Vid2Haircut: 3D Strand-Based Hairstyle Reconstruction from Video

Fatma Ben Ayed[3], Giorgio Becherini[1], Justus Thies[1,4], and Vanessa Sklyarova[1,2]

[1] Max Planck Institute for Intelligent Systems, [2] ETH Zurich
[3] Technical University of Munich, [4] Technical University of Darmstadt

Abstract

*We present Vid2Haircut, a novel approach for strand-based 3D hair reconstruction from monocular head-motion videos. While existing multi-view methods achieve high-fidelity results, they require controlled capture setups. In contrast, single-image approaches suffer from occlusion-driven ambiguities, particularly in unseen regions such as the back of the head. Recent monocular video methods improve reconstruction by leveraging learned priors, but may struggle under natural head motion. To address this, our approach reconstructs accurate geometry from a short monocular video by leveraging viewpoint variations induced by natural head motion to resolve occlusions. Specifically, we extend a learned hair prior [SZP*25] by jointly optimizing a shared, scalp-aligned hair map in a canonical space across multiple key-frames. To accommodate hair motion during capture, we incorporate a deformation MLP that predicts residual strand offsets, preventing frame-specific deformations from corrupting the canonical hairstyle. Additionally, we stabilize the reconstruction of poorly observed regions using visibility-aware updates and neighboring-strand smoothness constraints. Experiments on synthetic and real data demonstrate improved back-view consistency, scalp attachment, and overall reconstruction quality compared to state-of-the-art baselines, while requiring only casual monocular video as input. For additional results, please refer to the project page†.*

CCS Concepts

• ***Computing methodologies*** → ***3D imaging; Reconstruction;***

1. Introduction

Digital hair reconstruction is fundamental to realistic 3D avatar creation, playing a vital role in film, gaming, and virtual reality. Despite its importance, hair remains one of the most challenging structures to model due to its dense, anisotropic geometry, complex self-occlusions, and high style variability. Even small inaccuracies in hair volume, length, or scalp attachment are immediately perceptible and can significantly reduce the realism of a personalized avatar.

Existing solutions generally fall into three categories. High-fidelity multi-view methods achieve impressive results using dense camera setups, but are restricted to controlled environments and expensive hardware, limiting their practical deployment. Single-image approaches [RWF*25,ZJL*23,SZP*25] reduce hardware requirements and have demonstrated strong results in reconstructing strand-based hairstyles from a single image. However, they are fundamentally limited by occlusions and unseen regions, often leading to inaccuracies in the crown and back of the head, as well as inconsistencies in overall hair volume and length. Recent works explore hair reconstruction from monocular video of a static subject (e.g., MonoHair [WYK*24]), leveraging learned priors to improve reconstruction quality. While effective, these approaches typically assume stable capture conditions and require significant processing time. Rather than leveraging personalized observations, they tend to synthesize occluded regions based on learned priors, which can limit person-specific accuracy.

In this work, we explore a middle ground by leveraging short monocular videos with natural head motion, which reveal additional viewpoints beyond a single image. Such captures can be easily obtained in casual settings using commodity devices, yet they implicitly encode multi-view observations and motion cues that reveal hair length and help resolve occlusions. A similar capture paradigm is employed in commercial mobile hair scanning systems [GHH*25]; however, these methods are not publicly available and often produce lower-fidelity results. We propose Vid2Haircut, which extends the strand-based prior of Im2Haircut to multi-frame input. Our core contribution is a joint optimization framework that aggregates information across multiple keyframes into a shared, PCA-parameterized hair map. To handle mild non-rigid hair motion during capture, we introduce a deformation MLP that estimates residual per-frame offsets, preventing such variations from being encoded into the canonical hairstyle. Furthermore, we incorporate a visibility-aware optimization strategy that stabilizes poorly observed regions by restricting updates to frames with reliable observations. By leveraging multi-frame monocular video observations rather than relying solely on learned priors, our method achieves more accurate and person-specific hairstyle reconstruction.

† https://fatma18f.github.io/Vid2Haircut/

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.2312/egs.20261009

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

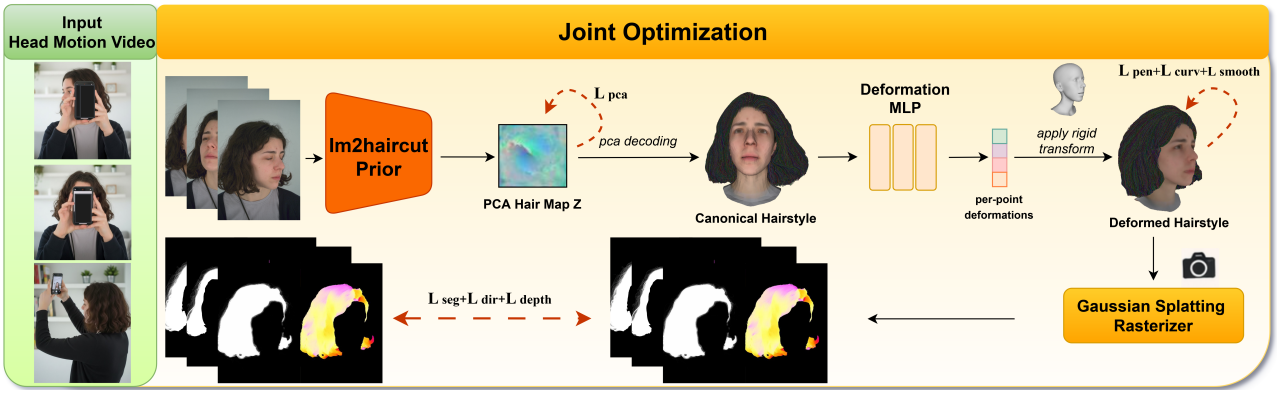

Figure 1: ***Method Pipeline.*** *Starting from a frontal reference frame, we initialize the canonical hairstyle using the Im2haircut prior and select keyframes from the input monocular video for joint optimization. A shared PCA hair map* $\mathbf{Z}$ *is refined across frames, while a deformation MLP accounts for non-rigid motion. Visibility-driven updates ensure that gradients modify the canonical map only in regions with reliable observations, with optimization stabilized by smoothness regularization and reprojection losses.*

2. Method

Our method reconstructs a 3D strand-based hairstyle from a short monocular head-motion video, in which the subject performs natural head rotations, spanning near-frontal to partial side views and providing viewpoints beyond a single image. We provide an example input sequence in the Appendix. We assume that hair undergoes mild motion during capture, while the overall hairstyle structure remains preserved. Following Im2Haircut [SZP*25], we represent hair using a PCA hair map $Z \in \mathbb{R}^{H\times W\times K}$ defined on the FLAME [SBFB19] scalp UV space, together with a baldness mask M. Each texel of Z encodes a strand as a polyline with a fixed number of points. In contrast to the single-image setting of Im2Haircut, we leverage multiple frames to refine a shared canonical hairstyle. We define a shared canonical hair map and model per-frame residual non-rigid motion as offsets Δx_t estimated by a deformation MLP f_θ. From the input sequence, we select M keyframes $\{I_t\}_{t=1}^{M}$ with camera parameters and estimate rigid head poses T_t via FLAME fitting. The most frontal frame is selected as the canonical reference view. Optimization proceeds in two stages:

Stage 1: Single-frame initialization. Given the canonical reference frame, we perform single-view optimization following the Im2Haircut pipeline. Specifically, we optimize the hair map $\mathbf{Z}$ for 100 iterations within the learned prior space using projection losses including the standard silhouette, direction, depth, and penetration terms [SZP*25]. This produces an initial 3D hairstyle that roughly matches the reference view.

Stage 2: Joint multi-frame refinement. We refine Z by integrating newly visible areas, such as the back and crown, revealed by the head motion. The canonical strands decoded from Z are transformed by the rigid head pose T_t, and then further adjusted by a learned non-rigid deformation model f_θ to account for hair motion. The strands are rasterized as strand-aligned 3D Gaussians and rendered to obtain a hair silhouette s_t, direction map o_t, and depth map d_t. We jointly optimize hair prior model weights and the deformation network parameters θ by minimizing our proposed video loss:

$$\begin{aligned}\mathcal{L}_{\text{video}} = (&\lambda_{\text{seg}}\|s_t-\hat{s}_t\|_1+\lambda_{\text{dir}}\|o_t-\hat{o}_t\|_1\\ &+\lambda_{\text{depth}}\|d_t-\hat{d}_t\|_1+\lambda_{\text{pen}}\mathcal{L}_{\text{pen}}\\ &+\lambda_{\text{smooth}}\mathcal{L}_{\text{smooth}}+\lambda_{\text{pca}}\mathcal{L}_{\text{pca}}),\end{aligned} \tag{1}$$

where $\hat{s}_t$, $\hat{o}_t$, and $\hat{d}_t$ are the ground-truth hair segmentation, direction, and depth maps, respectively. Similar to Img2haircut [SZP*25], the term $\mathcal{L}_{\text{pen}}$ penalizes hair-head intersections using a signed distance field defined on the FLAME mesh. The regularization term $\mathcal{L}_{\text{smooth}}$ enforces neighboring-strand smoothness in length and direction. $\mathcal{L}_{\text{pca}}$ enforces consistency of the PCA coefficients between observed and unobserved regions, encouraging plausible structure in areas lacking direct supervision. Further details on the loss terms are provided in the Appendix.

Deformation MLP. To model non-rigid hair motion, we introduce a deformation MLP f_θ that estimates per-frame residual offsets Δx_t conditioned on canonical points x, root-to-tip coordinates $u \in [0,1]$, frame specific latent code $\mathbf{z}_t$.

$$\Delta x_t = w_x \cdot f_\theta(x, u, \mathbf{z}_t), \tag{2}$$

where w_x is a weighting that suppresses motion near the roots and allows larger offsets near the tips. The deformed strand point in frame t is then given by $x_t = T_t(x+\Delta x_t)$. The network captures frame-dependent non-rigid variations, preventing them from being encoded into the canonical hairstyle.

Poorly-observed region stabilization. Regions such as the back and crown are only visible in a subset of frames. To stabilize these areas, we impose neighboring-strand smoothness on length and direction in canonical space, and apply a per-frame visibility mask v_t derived from the rasterizer. During optimization, gradients from frame t are applied only to canonical strand segments that are currently visible ($v_t = 1$), while invisible segments are updated indirectly through smoothness regularization and the PCA prior. This visibility-aware strategy prevents poorly observed regions from drifting due to inconsistent supervision across frames.

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

Figure 2: ***Comparison with baselines.*** *We compare against state-of-the-art hair reconstruction methods, including Im2Haircut [SZP*25], Difflocks [RWF*25], and Hairstep [ZJL*23], which operate in the single-image setting, as well as the commercial monocular video-based method MS-Hair [GHH*25]. Our method achieves more person-specific hairstyle reconstruction compared to all baselines.*

3. Experiments

Qualitative comparison. We compare our method against the single-view reconstruction methods Hairstep [ZJL*23], Difflocks [RWF*25], and Im2Haircut [SZP*25], as well as the commercial hair scanning solution MS-Hair [GHH*25]. Hairstep [ZJL*23] extracts global and local hairstyle features from an input image and reconstructs strand-based geometry through a hair-growing procedure. Difflocks [RWF*25] employs a diffusion model conditioned on image features to denoise the entire hairstyle. Im2Haircut [SZP*25] is an optimization-based approach that trains a prior model on synthetic data and optimizes in prior space to reconstruct the hairstyle. MS-Hair [GHH*25] extracts features from input images and applies fitting procedures to refine the hairstyle; however, implementation details are not publicly available. Our method extends Im2Haircut to handle monocular video input. In our experiments, we use M=6 keyframes, though the method supports an arbitrary number of input images.

We present results on five sequences from the NeRSemble dataset [KQG*23], visualizing frontal, side, and back views (see Figs. 2, 3). Our method produces more person-specific hairstyles compared to existing approaches. In particular, Hairstep [ZJL*23] often produces incorrect strand lengths, Difflocks [RWF*25] tends to bias reconstructions toward curlier hairstyles, and Im2Haircut [SZP*25] exhibits inter-strand smoothness artifacts. While some prior methods may produce more visually separated strands, this often comes at the cost of geometric inconsistencies, particularly in weakly observed or occluded regions. In contrast, our method produces more stable strand orientations and consistent geometry across viewpoints. Finally, although MS-Hair [GHH*25] takes video as input, it struggles to recover person-specific hairstyle details. Our method achieves higher reconstruction quality and can incorporate an arbitrary number of input images to further improve results.

Table 1: *Quantitative comparison on synthetic data.*

Method	Chamfer_pts ↓	Chamfer_angle ↓
Difflocks [RWF*25]	0.000241	0.3638
Im2Haircut [SZP*25]	0.000101	0.2846
Ours	**0.000033**	**0.2478**

Quantitative comparison. We evaluate Vid2Haircut on a synthetic benchmark generated using Unreal Engine, comprising realistic head motion sequences for four diverse hairstyles, providing controlled multi-view inputs for evaluation. As shown in Tab. 1, Vid2Haircut achieves the lowest Chamfer distance and Chamfer angle among all methods, indicating more accurate recovery of both strand geometry and orientation. This improvement is attributed to aggregating information across multiple frames, which resolves geometric ambiguities inherent in single-view inputs.

Ablation Studies. We provide qualitative comparisons on sev-

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.

Figure 3: ***Back-view comparison with baselines.*** *Note that ground-truth back views are not available.*

Figure 4: ***Ablation on deformation MLP network.*** *Omitting MLP leads to worse hair volume reconstruction.*

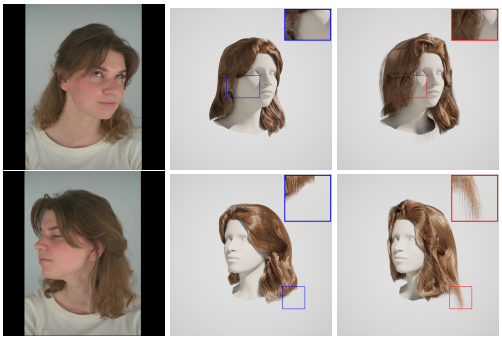

Figure 5: ***Ablation on the number of input images.*** *More input images yield improved reconstruction of unobserved regions.*

Figure 6: ***Ablation on*** $\mathcal{L}_{smooth}$. *Adding a smoothness loss regularizes the hair map and removes jagged artifacts.*

eral subjects from the NeRSemble [KQG*23] dataset to validate our design choices. We ablate the deformation MLP (w/o deform in Fig. 4) and strand smoothness regularization (w/o $\mathcal{L}_{\text{smooth}}$ in Fig. 6). The deformation MLP improves the overall volume and accuracy of the reconstructed hairstyle, while strand smoothness regularization ensures that adjacent strands maintain consistent orientations and curvatures. As shown in Fig. 6, removing this constraint results in localized noise and strand tangling. Finally, we demonstrate how reconstruction quality improves progressively with the number of input images (see Fig. 5).

4. Conclusion and Discussion

We present a novel method for reconstructing strand-based hairstyles from a monocular video. Our approach optimizes a hairstyle prior using visual guides extracted from the input images combined with 3D geometric regularization. To handle mild hair motion, we optimize a canonical hairstyle representation coupled with per-frame deformation fields. The formulation supports an arbitrary number of input images, enabling increasingly accurate, person-specific hairstyle reconstruction. Both qualitative and quantitative evaluations show that our method significantly outperforms single-image baselines when multiple input images are available. The reconstructed strand-based hairstyles are directly compatible with modern physics engines, enabling realistic rendering and simulation.

Acknowledgements. The project is supported by the ERC Starting Grant 101162081 "LeMo" and Max Planck ETH Center for Learning Systems (CLS).

References

[GHH*25] GRASSAL P.-W., HORMANN L., HAMLAOUI N., LEISTNER T., ARDIZZONE L.: A mobile scanning solution to reconstruct strand-based hairstyles. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Talks* (2025). doi:10.1145/3721239.3734111. 1, 3

[KQG*23] KIRSCHSTEIN T., QIAN S., GIEBENHAIN S., WALTER T., NIESSNER M.: Nersemble: Multi-view radiance field reconstruction of human heads. *ACM Trans. Graph. 42* (2023). doi:10.1145/3592455. 3, 4

[RWF*25] ROSU R. A., WU K., FENG Y., ZHENG Y., BLACK M. J.: Difflocks: Generating 3d hair from a single image using diffusion models, 2025. arXiv:2505.06166. 3

[SBFB19] SANYAL S., BOLKART T., FENG H., BLACK M. J.: Learning to regress 3d face shape and expression from an image without 3d supervision, 2019. URL: https://arxiv.org/abs/1905.06817, arXiv:1905.06817. 2

[SZP*25] SKLYAROVA V., ZAKHAROV E., PRINZLER M., BECHERINI G., BLACK M. J., THIES J.: Im2haircut: Single-view strand-based hair reconstruction for human avatars, 2025. arXiv:2509.01469. 1, 2, 3

[WYK*24] WU K., YANG L., KUANG Z., FENG Y., HAN X., SHEN Y., FU H., ZHOU K., ZHENG Y.: Monohair: High-fidelity hair modeling from a monocular video, 2024. URL: https://arxiv.org/abs/2403.18356, arXiv:2403.18356. 1

[ZJL*23] ZHENG Y., JIN Z., LI M., HUANG H., MA C., CUI S., HAN X.: Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling, 2023. arXiv:2303.02700. 1, 3

© 2026 The Author(s).
Proceedings published by Eurographics - The European Association for Computer Graphics.