apply endpoint attachment and spatial refinement.
return $\mathcal{H}$

Here, $\hat{O}(\mathbf{x}_t, \mathbf{d}_t)$ is the directional probability queried from the local SODF along tangent $\mathbf{d}_t$. The factor $1/|P|$ length-normalizes the accumulated directional support. This normalization prevents variable-length paths from being biased merely by their sample counts, ensuring that short and long strands are compared primarily by their average probabilistic support. Ψ_ρ and Ψ_κ represent the path-wise density confidence and curvature regularizer, respectively.

Multi-Hypothesis Path Proposal. Directly optimizing the objective over the complete path space $\mathcal{P}$ is intractable. Furthermore, greedy tracing that strictly follows the local maximum would discard the preserved multi-modal orientation information, failing to recover layered or tangled structures. Therefore, we utilize a field-guided beam search, illustrated in Fig. 6, to propose a compact, high-likelihood candidate subset $\mathcal{Q} \subset \mathcal{P}$.

For computational stability, we rank candidates in the log-domain:

$$\phi(P) = \frac{1}{|P|} \sum_{t=1}^{|P|} \log\left(\hat{O}(\mathbf{x}_t, \mathbf{d}_t) + \epsilon_O\right) + \log \Psi_\rho(P; \hat{\rho}) + \log \Psi_\kappa(P). \tag{16}$$

Starting from a set of terminals $\mathcal{T} = \mathcal{R} \cup \text{Endpoints}(C)$ (comprising scalp roots and chained-segment endpoints), the beam search maintains a state $q = (\mathbf{x}, \boldsymbol{\tau}, e)$, where $\mathbf{x}$ is the current endpoint, $\boldsymbol{\tau}$ is the current path tangent, and $e \in \mathcal{T}$ is the source terminal. At each step, we query the local field $(\hat{\rho}(\mathbf{x}), \hat{\mathbf{O}}(\mathbf{x}))$ and orient the unsigned axes $\mathbf{a}_k$ using the current tangent $\boldsymbol{\tau}$. The beam expands along several different directions, retaining the top-scoring partial paths. Beam search terminates a path when it reaches another compatible terminal or exits the valid SODF domain. The latter is retained as a free-tip candidate, whose exit point is interpreted as a strand tip rather than an additional terminal; this avoids forcing an unsupported terminal-to-terminal connection. These terminated paths form the finite candidate set:

$$\mathcal{Q} = \bigcup_{e \in \mathcal{T}} \text{BeamSearch}(e; \hat{\rho}, \hat{\mathbf{O}}). \tag{17}$$

Constrained Topology Selection. While beam search provides high-likelihood path fragments, not all candidates in $\mathcal{Q}$ can coexist. A valid hairstyle must satisfy the strict topological constraints $\mathfrak{I}$: no repeated endpoint usage, no direction conflicts within fixed chained segments, no cyclic structures, and no spatial collisions.

Because exact subset optimization with such coupled constraints is NP-hard, we deploy a score-ordered feasible insertion algorithm. Starting with an empty selected set $\mathcal{F} = \emptyset$, we iterate through candidates $P \in \mathcal{Q}$ in descending order of their log-likelihood $\phi(P)$. A candidate is greedily inserted only if its addition maintains global feasibility ($\mathcal{F} \cup \{P\} \in \mathfrak{I}$). This guarantees a globally consistent, collision-free assembly that approximates the maximum-likelihood objective.

Strand Concatenation and Refinement. Finally, we traverse the selected forest $\mathcal{F}$, concatenating the accepted paths with the fixed chained segments C into explicit strands. For any remaining short, disconnected components, we attach them to nearby scalp-connected strands if strongly supported by the SODF field. After removing overloaded roots and resampling, the pipeline outputs the final high-fidelity explicit strand asset $\mathcal{H}$.

The complete procedure is summarized in Algorithm 1.

4 Experiments

To rigorously validate our decoupled reconstruction pipeline and probabilistic field formulation, our evaluation is structured around three core dimensions. First, we compare SODF-Hair against state-of-the-art image-based hair reconstruction methods on established public benchmarks and synthetic datasets. Second, we conduct comprehensive ablations isolating the three pillars of our pipeline: visible hair shell reconstruction, multi-modal probabilistic field prediction, and field-guided strand assembly. Within the field analysis, we specifically validate the superiority of the SODF representation over deterministic and other compact parameterizations.



Fig. 7. Qualitative comparisons on public video captures from the Neural Haircut dataset [Sklyarova et al. 2023]. Each row shows a captured subject from two representative viewpoints, with the reference image followed by reconstructions from MonoHair [Wu et al. 2024], Gaussian Haircut [Zakharov et al. 2024], GroomCap [Zhou et al. 2024], and our method. Compared with prior methods, our results more faithfully reproduce hairstyle-dependent geometry, including curl clusters, layered strand flow, compact short-hair shapes, and overall hair volume across different viewpoints.

4.1 Experimental Setup and Metrics

Datasets and Baselines. Our evaluation utilizes three distinct data regimens to thoroughly assess geometric fidelity and real-world robustness. (1) *Public Video Benchmark:* For direct comparison with prior work, we utilize multi-view captures from the Neural Haircut dataset [Sklyarova et al. 2023]. We compare against strong baselines, including MonoHair [Wu et al. 2024], Gaussian Haircut [Zakharov et al. 2024], and GroomCap [Zhou et al. 2024]. (2) *Synthetic Benchmark:* For a fair quantitative geometric evaluation, we use held-out strand-based MetaHuman Groom assets [Epic Games 2026], whose asset source is entirely different from our training data.

We render calibrated input views and image-space cues along handheld orbit trajectories. The field-prediction generative model is strictly isolated from this test set. (3) *Real In-the-Wild Captures:* We evaluate on 15 custom handheld orbit videos featuring diverse hair topologies (short, medium, long, straight, curly) to assess robustness under unconstrained casual capture conditions.

Macro-Geometric Metrics. Following standard strand-reconstruction protocols, we uniformly sample reconstructed and ground-truth strands by arc length to obtain oriented point sets $\mathcal{P}_{\text{rec}} = \{(\mathbf{x}, \mathbf{t})\}$ and $\mathcal{P}_{\text{gt}} = \{(\mathbf{y}, \mathbf{s})\}$, where tangents are axial (unsigned). We report Accuracy (Acc.) and Completeness (Comp.) distances in millimeters:

$$d_{\text{acc}} = \frac{1}{|\mathcal{P}_{\text{rec}}|} \sum_{(\mathbf{x},\mathbf{t}) \in \mathcal{P}_{\text{rec}}} \min_{(\mathbf{y},\mathbf{s}) \in \mathcal{P}_{\text{gt}}} \|\mathbf{x} - \mathbf{y}\|_2, \tag{18}$$

$$d_{\text{comp}} = \frac{1}{|\mathcal{P}_{\text{gt}}|} \sum_{(\mathbf{y},\mathbf{s}) \in \mathcal{P}_{\text{gt}}} \min_{(\mathbf{x},\mathbf{t}) \in \mathcal{P}_{\text{rec}}} \|\mathbf{x} - \mathbf{y}\|_2, \tag{19}$$

alongside the symmetric Chamfer distance (Cham.) $d_{\text{cham}} = \frac{1}{2}(d_{\text{acc}} + d_{\text{comp}})$.

We also compute strand-level Precision (Prec.), Recall (Rec.), and F-score. A sampled point is considered matched only if both its position and orientation are within the chosen tolerances: $\|\mathbf{x}-\mathbf{y}\|_2 < \tau_d$ and $\arccos(|\mathbf{t}^\top \mathbf{s}|) < \tau_\alpha$. We report the main Precision (Prec.) and Recall (Rec.) at (2mm/20°) and include F-scores denoted as $F_{2/20}$, $F_{3/30}$, and $F_{4/40}$ at (2mm/20°), (3mm/30°), and (4mm/40°), respectively.

Probabilistic Field Fidelity Metrics. A core premise of our approach is the preservation of multi-modal directional ambiguity. Therefore, evaluating the field requires moving beyond deterministic vector

metrics. We voxelize ground-truth strands to extract the true orientation distribution $\mathbf{s}_i^{\mathrm{gt}}$. Given a predicted orientation distribution $\hat{\mathbf{s}}_i$, its unnormalized second moment tensor is:

$$Q(\mathbf{s}_i) = \sum_{k=1}^{K} s_{i,k}\mathbf{a}_k\mathbf{a}_k^{\top}. \tag{20}$$

While we report the standard deterministic *Orientation Error* (using the dominant eigenvector of $Q(\mathbf{s}_i)$), this metric intrinsically collapses multi-modal information. To rigorously evaluate the preservation of overlapping structures without penalizing multi-modality, we introduce the *Moment Error* (Moment Err.):

$$e_Q = \left\| Q(\hat{\mathbf{s}}_i) - Q(\mathbf{s}_i^{\mathrm{gt}}) \right\|_F. \tag{21}$$

For orientation-parameterization and latent-code ablations, we further measure distribution similarity on the exact antipodal icosphere using Spherical Jensen-Shannon Divergence (Sph. JSD), L_1 probability error (L1 Err.), Moment Error (Moment Err.), and Mode Error (Mode Err., angular error between dominant modes). We use the condition number of the latent whitening transform (Lat. Cond.) to reflect how well-conditioned the learned latent space is before SODF-Flow training.

For field-representation analysis, we additionally evaluate Density IoU (Dens. IoU), voxel Chamfer (Vox. Cham.), Axis Error (Axis Err.), and SODF Jensen-Shannon Divergence (SODF JSD). Dens. IoU and Vox. Cham. measure occupied support, Axis Err. compares dominant axial directions on the overlapping support, and SODF JSD compares orientation distributions after converting deterministic directions to soft one-hot distributions on the antipodal icosphere.

4.2 Implementation Details

Both learned components are trained once on hairstyles from the public DiffLocks dataset [Rosu et al. 2025], whose length, curliness, and density parameters are randomized.

Hair Shell Reconstruction. Each subject is initialized with 5.61 million Gaussian segment proxies on average and retains 1.92 million after optimization. We fix $\sigma_\perp = 5{\times}10^{-5}$ m and $\sigma_\parallel = 10^{-3}$ m, while optimizing position, rotation, opacity, and color. We use $\lambda_{\mathrm{rgb}} = \lambda_\alpha = \lambda_{\mathrm{o}} = 1$, with L_1 and SSIM weighted by 0.8 and 0.2 within $\mathcal{L}_{\mathrm{rgb}}$. The cross-view visibility pruning described in Sec. 3.2 is applied every 3,000 iterations.

SODF Auto-Encoder. We use 1,000 hairstyles and sample 4,000 voxels per hairstyle, balancing simple, high-entropy, and high-complexity distributions. Its four-block encoder and decoder are optimized with AdamW at a learning rate of 10^{-3}. We set $(\lambda_{\mathrm{sqrt}}, \lambda_{\mathrm{js}}, \lambda_{\mathrm{ten}}, \lambda_{\mathrm{lat}}) = (16, 0.25, 0.05, 10^{-4})$.

SODF-Flow. We use 22,388 DiffLocks hairstyles for training and reserve 2,523 for testing. SODF-Flow uses a DiT backbone with 20 transformer blocks operating on sparse voxel features. We train it with AdamW for 12K steps at $r = 64$ with a learning rate of 10^{-4}, then fine-tune it for 15K steps at $r = 128$ with a learning rate of $5{\times}10^{-5}$. At $r = 128$, δ_{v} spans 1.62–4.86 mm for synthetic training data and 1.74–4.68 mm for real captures.

Field-Guided Strand Assembly. We sample 100,000 scalp roots using Poisson-disk sampling and set W to 3 for root terminals and 4 for visible-chain terminals. The maximum number of tracing steps is $T_{\max} = 224$, with a step length of 0.75 voxel, and final strands are resampled at 3 mm intervals. Across 22 reconstructions, selected free-tip paths comprise roughly 5% of strands for short hairstyles and 12% for medium-to-long hairstyles; they occur mainly in occluded interiors and have limited visible impact.

Training Cost. The learned components are trained once. Training the SODF auto-encoder for 1.12M steps takes 5.5 h on one RTX 4090. On eight RTX 5090 GPUs, the 12K-step $r = 64$ stage of SODF-Flow takes approximately one day, followed by approximately two days for 15K-step fine-tuning at $r = 128$.

Runtime Cost. We benchmark reconstruction on six cases from the public Neural Haircut video dataset, running our method and the released implementations of MonoHair and Gaussian Haircut on the same workstation with one RTX 4090 and an i7-13700KF CPU. Our method takes 54.2 min per subject on average. Hair Shell Reconstruction accounts for 46.6 min: 13.9 min for oriented point-cloud initialization, 8.5 min for SMPL-X fitting, and 24.2 min for segment optimization and envelope construction. Latent SODF Inpainting and Field-Guided Strand Assembly take 1.7 min and 5.9 min, respectively. The directly measured runtimes are 4.7 h for MonoHair and 4.2 h for Gaussian Haircut. GroomCap is not publicly released and is therefore excluded from this benchmark; its paper reports approximately 33.5 h per subject [Zhou et al. 2024].

4.3 Comparison with Prior Work

Public Video Captures. Fig. 7 provides qualitative comparisons on the Neural Haircut benchmark. We specifically focus on regions where the preservation of subject-specific multi-modal geometry is critical, such as flyaways, local strand crossings, and dense curl clusters. While strong baselines like MonoHair and Gaussian Haircut capture general global shapes, they frequently smooth out intersecting strands or hallucinate overly uniform flows due to their deterministic orientation or diffusion priors. In contrast, by decoupling visible evidence preservation from probabilistic internal inference, our method faithfully reproduces complex layered strand flows and compact structural volumes across all viewpoints.

To further examine how each method completes self-occluded regions, Fig. 8 compares planar cutaway renderings under the same cutting plane for the examples corresponding to Fig. 7. The exposed interiors reveal different reconstruction behaviors that are less apparent from exterior views alone: the prior-free GroomCap result preserves visible surface evidence but often produces vertically aligned bundle artifacts in hidden regions, whereas methods with stronger hairstyle priors yield smoother interiors with attenuated local variation. Our SODF completion maintains interior flows that follow the visible curls and layers while preserving finer strand-level structure.

Held-Out Synthetic Data. We further conduct qualitative and quantitative analyses on held-out synthetic data. GroomCap [Zhou et al. 2024] is not included in these synthetic comparisons because its implementation is not publicly available. Fig. 9 visualizes one

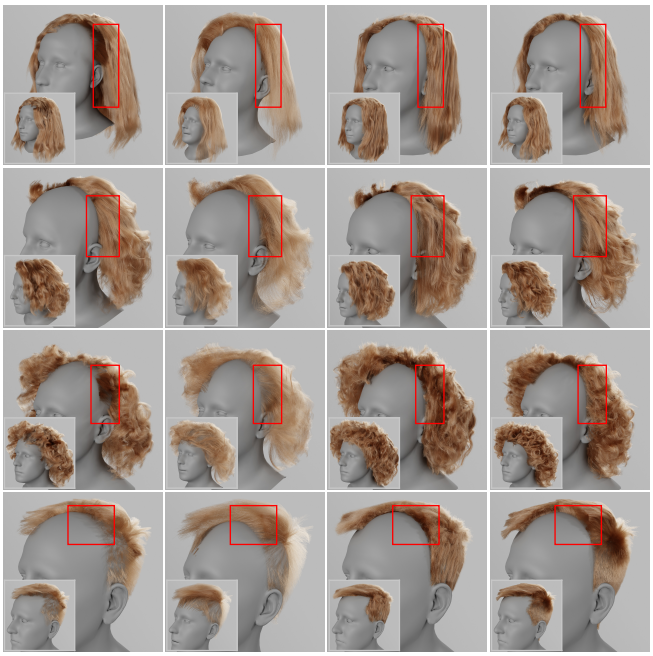


Fig. 8. Cutaway comparison on public video captures. Each column uses the same wedge-shaped cutaway oriented toward the viewing direction for one compared method or our result, with a small inset showing the corresponding uncut hairstyle. The exposed regions highlight how different representations complete internal hair geometry: GroomCap tends to form vertically aligned hidden bundles without a learned interior prior, while methods with stronger hairstyle priors produce smoother but less detailed interiors. Our method recovers internal structures that remain consistent with the exterior hairstyle and retain local strand variation.

held-out synthetic hairstyle together with its corresponding planar cutaway results. The comparison shows that our reconstruction preserves rich layered details on the visible surface while forming a coherent interior structure after cutting the hairstyle open. In contrast, competing methods either lose high-frequency exterior variation or expose sparse and less organized hidden strands. In addition, Table 1 shows that our method achieves the best performance across all hairstyle accuracy metrics on the held-out synthetic test set, further demonstrating its effectiveness for high-fidelity hair reconstruction.

Real Handheld Captures. Fig. 10 demonstrates our system's robustness on 15 unconstrained handheld orbit videos encompassing highly diverse hair topologies. Across extreme variations in length, curliness, and observation quality, our method consistently reconstructs explicit, scalp-connected 3D hair assets that maintain strict multi-view geometric consistency rather than degrading into view-dependent approximations.

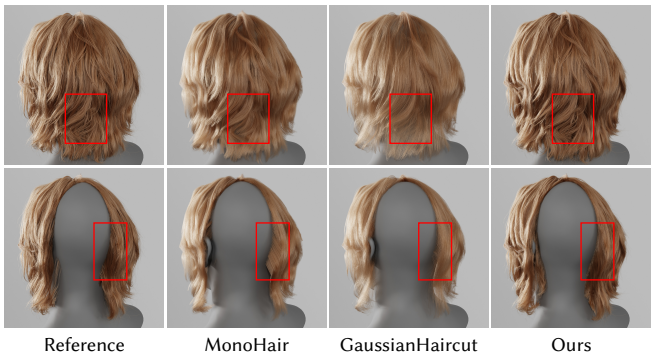


Fig. 9. Validation on held-out synthetic data. The first row shows the uncut hairstyle and the second row exposes the internal structure using the same cutaway rendering protocol. With the synthetic reference as ground truth, our result is closer to the reference both externally and internally: it retains layered surface details while producing a plausible dense interior that follows the hairstyle structure.

Table 1. Quantitative comparison of MonoHair [Wu et al. 2024], GaussianHaircut [Zakharov et al. 2024], and our SODF-Hair on held-out synthetic hairstyles. Distances (accuracy, completeness, and Chamfer) are reported in millimeters. Precision and recall use the 2mm/20° position-orientation threshold; F-scores under looser thresholds are also reported. Here, bold numbers denote the best performance and underlined numbers denote the second-best.

Method	Acc. ↓	Comp. ↓	Cham. ↓	Prec. ↑	Rec. ↑	$F_{2/20}$ ↑	$F_{3/30}$ ↑	$F_{4/40}$ ↑
MonoHair	<u>0.822</u>	1.266	<u>1.044</u>	<u>0.744</u>	0.597	0.662	0.803	0.898
GaussianHaircut	1.367	<u>0.813</u>	1.090	0.644	<u>0.817</u>	<u>0.720</u>	<u>0.836</u>	<u>0.899</u>
Ours	**0.639**	**0.495**	**0.567**	**0.856**	**0.820**	**0.838**	**0.935**	**0.972**

4.4 Analysis of Hair Shell Reconstruction

It is crucial to note that this stage is designed strictly to establish a watertight geometric boundary condition for downstream completion, rather than functioning as a standalone rendering technique. Fig. 11 isolates the architectural necessity of our optimization and anatomical constraints. The *w/o Optimization* baseline, which relies purely on voxel centers and MVS-derived directions, provides only noisy coarse exterior support and fails to form a continuous envelope near high-frequency regions like the hair part. While segment optimization improves local flow, removing the anatomical depth constraint (*w/o Body*) allows segments to erroneously leak into or behind the scalp, creating topological holes. By integrating fixed-scale segment optimization with rigorous body-aware visibility constraints, our full formulation successfully distills noisy observations into a continuous, high-fidelity boundary shell. This isolates the unknown interior from the exterior, establishing a robust completion domain $\mathcal{E}$ for the latent SODF.

We further evaluate the impact of observation density on the visible hair shell reconstruction. As shown in Fig. 12, we compare the geometric fidelity of the optimized shell using 16, 32, 64, 128, and 256 views. Under sparse observation conditions (e.g., 16 or 32 views), the segment proxies lack sufficient multi-view optical constraints, yielding fragmented exterior boundaries and unresolved



Fig. 10. Strand-level reconstruction results on casual handheld captures. For each example, we show a representative input frame and the corresponding reconstructed 3D hairstyle rendered from a similar viewpoint. The results cover a range of hair lengths, styles, and viewing directions. Fine strand-level details are most clearly visible in the full-resolution electronic version.

local structures. At 64 views, the optimization reliably converges to a plausible geometric boundary, successfully establishing the rigorous completion domain $\mathcal{E}$ required for downstream volumetric inpainting. Providing additional views beyond this threshold primarily refines high-frequency localized features, as complex hair regions receive valid stereoscopic coverage from a wider baseline, while the macroscopic shell topology remains stable. The improvement also gradually saturates: the change from 128 to 256 views is noticeably smaller than that from 64 to 128 views. Furthermore, the optimization runtime remains largely invariant to the input view count in this comparison, because all view-count settings are optimized with the same number of GS iterations. Thus, increasing the number of views changes the amount of available supervision, but not the optimization budget used for convergence. We therefore use 256 views for fixed-scale Gaussian segment optimization in all experiments, maximizing visible-detail coverage while keeping the optimization budget fixed.



Fig. 11. Ablation of visible hair geometry reconstruction. The *w/o Optimize* column renders the initial coarse exterior, which is noisy and has not yet formed a complete visible envelope. The *w/o Body* variant optimizes the same proxies without body-depth visibility, leading to holes near the hair part. The full stage keeps the anatomical visibility constraint, producing more continuous visible support and a better hair-volume envelope for downstream completion. The bottom row shows zoomed-in crops of the marked region.

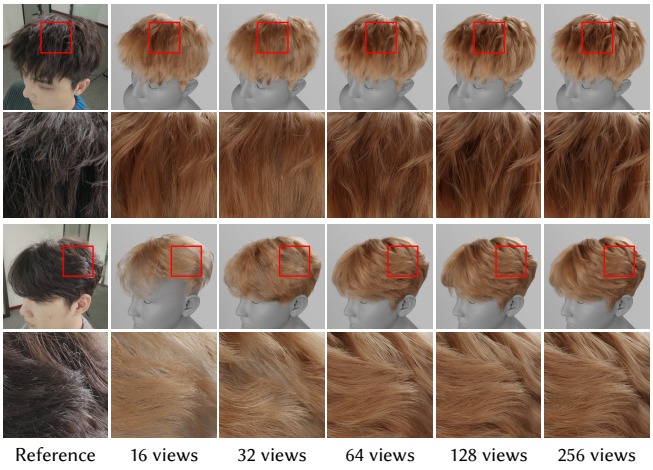


Fig. 12. Effect of view count on visible hair-segment shell reconstruction. From left to right: reference image and optimized results using 16, 32, 64, 128, and 256 views. Increasing the number of views improves visible coverage and recovers more high-frequency hair details.

4.5 Analysis of Latent SODF Inpainting

We analyze the latent SODF inpainting stage along three axes: the compact parameterization of SODF distributions, the design of the latent SODF auto-encoder, and the effectiveness of SODF-Flow for completing invisible interior hair structures.

Compact Distribution Parameterization. Fig. 13 compares three compact parameterizations at comparable intrinsic capacity. Antipodal SH with $L = 4$ uses 15 coefficients, a symmetric vMF mixture with $K = 4$ uses 20 explicit parameters but has 15 effective degrees of freedom, and our latent SODF uses a 16-dimensional code. Low-order SH smooths sharp modes, while a fixed-size vMF mixture captures only a few lobes; our codec better preserves localized multi-modal structure. Higher-capacity SH bases or vMF mixtures improve accuracy but sacrifice compactness; compressing their larger parameter sets with an auto-encoder would still require basis or mixture evaluation during strand assembly, whereas ours directly decodes queried bins.

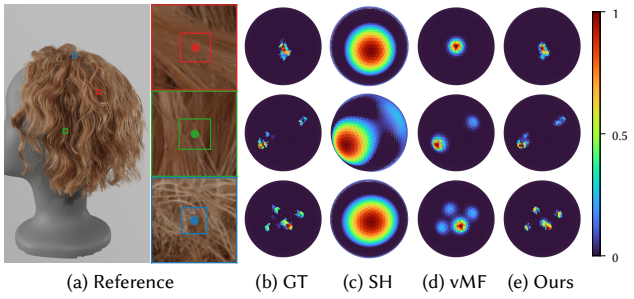


Fig. 13. Ablation of orientation-distribution parameterization. The left column shows a reference view and crop regions for three selected visible voxels. For each voxel, we compare the voxelized ground-truth antipodal orientation distribution (GT), antipodal spherical harmonics with degree $L = 4$ (SH), a symmetric von Mises–Fisher mixture with $K = 4$ components (vMF), and our 16-dimensional latent SODF. Each row uses a voxel-specific sphere view chosen from the GT distribution, shared by all methods.

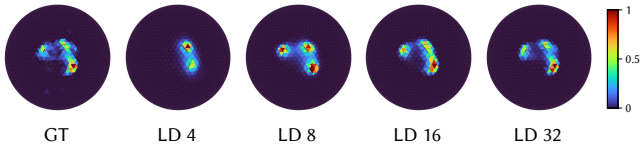


Fig. 14. Qualitative Ablation of Latent SODF Code Dimension on a representative SODF sample. Low-dimensional codes preserve only fragmented peaks, whereas 16 dimensions recover the main asymmetric structure and 32 dimensions further refine the support and probability mass.

Table 2. Quantitative Ablation of Latent SODF Code Dimension. The same auto-encoder architecture and loss are used for all reported variants.

Latent dim.	Sph. JSD ↓	L1 Err. ↓	Moment Err. ↓	Mode Err. ↓
4	0.1626	0.7389	0.0868	6.08
8	0.1138	0.5760	0.0409	5.90
16	0.0856	0.4767	0.0214	3.88
32	0.0676	0.4129	0.0127	3.04

Latent Dimension. As shown in Table 2, increasing the latent dimensionality consistently improves reconstruction accuracy. However, the visual comparison in Fig. 14 indicates that the major perceptual gain comes from increasing the code size from 8 to 16 dimensions, while the difference between 16 and 32 dimensions is relatively minor. Considering this accuracy–compactness trade-off and GPU memory constraints, we use a 16-dimensional latent SODF code in our experiments.

Table 3. Ablation of the latent SODF auto-encoder loss function. All variants are evaluated on the same architecture and latent dimensionality. Bold numbers denote the best performance and underlined numbers denote the second-best.

Variant	Sph. JSD ↓	L1 Err. ↓	Moment Err. ↓	Lat. Cond. ↓
w/o $\mathcal{L}_{\text{js}}$	0.1011	0.5398	**0.0050**	4.75
w/o $\mathcal{L}_{\text{ten}}$	0.0876	0.4862	0.0343	<u>4.15</u>
w/o $\mathcal{L}_{\text{lat}}$	<u>0.0870</u>	<u>0.4814</u>	0.0242	4.82
Full	**0.0856**	**0.4767**	<u>0.0214</u>	**3.63**

Table 4. Field-representation analysis on the held-out synthetic test set. Bold numbers denote the best performance.

Method / Variant	Dens. IoU ↑	Vox. Cham. ↓	Axis Err. ↓	SODF JSD ↓
MonoHair deterministic field	0.424	2.505	16.483	0.630
Ours SODF field	**0.771**	**0.385**	**16.293**	**0.499**

Auto-Encoder Loss Ablation. Table 3 ablates the SODF auto-encoder objective under the same architecture and latent dimensionality. Removing $\mathcal{L}_{\text{js}}$ lowers the Moment Err. but substantially worsens Sph. JSD, L1 Err., and Lat. Cond. Across all reported metrics, the full objective with $\mathcal{L}_{\text{js}}$, $\mathcal{L}_{\text{ten}}$, and $\mathcal{L}_{\text{lat}}$ provides the most balanced performance. We therefore use the full loss in our experiments.

SODF-Flow Effectiveness. MonoHair shares our philosophy of decoupling exterior reconstruction from interior inference. However, its interior module predicts a deterministic voxel-wise orientation field from multi-view image orientation cues, whereas our method formulates the occluded interior as geometry-conditioned inpainting over latent SODF-density fields. Table 4 compares our field representation with MonoHair's deterministic voxel field on held-out synthetic data, showing better occupied support, axial accuracy, and distribution matching. To assess hidden-field completion under stronger local curvature, we evaluate DiffLocks test samples with its procedural curliness parameter above 0.75. The resulting Density IoU, voxel Chamfer, Axis Error, and SODF JSD are 0.891, 0.402 mm, 24.69°, and 0.591, respectively, indicating stable occupancy recovery while local orientation completion remains more challenging. Fig. 15 further qualitatively validates SODF-Flow on a held-out synthetic hairstyle with ground-truth strands. Starting from sparse visible evidence, SODF-Flow completes the density and SODF field inside the occluded volume, and the sliced visualizations show that both completed SODFs and density follow the layered ground-truth structure. For clarity, the SODF and density panels use matched voxel support when comparing ours with the reference.

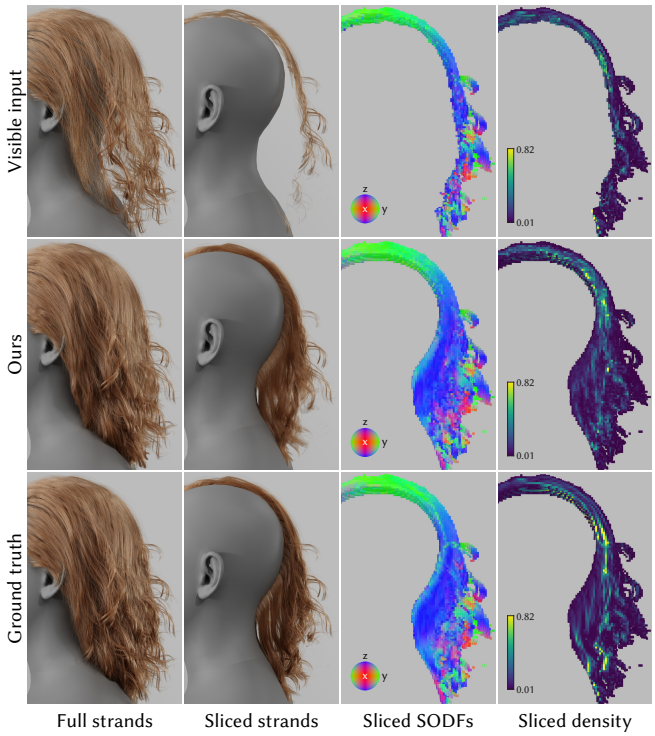


Fig. 15. Qualitative validation of SODF-Flow inpainting on synthetic data. From top to bottom, we show the visible input, our completed field, and the ground truth. The columns show full strands, sliced strands, sliced SODFs, and sliced density. SODF colors encode the dominant moment axis, and density colors share one scale across rows. The sliced views show that our network completes coherent interior density and orientation structure from sparse visible evidence.

4.6 Analysis of Strand Assembly

We evaluate the effect of our multi-hypothesis strand assembly by comparing it with two single-trajectory tracing baselines under the same boundary conditions and endpoint attachments. *Mean-axis tracing*, adapted from the deterministic ODE-style tracing in MonoHair [Wu et al. 2024], integrates along the probability-weighted expected axis of the local SODF. This collapses each multi-modal distribution into a single averaged direction, making the traced strands overly smooth and rigid in crossing or layered regions. *Stochastic tracing* samples one discrete axis from the SODF at each step, allowing non-dominant modes to be selected locally. However, because each step is sampled independently, it still lacks a mechanism for maintaining coherent long-range strand structure and often accumulates directional noise.

As shown in Fig. 16, mean-axis tracing produces smooth but stiff trajectories that fail to follow complex local hair flow. Stochastic tracing introduces more local variation, but this variation is not structurally organized and therefore does not reliably resolve ambiguous regions. Our assembly retains multiple candidate continuations during growth and selects compatible strand paths, producing more coherent structures in layered and tangled hairstyles.

5 Conclusion and Future Work

In this work, we introduced a novel framework that addresses the challenge of balancing reliable visible evidence preservation with the inference of occluded hair geometries from casual multi-view captures. By shifting the representational paradigm from deterministic vector fields to Spherical Orientation Distribution Functions (SODFs), we explicitly model the multi-modal nature of intersecting

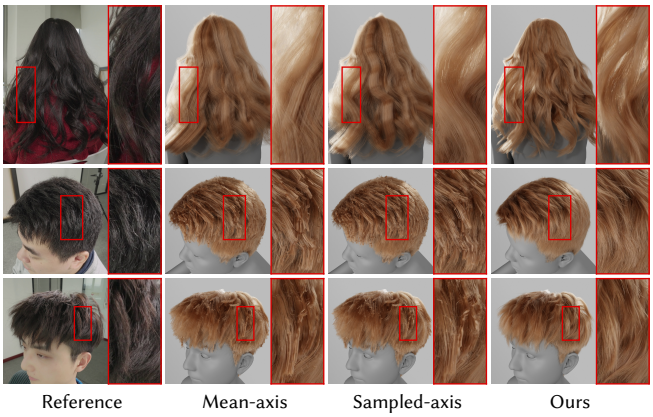


Fig. 16. Qualitative ablation of strand assembly from the same predicted probabilistic field. Mean-axis tracing follows the expected local axis, while stochastic tracing samples one local axis at each step. Both baselines commit to a single continuation during growth. Our assembly keeps multiple hypotheses and selects a globally compatible set of scalp-connected strands, improving coverage and flow in the highlighted regions.

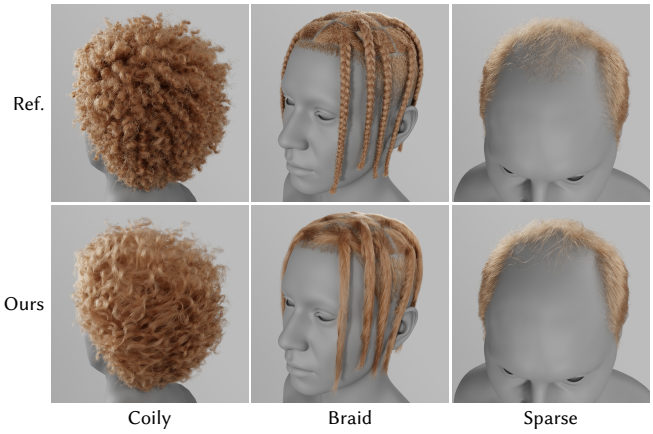


Fig. 17. Limitations. Top and bottom rows show references and our reconstructions, respectively. Coily and braided hairstyles exhibit simplified fine-scale winding and interleaving, while the synthetic sparse/balding case correctly leaves fully exposed scalp uncovered but misses some fine strands in thinning regions.

and layered fibers. Our decoupled pipeline distills noisy optical observations into a visible hair-segment shell, formulates the recovery of internal structures as geometry-conditioned latent flow matching, and replaces premature deterministic tracing with a global multi-hypothesis strand assembly. Extensive evaluations demonstrate that this probabilistic formulation effectively prevents the structural collapse associated with conventional unimodal assumptions, establishing a robust foundation for high-fidelity 3D hair reconstruction. The explicit strand output can also serve as geometric input to multi-view inverse rendering for hair appearance estimation [Sun et al. 2021].

While our probabilistic framework expands the representational capacity for complex hairstyles, several geometric and optical challenges remain. As illustrated by the representative failure cases in Fig. 17, these limitations open important avenues for future research: First, the fidelity of the visible hair shell heavily depends on the quality of passive optical evidence. Highly specular or uniformly dark hair can compromise multi-view orientation estimates and image-space constraints, propagating geometric uncertainty into the downstream completion domain. Very sparse or thinning hair remains challenging: fine strands may be missed during segmentation, and extremely sparse observations may not form a stable, scalp-connected shell, leading to incomplete reconstruction in these regions. Future work could combine more robust segmentation with uncertainty-aware optical features or active illumination. Second, our current spatial discretization has physical limits. Although per-voxel SODFs accommodate intersecting modes, extremely coily or tightly textured hair exhibits sub-voxel winding topologies that exceed the spatial Nyquist limit of our grid, which can lead to simplified geometry. Overcoming this requires exploring continuous latent representations or adaptive continuous-discrete hybrid structures. Finally, highly structured topologies, such as braids, remain challenging. These structures are governed by macroscopic interleaving rules that are difficult to fully capture using local probabilistic orientation distributions alone. To reconstruct such complex deterministic routing, future pipelines could integrate higher-order procedural or topological priors into the multi-hypothesis strand assembly stage.

Ultimately, we hope our shift towards probabilistic volumetric modeling and delayed geometric commitment can provide a promising direction for the high-fidelity digitization of complex human geometries.

Acknowledgments

We thank the authors of Neural Haircut for providing access to their captured video data, and the authors of DiffLocks for releasing their hairstyle dataset. We thank the authors of GroomCap for providing the comparison results used in our evaluation. We also thank the authors of MonoHair and Gaussian Haircut for releasing their implementations, which facilitated our comparative evaluation. This work is partially supported by NSF China (No. 62227806, 62572430, 62421003) and the XPLORER PRIZE.

References

Menglei Chai, Linjie Luo, Kalyan Sunkavalli, Nathan Carr, Sunil Hadap, and Kun Zhou. 2015. High-Quality Hair Modeling from a Single Portrait Photo. *ACM Trans. Graph.* 34, 6, Article 204 (2015), 10 pages. doi:10.1145/2816795.2818112

Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. 2016. AutoHair: Fully Automatic Hair Modeling from a Single Image. *ACM Trans. Graph.* 35, 4, Article 116 (2016), 12 pages. doi:10.1145/2897824.2925961

Menglei Chai, Lvdi Wang, Yanlin Weng, Yizhou Yu, Baining Guo, and Kun Zhou. 2012. Single-View Hair Modeling for Portrait Manipulation. *ACM Trans. Graph.* 31, 4, Article 116 (2012), 8 pages. doi:10.1145/2185520.2185612

Wesley Chang, Andrew L. Russell, Stephane Grabli, Matt Jen-Yuan Chiang, Christophe Hery, Doug Roble, Ravi Ramamoorthi, Tzu-Mao Li, and Olivier Maury. 2025. Transforming Unstructured Hair Strands into Procedural Hair Grooms. *ACM Trans. Graph.* 44, 4, Article 105 (2025), 20 pages. doi:10.1145/3731168

Yunlu Chen, Francisco Vicente Carrasco, Christian Häne, Giljoo Nam, Jean-Charles Bazin, and Fernando De la Torre. 2024. Doubly Hierarchical Geometric Representations for Strand-Based Human Hairstyle Generation. In *NeurIPS*, Vol. 37. Curran Associates, USA, 89728–89751.

Francesco Di Sario, Daniel Rebain, Dor Verbin, Marco Grangetto, and Andrea Tagliasacchi. 2026. Spherical Voronoi: Directional Appearance as a Differentiable Partition of the Sphere. In *CVPR*. IEEE, USA, 22529–22538.

Axel Elaldi, Guido Gerig, and Neel Dey. 2024. Equivariant Spatio-Hemispherical Networks for Diffusion MRI Deconvolution. In *NeurIPS*, Vol. 37. Curran Associates, USA, 52095–52126.

Epic Games. 2026. MetaHuman Creator. Retrieved May 12, 2026 from https://dev.epicgames.com/documentation/en-us/metahuman/metahuman-creator

Chengan He, Xin Sun, Zhixin Shu, Fujun Luan, Sören Pirk, Jorge Alejandro Amador Herrera, Dominik L. Michels, Tuanfeng Y. Wang, Meng Zhang, Holly Rushmeier, and Yi Zhou. 2025. PERM: A Parametric Representation for Multi-Style 3D Hair Modeling. In *ICLR 2025*. OpenReview.net, USA. https://openreview.net/forum?id=WKfb1xGXGx

Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2015. Single-View Hair Modeling Using a Hairstyle Database. *ACM Trans. Graph.* 34, 4, Article 125 (2015), 9 pages. doi:10.1145/2766931

Wenzel Jakob, Jonathan T. Moon, and Steve Marschner. 2009. Capturing Hair Assemblies Fiber by Fiber. *ACM Trans. Graph.* 28, 5, Article 164 (2009), 9 pages. doi:10.1145/1618452.1618510

Ben Jeurissen, Jacques-Donald Tournier, Thijs Dhollander, Alan Connelly, and Jan Sijbers. 2014. Multi-Tissue Constrained Spherical Deconvolution for Improved Analysis of Multi-Shell Diffusion MRI Data. *NeuroImage* 103 (2014), 411–426. doi:10.1016/j.neuroimage.2014.07.061

Leyang Jin, Yujian Zheng, Bingkui Tong, Yuda Qiu, Zhenyu Xie, and Hao Li. 2026. *HairOrbit: Multi-View Aware 3D Hair Modeling from Single Portraits.* arXiv:2604.02867 doi:10.48550/arXiv.2604.02867

Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. DeepMVSHair: Deep Hair Modeling from Sparse Views. In *SIGGRAPH Asia 2022 Conference Papers*. ACM, USA, Article 10, 8 pages.

Pengyu Long, Zijun Zhao, Min Ouyang, Qingcheng Zhao, Wei Yang, Lan Xu, and Jingyi Yu. 2025. Generating 3D Hair Strands from Images with Diverse Styles and Viewpoints. In *ACM MM 2025*. ACM, USA, 10602–10611.

Haimin Luo, Min Ouyang, Lan Xu, and Jingyi Yu. 2026. HairGPT: Strand-as-Language Autoregressive Modeling for Realistic 3D Hairstyle Synthesis. *ACM Trans. Graph.* 45, 4, Article 91 (2026), 14 pages. doi:10.1145/3811350

Haimin Luo, Min Ouyang, Zijun Zhao, Suyi Jiang, Longwen Zhang, Qixuan Zhang, Wei Yang, Lan Xu, and Jingyi Yu. 2024. *GaussianHair: Hair Modeling and Rendering with Light-Aware Gaussians.* arXiv:2402.10483 doi:10.48550/arXiv.2402.10483

Linjie Luo, Hao Li, Sylvain Paris, Thibaut Weise, Mark Pauly, and Szymon Rusinkiewicz. 2012. Multi-View Hair Capture Using Orientation Fields. In *CVPR*. IEEE, USA, 1490–1497.

Linjie Luo, Hao Li, and Szymon Rusinkiewicz. 2013a. Structure-Aware Hair Capture. *ACM Trans. Graph.* 32, 4, Article 76 (2013), 12 pages. doi:10.1145/2461912.2462026

Linjie Luo, Cha Zhang, Zhengyou Zhang, and Szymon Rusinkiewicz. 2013b. Wide-Baseline Hair Capture Using Strand-Based Refinement. In *CVPR*. IEEE, USA, 265–272.

Ryota Maeda, Kenshi Takayama, and Takafumi Taketomi. 2023. Refinement of Hair Geometry by Strand Integration. *Comput. Graph. Forum* 42, 7 (2023), e14970. doi:10.1111/cgf.14970

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 *ECCV 2020 (LNCS, Vol. 12346)*. Springer, Switzerland, 405–421.

Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheikh. 2019. Strand-Accurate Multi-View Hair Capture. In *CVPR*. IEEE, USA, 155–164.

Sylvain Paris, Héctor M. Briceño, and François X. Sillion. 2004. Capture of Hair Geometry from Multiple Images. *ACM Trans. Graph.* 23, 3 (2004), 712–719. doi:10.1145/1015706.1015784

Sylvain Paris, Will Chang, Oleg I. Kozhushnyan, Wojciech Jarosz, Wojciech Matusik, Matthias Zwicker, and Frédo Durand. 2008. Hair Photobooth: Geometric and Photometric Acquisition of Real Hairstyles. *ACM Trans. Graph.* 27, 3, Article 30 (2008), 9 pages. doi:10.1145/1360612.1360629

Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *CVPR*. IEEE, USA, 10975–10985.

Jonathan Rafael-Patino, Gabriel Girard, Raphael Truffet, Marco Pizzolato, Emmanuel Caruyer, and Jean-Philippe Thiran. 2021. The Diffusion-Simulated Connectivity (DiSCo) Dataset. *Data in Brief* 38 (2021), 107429. doi:10.1016/j.dib.2021.107429

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. 2025. SAM 2: Segment Anything in Images and Videos. In *ICLR 2025*. OpenReview.net, USA. https://openreview.net/forum?id=Ha6RTeWMd0

Radu Alexandru Rosu, Shunsuke Saito, Ziyan Wang, Chenglei Wu, Sven Behnke, and Giljoo Nam. 2022. Neural Strands: Learning Hair Geometry and Appearance from Multi-View Images. In *ECCV 2022 (LNCS, Vol. 13693)*. Springer, Switzerland, 73–89.

Radu Alexandru Rosu, Keyu Wu, Yao Feng, Youyi Zheng, and Michael J. Black. 2025. DiffLocks: Generating 3D Hair from a Single Image Using Diffusion Models. In *CVPR*. IEEE, USA, 10847–10857.

Shunsuke Saito, Liwen Hu, Chongyang Ma, Hikaru Ibayashi, Linjie Luo, and Hao Li. 2018. 3D Hair Synthesis Using Volumetric Variational Autoencoders. *ACM Trans. Graph.* 37, 6, Article 208 (2018), 12 pages. doi:10.1145/3272127.3275019

Yuefan Shen, Shunsuke Saito, Ziyan Wang, Olivier Maury, Chenglei Wu, Jessica Hodgins, Youyi Zheng, and Giljoo Nam. 2023. CT2Hair: High-Fidelity 3D Hair Modeling Using Computed Tomography. *ACM Trans. Graph.* 42, 4, Article 75 (2023), 13 pages. doi:10.1145/3592106

Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. 2023. Neural Haircut: Prior-Guided Strand-Based Hair Reconstruction. In *ICCV*. IEEE, USA, 19762–19773.

Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, and Justus Thies. 2024. Text-Conditioned Generative Model of 3D Strand-Based Human Hairstyles. In *CVPR*. IEEE, USA, 4703–4712.

Vanessa Sklyarova, Egor Zakharov, Malte Prinzler, Giorgio Becherini, Michael J. Black, and Justus Thies. 2025. Im2Haircut: Single-View Strand-Based Hair Reconstruction for Human Avatars. In *ICCV*. IEEE, USA, 10656–10665.

Tiancheng Sun, Giljoo Nam, Carlos Aliaga, Christophe Hery, and Ravi Ramamoorthi. 2021. Human Hair Inverse Rendering Using Multi-View Photometric Data. In *EGSR 2021*. Eurographics Association, Germany, 179–190.

Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, and Bo Zheng. 2024. Dr.Hair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments. In *CVPR*. IEEE, USA, 20601–20611.

Jacques-Donald Tournier, Fernando Calamante, and Alan Connelly. 2007. Robust Determination of the Fibre Orientation Distribution in Diffusion MRI: Non-Negativity Constrained Super-Resolved Spherical Deconvolution. *NeuroImage* 35, 4 (2007), 1459–1472. doi:10.1016/j.neuroimage.2007.02.016

David S. Tuch. 2004. Q-Ball Imaging. *Magnetic Resonance in Medicine* 52, 6 (2004), 1358–1372. doi:10.1002/mrm.20279

Lvdi Wang, Yizhou Yu, Kun Zhou, and Baining Guo. 2009. Example-Based Hair Geometry Synthesis. *ACM Trans. Graph.* 28, 3, Article 56 (2009), 9 pages. doi:10.1145/1531326.1531362

Yichen Wei, Eyal Ofek, Long Quan, and Heung-Yeung Shum. 2005. Modeling Hair from Multiple Views. *ACM Trans. Graph.* 24, 3 (2005), 816–820. doi:10.1145/1073204.1073267

Philippe Weier, Lukas Bode, Philipp Slusallek, Adrián Jarabo, and Sébastien Speierer. 2026. Beyond Positional Encoding: A 5D Spatio-Directional Hash Encoding. In *SIGGRAPH 2026 Conference Papers*. ACM, USA, Article 118, 10 pages.

Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2024. MonoHair: High-Fidelity Hair Modeling from a Monocular Video. In *CVPR*. IEEE, USA, 24164–24173.

Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. NeuralHDHair: Automatic High-Fidelity Hair Modeling from a Single Image Using Implicit Neural Representations. In *CVPR*. IEEE, USA, 1516–1525.

Shuang Wu, Youtian Lin, Feihu Zhang, Yifei Zeng, Yikang Yang, Yajie Bao, Jiachen Qian, Siyu Zhu, Xun Cao, Philip Torr, and Yao Yao. 2025. Direct3D-S2: Gigascale 3D Generation Made Easy with Spatial Sparse Attention. In *NeurIPS*, Vol. 38. Curran Associates, USA, 170778–170804.

Egor Zakharov, Vanessa Sklyarova, Michael J. Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. 2024. Human Hair Reconstruction with Strand-Aligned 3D Gaussians. In *ECCV 2024 (LNCS, Vol. 15074)*. Springer, Switzerland, 409–425.

Meng Zhang, Pan Wu, Hongzhi Wu, Yanlin Weng, Youyi Zheng, and Kun Zhou. 2018. Modeling Hair from an RGB-D Camera. *ACM Trans. Graph.* 37, 6, Article 205 (2018), 10 pages. doi:10.1145/3272127.3275039

Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. 2023. HairStep: Transfer Synthetic to Real Using Strand and Depth Maps for Single-View 3D Hair Modeling. In *CVPR*. IEEE, USA, 12726–12735.

Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus Gross, and Thabo Beeler. 2023. GroomGen: A High-Quality Generative Hair Model Using Hierarchical Latent Representations. *ACM Trans. Graph.* 42, 6, Article 267 (2023), 16 pages. doi:10.1145/3618309

Yuxiao Zhou, Menglei Chai, Daoye Wang, Sebastian Winberg, Erroll Wood, Kripasindhu Sarkar, Markus Gross, and Thabo Beeler. 2024. GroomCap: High-Fidelity Prior-Free Hair Capture. *ACM Trans. Graph.* 43, 6, Article 254 (2024), 15 pages. doi:10.1145/3687768

Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. 2018. HairNet: Single-View Hair Reconstruction Using Convolutional Neural Networks. In *ECCV 2018 (LNCS, Vol. 11215)*. Springer, Switzerland, 249–265.