MonoHair: High-Fidelity Hair Modeling from a Monocular Video

Supplementary Material

1. Implementation Details and Data Preprocessing

1.1. Datasets preprocessing

Synthetic data. We train the DeepMVSHair* on USC-HairSalon[2], which contains 343 hairstyles aligned with a template head. We augment the dataset to 2,744 samples with three degrees of freedom: scaling, rotation, and translation. Then, for each sample, we generate a strand model, the ground-truth 3D orientation, and a rendered 2D unidirectional map with 16 fixed views. We also evaluate our method on another synthetic dataset [10]. Specifically, for each model, we use Blender [1] to render 150 RGB images with a resolution of 1920×1080, employing different camera poses. Then, we reconstruct the corresponding strand model and evaluate it with the ground truth.

Real-world data. We also evaluate our method utilizing a public multi-view H3DS dataset [10] and a set of real-world captured monocular videos. For the H3DS dataset, each scene has 32 views with their corresponding camera parameters. For monocular videos, to better initialize the coarse geometry, we select 150-200 frames around the subject using an image quality assessment network [8] and employ structure-from-motion with COLMAP [6] to obtain the camera intrinsic and extrinsic parameters. Then, we calculate the 2D orientation map utilizing 180 Gabor filters with $\sigma_x = 1.8$, $\sigma_y = 2.4$, frequency $\omega = 0.25$, and kernel size $k = 17$. Lastly, we also generate hair masks using human matting [3] and semantic segmentation [4].

1.2. Implementation details

To initialize the coarse point cloud, we densely sample around the coarse geometry produced by Instant-NGP [5] at a high resolution of $512 \times 512 \times 384$. For the PMVO, we remove the low confidence area when performing the patch sample strategy. Besides, we select 10 reference frames based on confidence values for different initializations of $L^{\mathbf{p}}$ to calculate $\mathcal{L}_{opt}$ more robustly. For interior hair geometry inference, we grow short segments using $\mathcal{P}_{\text{out}}$ to render the undirectional strand maps. Then, we train DeepMVSHair* on 16 fixed views at 1280×720 resolution with the learning rate of 0.0005. The training process takes about 6 days on a single NVIDIA RTX 3090, 3 days for the 3D orientation prediction network and 3 days for the 3D occupancy prediction network.

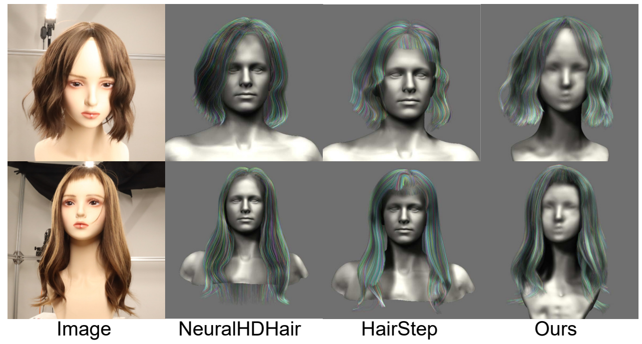

Figure 1. Qualitative comparison with single view-based hair modeling methods. The reconstruction results are produced by the open source code and pre-trained models.

2. Additional Experiments and Results

2.1. More comparisons

Comparison with single view-based image methods. We provide an extended qualitative comparison with the single view-based reconstruction methods [9, 11]. As shown in Fig. 1, obviously, our method achieves higher fidelity results.

Extended comparison with Neural Haircut. We provide an additional comparison with Neural HairCut [7] in Fig. 2. While their method can handle both long and short hair, it struggles to accurately reconstruct some complex hairstyles, such as curly hair. In contrast, our approach is versatile and adaptable to more hairstyles.

2.2. Additional ablation studies

Ablation studies on synthetic data. We also evaluate each component of our method on a synthetic dataset [10]. We render the ground-truth 3D strand model into RGB images using Blender and reconstruct it. As shown in Fig. 3. PMVO helps us maintain high-fidelity exterior hair geometry while DeepMVSHair* assists in reconstructing the complete strand model. As indicated in Tab. 1, without PMVO, both precision and recall are greatly reduced due to the errors of camera parameters. On the other hand, without DeepMVSHair*, we can achieve the highest precision but the lowest recall, since the reconstruction results contain only the outer layer of hair. Finally, our full method combining PMVO and DeepMVSHair* achieves the highest recall and F-score.

Evaluation of DeepMVSHair*. To evaluate the effectiveness of our improvements to DeepMVSHair, we con-

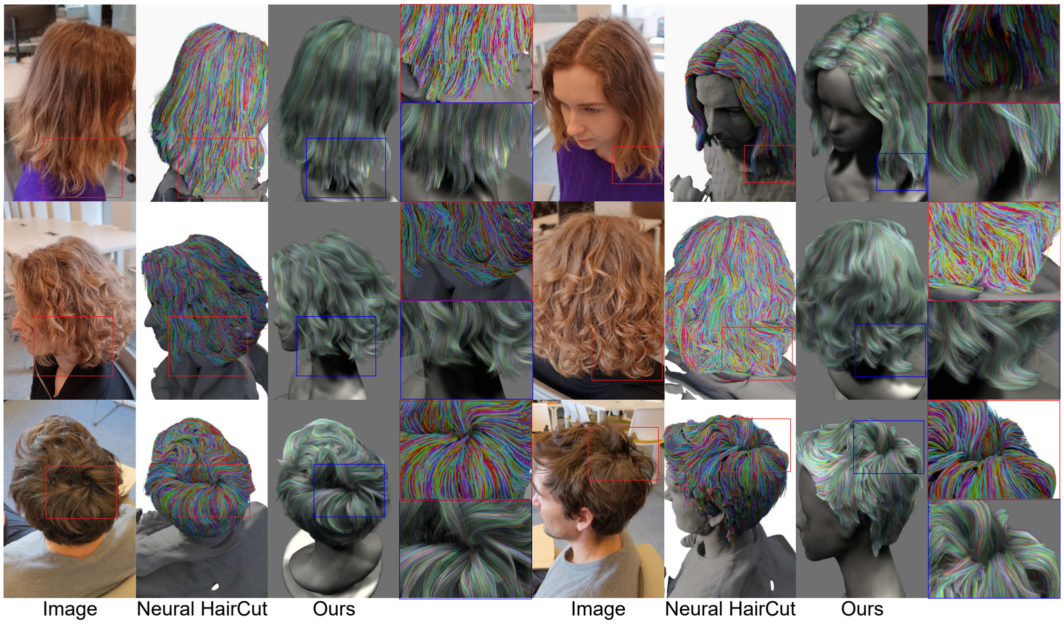

Figure 2. Extended qualitative comparison with Neural Haircut [7].

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision			Recall			F-score		
w/o PMVO	30.3	54.5	67.3	6.7	14.5	21.6	11.0	22.9	32.7
w/o DeepMVSHair*	**65.2**	**87.4**	**95.6**	6.3	15.2	19.8	11.5	25.9	32.8
Full	60.8	83.3	92.1	**10.4**	**19.3**	**25.9**	**17.8**	**31.3**	**40.4**

Table 1. Quantitative evaluation of individual components of our method on the synthetic dataset [10].

ducted two sets of ablation studies: 1) without the undirectional strand map (w/o Map) and 2) without the proposed new loss function $\mathcal{L}_{ori}$. The comparison results are shown in Fig. 4. With no undirectional strand map, the reconstructed results are more susceptible to the loss of geometric details due to ambiguity. While the loss function $\mathcal{L}_{ori}$ helps us train DeepMVSHair* more robustly.

Comparison with different patch sizes. We also compared the effects of different sample sizes (patch size) in PMVO on the reconstruction of the exterior hair geometry. As shown in Fig. 5 and Tab. 2, setting the patch size to 5 better balances performance and efficiency.

2.3. More results

Lastly, we provide additional reconstruction examples from a monocular video showcasing various hairstyles such as short, long, straight, curly, and wavy hair. As shown in Figs. 6 to 8, MonoHair can handle diverse, complex hairstyles and achieve high-fidelity results.

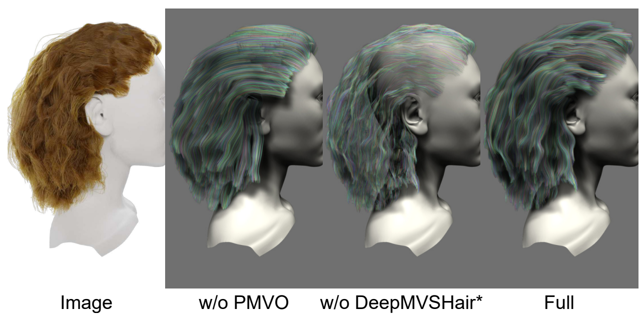

Figure 3. Qualitative evaluation of each key component of our method on synthetic data.

Figure 4. Qualitative evaluation of the improvements to DeepMVSHair.

Patch size	1	3	5	7
Time	∼30min	∼ 90min	3h-4h	6h-8h

Table 2. Comparison of time consumption under different patch sizes. As the patch size increases, the time consumption increases significantly.

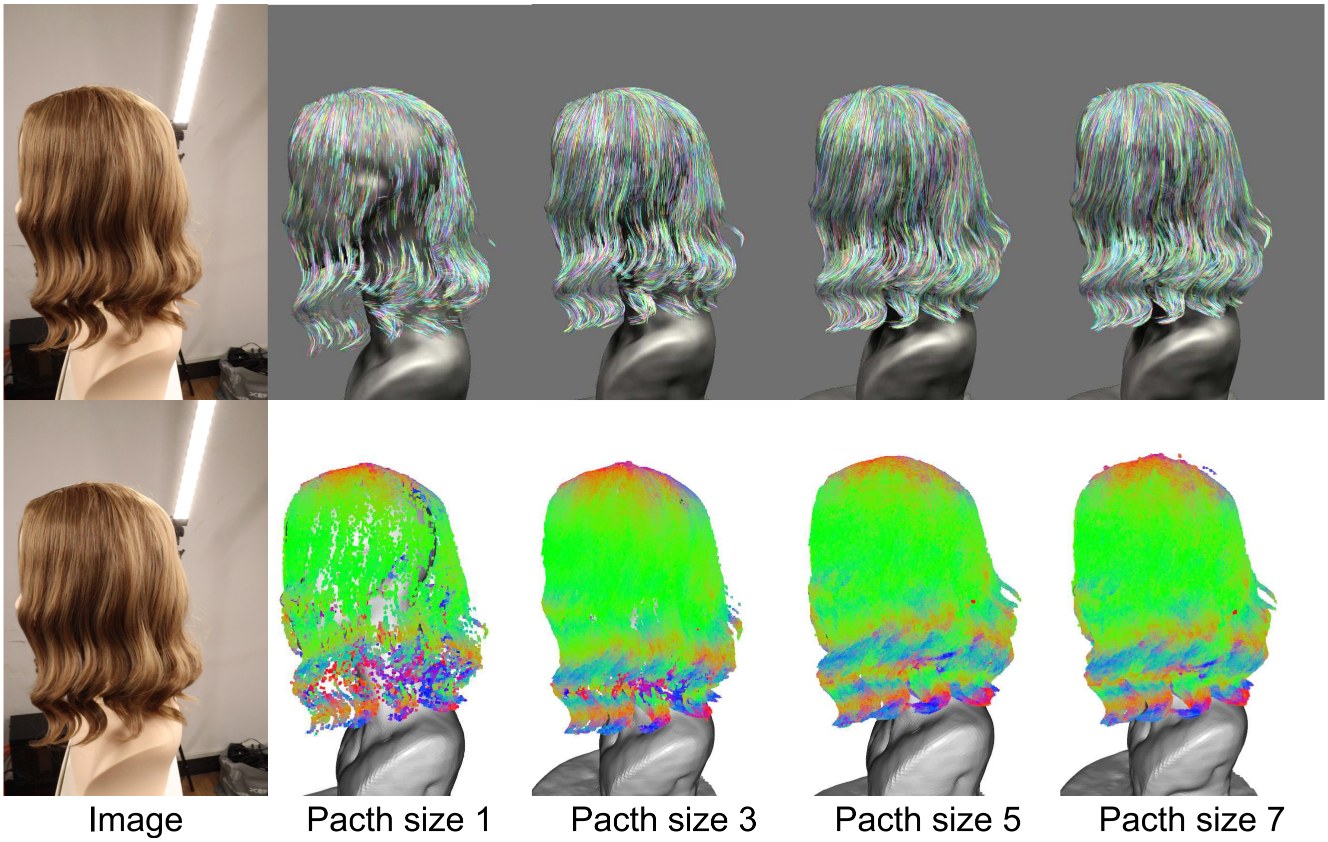

Figure 5. Qualitative comparison of PMVO with different patch sizes. When the patch size is greater than 5, there is no significant improvement in the reconstructed geometry. The above results are all the exterior geometry of the hair.

References

[1] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2023. 1, 4

[2] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (ToG)*, 34(4):1–9, 2015. 1

[3] Zhanghan Ke, Jiayu Sun, Kaican Li, Qiong Yan, and Rynson WH Lau. Modnet: Real-time trimap-free portrait matting via objective decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 1140–1147, 2022. 1

[4] Kunliang Liu, Ouk Choi, Jianming Wang, and Wonjun Hwang. Cdgnet: Class distribution guided network for human parsing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4473–4482, 2022. 1

[5] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 1

[6] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 1

[7] Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. Neural haircut: Prior-guided strand-based hair reconstruction. In *ICCV*, 2023. 1, 2

[8] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3667–3676, 2020. 1

[9] Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. Neuralhdhair: Automatic high-fidelity hair modeling from a single image using implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1526–1535, 2022. 1

[10] Cem Yuksel, Scott Schaefer, and John Keyser. Hair meshes. *ACM Transactions on Graphics (TOG)*, 28(5):1–7, 2009. 1, 2

[11] Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling. In *CVPR*, 2023. 1

Figure 6. Additional reconstruction results of our method based on monocular videos. We use two rendering styles to showcase our results: colorful strands for displaying hair geometry and Blender [1] for realistic rendering.

Figure 7. Additional reconstruction results of our method based on monocular videos.

Figure 8. Additional reconstruction results of our method based on monocular videos.