Multi-View Hair Capture using Orientation Fields

Linjie Luo[1]   Hao Li[2,1]   Sylvain Paris[3]   Thibaut Weise[4]   Mark Pauly[4]   Szymon Rusinkiewicz[1]

[1]Princeton University   [2]Columbia University   [3]Adobe Systems, Inc.   [4]EPFL

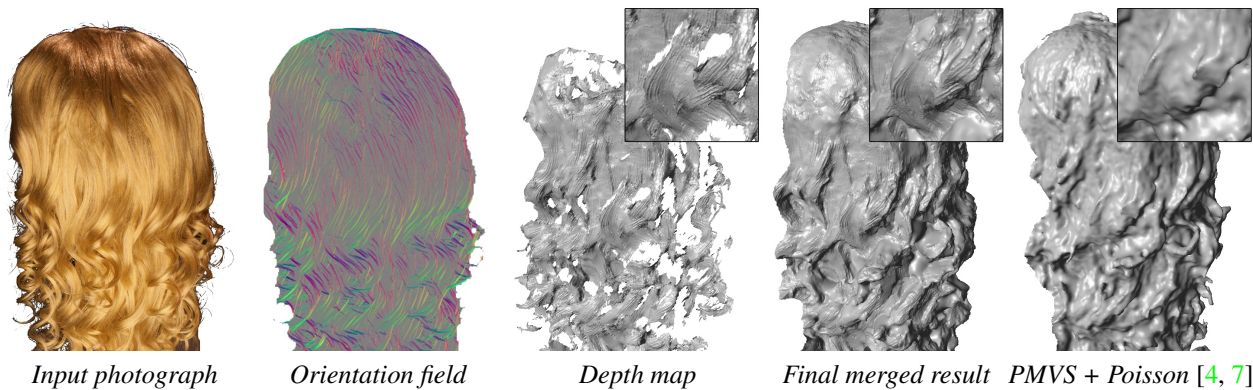| *Input photograph* | *Orientation field* | *Depth map* | *Final merged result* | *PMVS + Poisson* [4, 7] |

Figure 1: *We begin with many high-resolution photographs (with unconstrained lighting), compute an orientation field for each, and perform multi-view stereo matching using a metric based on orientation similarity. The resulting depth maps show high-resolution details of hair strands and we integrate them into a single merged model. In contrast, conventional multi-view stereo algorithms and merging techniques* [4, 7] *fail at capturing the fine hair structures.*

Abstract

Reconstructing realistic 3D hair geometry is challenging due to omnipresent occlusions, complex discontinuities and specular appearance. To address these challenges, we propose a multi-view hair reconstruction algorithm based on orientation fields with structure-aware aggregation. Our key insight is that while hair's color appearance is view-dependent, the response to oriented filters that captures the local hair orientation is more stable. We apply the structure-aware aggregation to the MRF matching energy to enforce the structural continuities implied from the local hair orientations. Multiple depth maps from the MRF optimization are then fused into a globally consistent hair geometry with a template refinement procedure. Compared to the state-of-the-art color-based methods, our method faithfully reconstructs detailed hair structures. We demonstrate the results for a number of hair styles, ranging from straight to curly, and show that our framework is suitable for capturing hair in motion.

1 Introduction

Despite its aesthetic importance in defining a person's look, the reconstruction of realistic hair is often neglected in methods for 3D acquisition. Hair is one of the most challenging objects to capture using standard computer vision techniques. Occlusions are omnipresent, and hair's strand geometry precludes general surface-based smoothness priors for stereo (e.g., [17]). Besides, the highly specular nature of hair fibers [9] is not well modeled by standard appearance models. Even the latest facial reconstruction techniques (e.g., [1]) exclude hair from the region of interest, and so hair modeling largely remains a manual task in practice [10]. Dedicated acquisition methods have been proposed [12, 6], but they rely on scanning rigs that are costly and difficult to build. Furthermore, these methods require lengthy capture sessions that limit their suitability to treat hair in motion. While Hair Photobooth is perhaps the best example of this approach, we argue that in fact all active illumination methods will be challenging to apply to hair capture, especially in the dynamic setting. Single-frame methods such as noise patterns are ineffective because of the complex geometry and occlusion of hair, while multi-frame structured light methods such as Gray codes fail for moving hair.

To address these difficulties we investigate a single-shot passive multi-view stereo approach that requires only consumer-level hardware and produces results on par with or superior to existing techniques.

Naively applying existing stereo techniques fails because hair is specular, and hence, observed color varies quickly with changes in viewpoint. Our key insight is that hair orientation (which can be computed using a
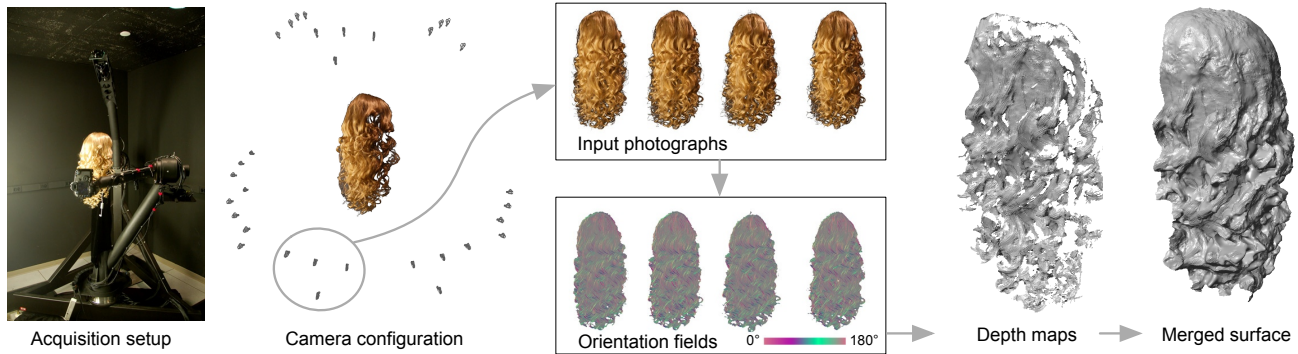
1

Figure 2: *Our reconstruction pipeline. We use a robotic gantry to capture static images from different views. Each 4 views are grouped into a cluster to reconstruct a single accurate depth map based on orientation fields computed from input photographs. (Section 2 and 3) Finally, the depth maps are merged into a single hair surface (Section 4).*

filter bank of many — e.g. 180 — oriented Difference-of-Gaussians (DoG) kernels) is a stable feature across nearby viewpoints and can be used as a reliable matching criterion. We leverage this idea by defining a stereo matching metric based on *similarity of local orientation* that is insensitive to local changes in brightness. We further improve matching robustness by aggregating local evidence with a scheme that accounts for the local hair structure; i.e., we aggregate matching costs *along* strands but not perpendicular to them. We incorporate the resulting matching score into a coarse-to-fine multi-view stereo framework based on Markov Random Fields (MRF). We operate on small sets of nearby viewpoints at a time to minimize the effect of inter-view differences in foreshortening, then merge all the resulting depth maps into a globally consistent detail-preserving model of the whole head of hair.

We demonstrate that our approach can handle a wide variety of hairstyles, that strands can be grown within the reconstructed hair volume which is suitable for rendering, and that our approach is capable of capturing hair in motion when using video cameras. We quantitatively evaluate the precision of our approach using synthetic data.

1.1 Related Work

Multi-view stereo has received significant attention [15] but applying a generic algorithm to hair images yields unsatisfying results, as we illustrate in Figure 6.

A few dedicated techniques have been designed to capture hair. Paris et al. [11] introduced the idea of estimating the orientation of hair in images, coupled with an analysis of the highlights on the hair. This analysis requires a light source to move along predefined known trajectories. Paris et al. [12] later described *Hair Photobooth*, a complex system made of several light sources, projectors, and video cameras that captures a rich set of data to extract the hair geometry and appearance. Jakob et al. [6] showed how to capture individual hair strands using focal sweeps with a camera controlled by a robotic gantry. While accurate, these active techniques are expensive and the capture is inherently slow because of their design. For example, the method of Paris et al. [12] relies on time multiplexing and requires thousands of images for a single reconstruction. Similarly, Jakob et al. [6] uses many input photographs with a sweeping focus plane across the hair volume.

Wei et al. [16] proposed a purely passive technique based on several handheld photographs. Their approach also relies on hair orientation fields, but uses a coarse *visual hull* as the approximate bounding geometry for hair growing.

The numerical accuracy of existing hair acquisition techniques remains unexplored since only visual evaluations were conducted. We address this shortcoming by performing ground-truth analyses using synthetically generated data (see Section 5).

1.2 Contributions

Compared to previous work, we introduce the following contributions:

- a passive multi-view stereo approach capable of reconstructing finely detailed hair geometry;
- a robust matching criterion based on the local orientation of hair;
- an aggregation scheme to gather local evidence while taking hair structure into account;
- a progressive template fitting procedure to fuse multiple depth maps; and
- a quantitative evaluation of our acquisition system.

2 Local Hair Orientation

Because hair is highly specular, its color varies quickly when the viewpoint moves, making standard multi-view stereo fail, as shown in Figure 6. We address this issue by replacing colors with local orientations. In this section, we describe how we reliably estimate local directions of hair strands.
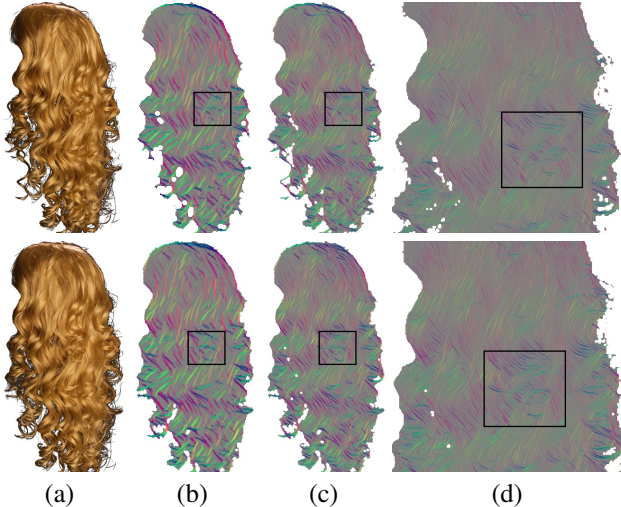
Figure 3: *The multi-resolution orientation fields (b)-(d) computed from two nearby input images (a) shown in separate rows. Note that the local hair orientations highlighted by boxes correspond naturally between the two views. Finer levels of orientation field reveal high resolution hair details.*

Paris et al. [11] first introduced dense orientation fields for hair modeling. Our orientation field computation differs from the prior one in that we only consider highlighted hair strands (i.e., positive filter response). We observe in our experiments that reliable orientations are difficult to obtain in dark regions due to the poor signal-to-noise ratio of negative parts of the filter response. Instead, we recover the orientations in dark regions using available data obtained from the coarser resolution level where more data is aggregated and/or from a finer resolution level where finer and brighter structures are revealed.

Formally, given oriented filters K_θ generated by rotating an original x-aligned filter K_0 by angles $\theta \in [0, \pi)$, we define the orientation $\Theta(x, y)$ of image I at pixel (x, y) as $\Theta(x, y) = \arg\max_\theta |K_\theta * I(x, y)|$. To eliminate the $\pm\pi$ ambiguity of the orientation, we map Θ to the complex domain as in [11] by $\Phi(x, y) = \exp(2i\Theta(x, y))$. We also use the (nonlinearly mapped) maximum response $F(x, y) = \max_\theta |K_\theta * I(x, y)|^\gamma$ as a confidence measure in our stereo algorithm: it captures both the strength of the image intensity and confidence in the orientation, at the filter's characteristic scale. The power-law mapping enhances weak responses and improves reconstruction quality. We use $\gamma = 0.5$ for all our datasets. Finally, our orientation field $O(x, y)$ is defined by taking the product of $\Phi(x, y)$ and $F(x, y)$, where the maximum filter response was positive:

$$O(x, y) = \begin{cases} F(x, y)\, \Phi(x, y), & \max_\theta\big(K_\theta * I(x, y)\big) > 0 \\ 0, & \max_\theta\big(K_\theta * I(x, y)\big) \leq 0 \end{cases} \tag{1}$$

We select a DoG filter for K_0. Specifically, $K_0(x, y) = \big(G_\sigma(x) - G_{\sigma'}(x)\big) G_{\sigma''}(y)$, where G_σ is a 1D zero-mean Gaussian with standard deviation σ. We use filters 1 degree apart, for a total of 180 filters.

Next, we describe a coarse-to-fine optimization strategy that requires orientation fields at multiple resolutions. We generate these fields with a pyramid structure to accelerate the computation: we recursively downsample the image for coarse levels in the pyramid and apply each oriented filter K_θ. We use a fixed sized K_θ with $\sigma = 0.5$, $\sigma' = 1$ and $\sigma'' = 4$ for all levels of the orientation field. The multi-resolution oriented pyramid is visualized in Figure 3.

3 Partial Geometry Reconstruction

In this section, we assume that we have a set of images of the hair from a few nearby viewpoints. We will discuss in Section 5 the specific setups we have used in our experiments. For now, we focus on reconstructing the partial geometry of the hair seen from these viewpoints. In Section 4 we will describe how to merge the pieces coming from several groups of cameras to form a full head of hair.

We formulate the partial reconstruction process as an MRF optimization based on the computed orientation fields, i.e., we seek to reconstruct the geometry that best explains the orientations observed in each view. To make this process robust, we use a coarse-to-fine strategy and locally aggregate evidence using a scheme inspired by [13].

We reconstruct a depth value $D(p)$ for each pixel p of the center reference view using orientation fields computed from all cameras. The reconstruction volume is bounded by the nearest and farthest depths, d_{near} and d_{far}.

3.1 Energy Formulation

We use an MRF energy minimization framework to optimize for D. The total MRF energy $E(D)$ with respect to D consists of a data term $E_d(D)$ and a smoothness term $E_s(D)$:

$$E(D) = E_d(D) + \lambda E_s(D), \tag{2}$$

where λ is the smoothness weight. The data energy is the sum of the per-pixel data cost $e_d(p, D)$ for each pixel p of the reference view while the smoothness energy is the weighted sum of the depth deviation between p and its 4-connected neighbors $\mathcal{N}(p)$:

$$E_d(D) = \sum_{p \in pixels} e_d(p, D)$$

$$E_s(D) = \sum_{p \in pixels} \sum_{p' \in \mathcal{N}(p)} w_s(p, p') \big|D(p) - D(p')\big|^2. \tag{3}$$

The MRF cues $w_s(p, p')$ encode different depth continuity constraints between adjacent pixels p and p'. To enforce a strong depth continuity along the hair strands where orientations are similar, we define $w_s(p, p')$ as a Gaussian of the

orientation distance in the reference image:

$$w_s(p, p') = \exp\left(-\frac{|O_{\text{ref}}(p) - O_{\text{ref}}(p')|^2}{2\sigma_o^2}\right). \quad (4)$$

The parameter σ_o controls the constraint sensitivity and is set to $\sigma_o = 0.15$ for all our datasets.

Similar to [14], we formulate the data term e_d based on the multi-resolution orientation field computed in Section 2. We define e_d as the sum of the matching costs $e_d^{(l)}$ of each level l from the orientation field for all views:

$$e_d(p, D) = \sum_{l \in levels} e_d^{(l)}(p, D)$$

$$e_d^{(l)}(p, D) = \sum_{v \in views} c_v\left(O_{\text{ref}}^{(l)}(p), O_v^{(l)}\left(P_v(p, D)\right)\right), \quad (5)$$

where $O_{\text{ref}}^{(l)}$ and $O_v^{(l)}$ are the orientation fields at level l of the reference view and of adjacent view v, respectively. $P_v(p, D)$ is the projection of the 3D point defined by the depth map D at pixel p onto view v. The cost function c_v for adjacent view v is defined as:

$$c_v(O, O') = -\Re\left\{O^* O' \exp\left(2i(\phi_{\text{ref}} - \phi_v)\right)\right\}, \quad (6)$$

where $\Re(z)$ denotes the real part of a complex number z, ϕ_{ref} and ϕ_v are the angles between image x-axis and the vector from the image principal point to the epipole of the other view for reference view and adjacent view v. Intuitively, $c_v(O, O')$ measures the deviation of the orientation fields for corresponding pixels as the negative correlation of the two orientation vectors O and O', and the correction factor $\exp\left(2i(\phi_{ref} - \phi_v)\right)$ compensates for the influence of the camera pair's different tilting angles on the orientation field comparison.

The data term $e_d(p, D)$ is a function on the volume defined by the pixel image of the reference view and each possible depth value d in the interval $[d_{\text{near}}, d_{\text{far}}]$

3.2 Structure-Aware Aggregation

To improve robustness and adaptivity of the data term energy to the local structure of the orientation field, we perform cross guided filtering [5] on each level l based on the orientation field of the reference view on that level. This process builds upon the idea of structure-aware aggregation introduced by Yoon and Kweong [18] for stereo and applied to other problems by Rhemann et al. [13]. However, none of these techniques can be directly applied to hair because they rely on color data. We address this issue in the rest of this section.

Before the data energy of each level $e_d^{(l)}(p, D)$ is summed in 5, each data energy at depth D of pixel p in the energy volume is aggregated as a weighted average of data energies of neighboring pixels p':

$$e_d^{(l)}(p, D) \leftarrow \sum_{p' \in \omega_p} W^{(l)}(p, p')\, e_d^{(l)}(p', D), \quad (7)$$

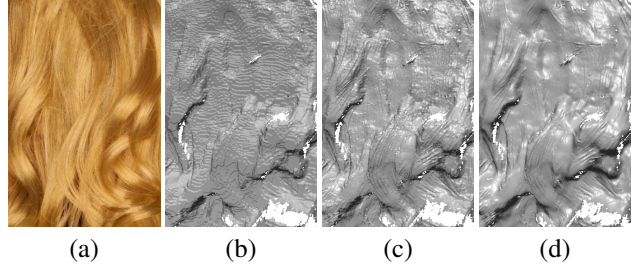
(a)   (b)   (c)   (d)

Figure 4: *The stages of depth map refinement improve reconstructed quality. The reference input image (a). The surface reconstructed from the initial MRF-optimized depth map (b) shows quantization artifacts that are removed by sub-pixel refinement (c). A post-reconstruction guided filtering step further improves quality (d).*

where $W^{(l)}(p, p')$ is the guided filter weight and ω_p is a local window centered at p.

Generalizing the expressions derived in [5], we define the weight $W^{(l)}(p, p')$ based on the orientation field:

$$W^{(l)}(p, p') =$$
$$\frac{1}{|\omega|^2} \sum_{k:(p,p') \in \omega_k} \left(1 + \frac{\Re\{(O(p) - \mu_k)^*(O(p') - \mu_k)\}}{\sigma_k^2 + \varepsilon}\right), \quad (8)$$

where the summation is over all pixels k such that p and p' are in a local window ω_k around k, $|\omega|$ is the number of pixels in the window, ε controls the structure-awareness based on the orientation, and μ_k and σ_k are the average and standard deviation of orientation, respectively, within ω_k.

Note that the computation in Equation (7) can be done efficiently regardless of the size of ω_k [5]. This enables efficient aggregation with large-kernel windows on high resolution datasets. We use 7×7, 11×11, and 15×15 kernel windows for the 3 levels of orientation fields, and we set $\varepsilon = 0.1^2$ for all of our examples.

After aggregation, the resulting energy in Equation (2) can be efficiently minimized by graph cuts [2].

3.3 Depth Map Refinement

We employ a sub-pixel refinement technique similar to [1] to refine the integer depth map optimized by graph cuts. To be specific, for each pixel p on the reference view and its associated depth $D(p)$, we look up its data cost $e_0 = e_d(p, D(p))$ and the data cost $e_{-1} = e_d(p, D(p) - 1)$ and $e_{+1} = e_d(p, D(p) + 1)$ for the adjacent depth values $D(p) - 1$ and $D(p) + 1$. The subpixel depth $D'(p)$ is computed as:

$$D'(p) = \begin{cases} D(p) - 0.5, & e_{-1} < e_0, e_{+1} \\ D(p) + 0.5\frac{e_{-1} - e_{+1}}{e_{-1} - 2e_0 + e_{+1}}, & e_0 < e_{-1}, e_{+1} \\ D(p) + 0.5, & e_{+1} < e_0, e_{-1} \end{cases} \quad (9)$$

We then apply guided filtering once again on the depth map based on the finest orientation level to further reduce the stereo noise with the same weights as in Equation 8:

$$D'(p) \leftarrow \sum_{p' \in \omega_p} W(p, p') \, D'(p'). \qquad (10)$$

Figure 4 shows how the reconstructed surface evolves after applying each of the refinement steps discussed above. Note the importance of subpixel refinement, without which the features are overwhelmed by quantization artifacts. The post-reconstruction guided filtering step increases surface quality modestly, but is not a replacement for the structure-aware aggregation.

4 Final Geometry Reconstruction

The previous section described how we produce a set of partial reconstructions (depth maps) of the hair volume using small groups of nearby views. We combine these into a model of the full head of hair by aligning and merging these pieces.

We begin by forming a *coarse template* that establishes the topology and overall geometry of the merged reconstruction. This step accounts for the fact that different depth maps see different portions of the hair and, in overlapping regions, may have misalignment. We use *Poisson surface reconstruction* [7] on the union of all input points, but restrict the depth of the octree to level 6 (i.e., reconstructing the surface at a resolution of $2^6 \times 2^6 \times 2^6$ voxels) in order to capture only the coarsest geometry of the hair.

We then form a *refined template* by deforming the coarse template towards each of the original depth maps, moving along each depth map's line of sight. The warping is done using a graph-based non-rigid registration algorithm [8] that maximizes local rigidity, so that missing data in each depth map does not result in a bumpy surface, and so that the amount of deformation may be controlled (i.e., regularized). We repeat this deformation step by reinitializing its rest energy state 10 times, reducing the amount of regularization for later iterations. The resulting template effectively av-
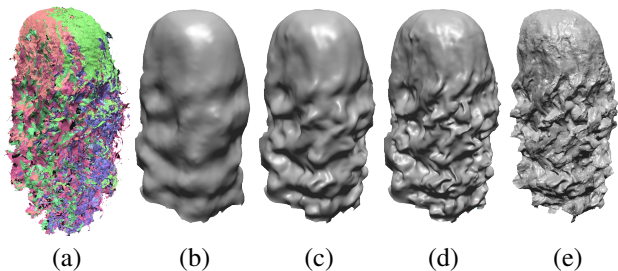
erages the shape of all the input depth maps where they overlap, while remaining smooth.

The *final mesh* is obtained by re-introducing high-frequency details from the depth maps onto the refined template. This is done using a more efficient linear deformation model based also described in [8] where the high-frequency details are represented as offsets along the mesh vertex normals.

We have found that this three-stage process successfully combines the goals of establishing a consensus topology (rejecting spurious disconnected components), averaging geometry in overlapping regions, and maintaining the details present in the original depth maps to the maximum extent possible (Figure 5). In contrast, previous (single-step) surface merging algorithms average away details in the presence of misalignment in overlap regions, and are often unable to cope with spurious topology.

5 Evaluation

We demonstrate the performance of our approach using a variety of setups. First, we show results for a large (8×4 camera positions) set of high-resolution (21 Mpix) images of static subjects: wigs. We use a robotic camera gantry for our setup, fix a wig on the central turn table, and use a Canon EOS 5D Mark II SLR camera to capture images (Figure 2). The arm of the gantry is about 60cm long and can be rotated freely, horizontally and vertically, to arbitrary latitudes and longitudes around the hair.

The camera is positioned in groups to obtain aggregate views of the hair for the depth map computation in Section 3. Each group consists of four different camera po-
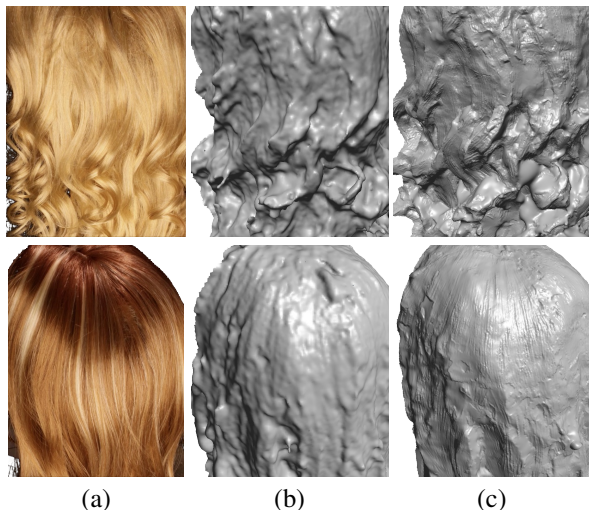

(a) (b) (c)

Figure 6: *Qualitative evaluation on the two real captured datasets of different hair styles (a) between the state-of-the-art multi-view stereo methods: PMVS + Poisson [4, 7] (b) and our method (c). Note that our method preserves hair strand details.*
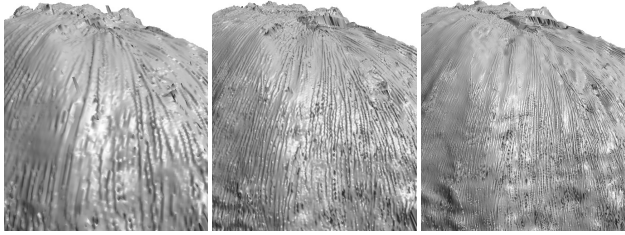

(a) (b) (c) (d) (e)

Figure 5: *The stages of final geometry reconstruction. (a) Partial depth maps from all views. (b) Coarse template. (c)-(d) Template refinement. (e) Re-introducing high frequency details.*

Figure 7: *Reconstruction results on different levels. From left to right the resolution of the depth map increases from 0.4M to 1.5M and 6M pixels, respectively.*

sitions in a T-pose: center, left, right and bottom. Each position is 10 degrees apart from the neighboring position in terms of gantry arm rotation. The left and right cameras in the T-pose provide balanced coverage with respect to the center reference camera. Since our system employs orientation-based stereo, matching will fail for horizontal hair strands (more specifically, strands parallel to epipolar lines). To address this problem, a bottom camera is added to extend the stereo baselines and prevent the "orientation blindness" for horizontal strands.

We use 8 groups of 32 views for all examples in this paper. Three of these groups are in the upper hemisphere, while a further five are positioned in a ring configuration on the middle horizontal plane, as shown in Figure 2. We calibrate the camera positions with a checkerboard pattern [19], then perform foreground-background segmentation by background color thresholding combined with a small amount of additional manual keying. A large area light source was used for these datasets.

Qualitative Evaluation The top two rows of Figure 11 show reconstructions for two different hairstyles, demonstrating that our method can accommodate a variety of hairstyles — straight to curly — and handle various hair colors. We also compare our results on these datasets with [4] and [7] in Figure 6. Note the significant details present in our reconstructions: though we do not claim to perform reconstruction at the level of individual hair strands, small groups of hair are clearly visible thanks to our structure-aware aggregation and detail-preserving merging algorithms.

In Figure 7 and Figure 8, we show how our reconstruction algorithm scales with higher resolution input and more camera views. Higher resolution and more views greatly increase the detail revealed in the reconstructed results.

Quantitative Evaluation To evaluate our reconstruction accuracy quantitatively, we hired a 3D artist to manually create a highly detailed hair model as our ground truth model. We then rendered 8 groups of 32 images of this model with the same camera configuration as in the real capture session. We ran our algorithm on the images and compared the depth maps of our reconstruction and the
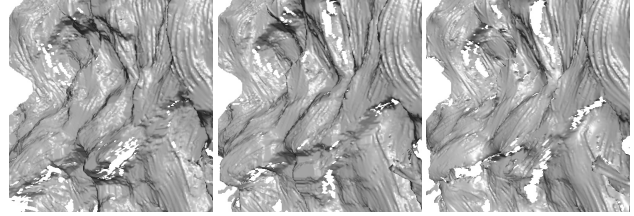

Figure 8: *Comparison between the depth map reconstructed with 2, 3, 4 cameras.*

ground truth model from the same viewpoints. The results are shown in Figure 9. On average, the distance between our result and the ground truth model is 5 mm, and the median distance is 3 mm. We also ran a state-of-the-art multi-view algorithm [4, 7, 1] on the synthetic dataset, and the statistics of its numerical accuracy are similar to ours. However, as shown in Figure 9, their visual appearance is a lot worse with the presence of blobs and spurious discontinuities.

Timings Our algorithm performs favorably in terms of efficiency. On a single thread of a Core i7 2.3GHz CPU, each

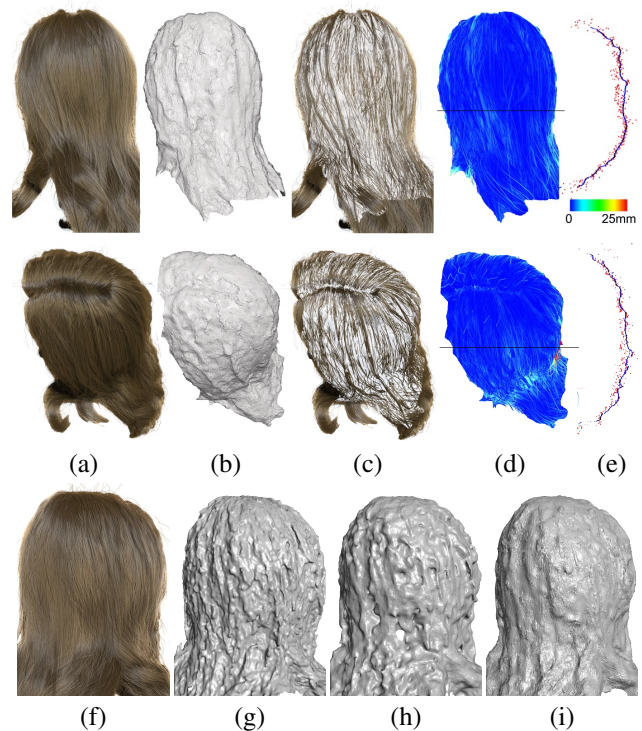(a)　　(b)　　(c)　　(d)　　(e)

(f)　　(g)　　(h)　　(i)

Figure 9: *We evaluate the accuracy of our approach by running it on synthetic data (a), (f). The result is shown in (b), and is overlaid to the synthetic 3D model in (c). The difference between our reconstruction in the ground-truth 3D model is on the order of a few millimeters (d). We show a horizontal slice of the depth map in (e): the ground-truth strands are shown in red and our reconstruction result in blue. Compared to PMVS + Poisson [4, 7] (g) and [1] (h), our reconstruction result (i) is more stable and accurate.*
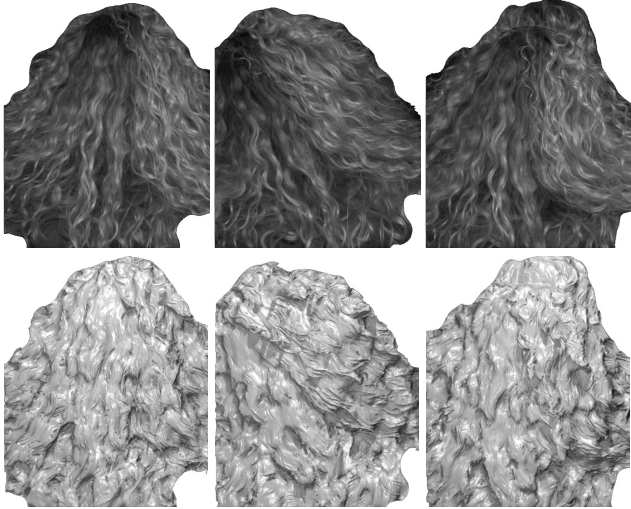
Figure 10: *Sample frames (first row) and the reconstructed depth maps (second row) from our dynamic hair capture setup.*

full-resolution (1404×936) depth map reconstruction takes 4.5 minutes. Multiple depth map reconstructions can be easily parallelized using more cores. In the final geometry reconstruction stage, the coarse template reconstruction (via Poisson) takes on average 30 seconds, template refinement 5 to 6 minutes, and final detail synthesis 20 seconds. In comparison, the Hair Photobooth [12] timings are on the order of several hours.

Dynamic Hair Capture A major advantage of our approach over previous work [12, 6] is that, being completely passive, it is amenable to simultaneous multi-view acquisition. This paves the way towards capturing hair in motion. As a proof of concept, we built a dynamic capture setup made of 4 high-speed video cameras capturing 640×480 pixels at 100 frames per second. Figure 10 shows the reconstruction results of a few sample frames from 4 captured 200-frame videos (see accompanying video for full input videos and the reconstruction results). However, because of the limited resolution of the high-speed video cameras, we were not able to achieve similar quality to our static reconstructions.

6 Conclusion and Future Work

We have proposed a passive multi-view reconstruction algorithm based on multi-resolution orientation fields. We demonstrated quantitatively that accurate measurements can be achieved by using orientation-based stereo. Combined with structure-aware aggregation, our method faithfully recovers detailed hair structures that surpass the quality of previous state-of-the-art methods. We also demonstrate the capability of our method to capture hair geometry in motion.

However, there are several limitations in our current method that needs to be addressed in the future. The structure-aware aggregation step, performed on the matching energy volume, imposes a fronto-parallel bias on the reconstruction result. This bias becomes considerable towards the edges of the reconstructed depth map, resulting in spurious and misaligned geometry. A possible solution is to use slanted window matching similar to [3] that fuses information from different depth layers.

As mentioned in Section 5, our orientation-based matching metric is "blind" for strands parallel to the epipolar plane. This can be addressed by adding more cameras of different baselines. Although we have experimented on a few different camera configurations, it still requires further investigation to find out the best possible configuration for our orientation-based hair capture system for optimal coverage of the entire head.

References

[1] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-quality single-shot capture of facial geometry. *ACM Trans. Graph.*, 29(4):40:1–40:9, July 2010. 1, 4, 6

[2] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, November 2001. 4

[3] D. Bradley, T. Boubekeur, and W. Heidrich. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Proceedings of CVPR*, 2008. 7

[4] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *PAMI*, 32:1362–1376, 2010. 1, 5, 6

[5] K. He, J. Sun, and X. Tang. Guided image filtering. In *ECCV (1)*, pages 1–14, 2010. 4

[6] W. Jakob, J. T. Moon, and S. Marschner. Capturing hair assemblies fiber by fiber. *ACM Trans. Graph.*, 28(5):164:1–164:9, December 2009. 1, 2, 7

[7] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of SGP*, 2006. 1, 5, 6

[8] H. Li, B. Adams, L. J. Guibas, and M. Pauly. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph.*, 28(5):175:1–175:10, December 2009. 5

Figure 11: *Final results for the three datasets (Alice, Lindsay and synthetic) reconstructed using our method. For each, we show one overview and two close-up views along with their reference images.*

[9] S. Marschner, H. W. Jensen, M. C. andS. Worley, and P. Hanrahan. Light scattering from human hair fibers. *ACM Trans. Graph.*, 22(3):780–791, July 2003. 1

[10] T. Mihashi, C. Tempelaar-Lietz, and G. Borshukov. Generating realistic human hair for *The Matrix Reloaded*. In *ACM SIGGRAPH Sketches and Applications Program*, 2003. 1

[11] S. Paris, H. Briceño, and F. Sillion. Capture of hair geometry from multiple images. *ACM Trans. Graph.*, 23(3):712–719, August 2004. 2, 3

[12] S. Paris, W. Chang, O. I. Kozhushnyan, W. Jarosz, W. Matusik, M. Zwicker, and F. Durand. Hair Photobooth: Geometric and photometric acquisition of real hairstyles. *ACM Trans. Graph.*, 27(3):30:1–30:9, August 2008. 1, 2, 7

[13] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *CVPR*, 2011. 3, 4

[14] K. Sasaki, S. Kameda, H. Ando, M. Sasaki, and A. Iwata. Stereo matching algorithm using a weighted average of costs aggregated by various window sizes. In *Proceedings of ACCV*, 2006. 4

[15] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of CVPR*, 2006. 2

[16] Y. Wei, E. Ofek, L. Quan, and H.-Y. Shum. Modeling hair from multiple views. *ACM Trans. Graph.*, 24(3):816–820, July 2005. 2

[17] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon. Global stereo reconstruction under second order smoothness priors. In *Proceedings of CVPR*, 2008. 1

[18] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *PAMI*, 28(4):650–656, April 2006. 4

[19] Z. Zhang. A flexible new technique for camera calibration. *PAMI*, 22(11):1330–1334, November 2000. 6