Dynamic Hair Capture

Linjie Luo
Princeton University

Hao Li
Columbia University / Princeton University

Thibaut Weise
EPFL

Sylvain Paris
Adobe Systems Inc.

Mark Pauly
EPFL

Szymon Rusinkiewicz
Princeton University

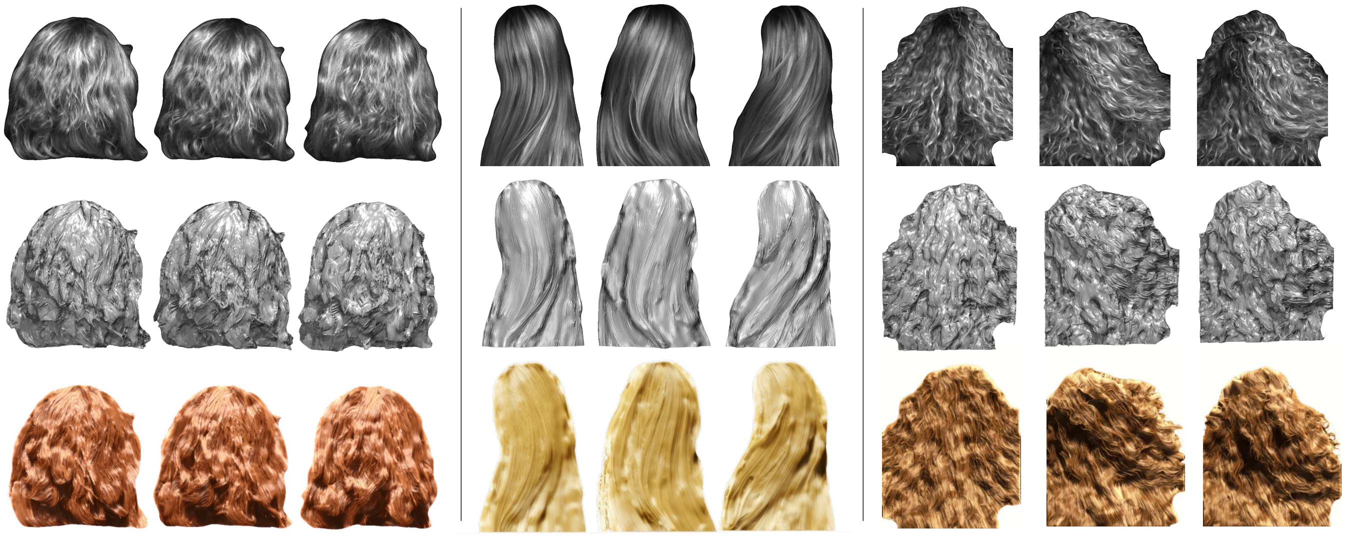

Figure 1: *Our system reconstructs a temporally coherent set of hair fibers for real-world dynamic hair. It accommodates a variety of hair types and styles, as well as nontrivial motion (top: input video, middle: reconstructed envelope surface, bottom: synthesized hair strands).*

Abstract

The realistic reconstruction of hair motion is challenging because of hair's complex occlusion, lack of a well-defined surface, and non-Lambertian material. We present a system for passive capture of dynamic hair performances using a set of high-speed video cameras. Our key insight is that, while hair color is unlikely to match across multiple views, the response to oriented filters will. We combine a multi-scale version of this orientation-based matching metric with bilateral aggregation, a MRF-based stereo reconstruction technique, and algorithms for temporal tracking and de-noising. Our final output is a set of hair strands for each frame, grown according to the per-frame reconstructed rough geometry and orientation field. We demonstrate results for a number of hair styles ranging from smooth and ordered to curly and messy.

1 Introduction

The hairstyle is one of a person's most noticeable features and accentuates one's face and overall appearance. However, despite this prominence, hair and in particular hair in motion remains difficult to handle in the digital world. For this reason, special-effect movies involving digital clones, such as The Matrix, frequently feature short-haired actors [Mihashi et al. 2003], thereby sidestepping the difficulties of modeling or capturing moving hair. While several successful capture systems exist for body poses, facial expressions, and clothes [Vlasic et al. 2009; Li et al. 2009; de Aguiar et al. 2008; Vlasic et al. 2008], the capture of dynamic hair remains mostly unexplored. And while motion capture is routinely used in production, as far as we know, animated CG hair is always simulated, e.g. [Ward et al. 2010].

Though hair simulation has increased in practicality and realism, we envision that both of these desirable features can be extended even further through a framework based on direct capture of real-life moving hair. Densely captured hair geometry offers several advantages over a pure simulation approach [Selle et al. 2008; McAdams et al. 2009]:

- **Generality:** Hair's complex dynamics can be recreated accurately, independently of its geometric complexity and style. Physically-based hair simulations, on the other hand, are often restricted to specific hair styles (flat and straight) and do not generalize to complex structures (curly and fluffy).
- **Flexibility:** While a physics simulation requires precise estimation of material properties (weight, density…) and knowledge of effecting forces (constraints, wind…), the dynamics of directly captured hair data come for free and can be easily integrated with other non-hair geometries.
- **Efficiency:** Because the animation of individual hair strands is fully automated through capture, the artist no longer needs to spend time on adjusting simulation parameters in attempt to accurately match real-world observations (which, in most cases, is not even possible).

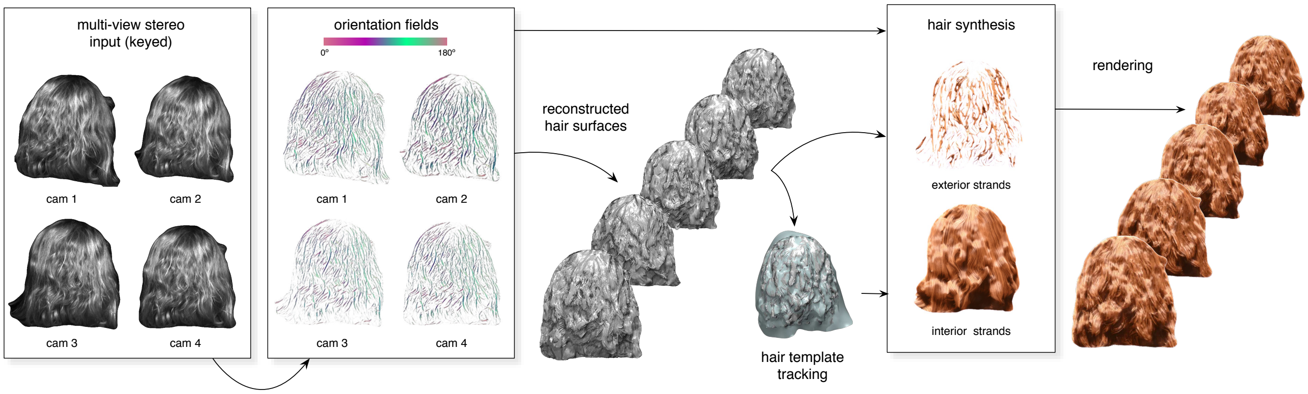

Figure 2: *Several stages of our dynamic hair reconstruction pipeline. We compute orientation images of keyed monochrome images recorded from four different viewpoints, perform multi-view stereo reconstruction to obtain hair envelope meshes, track these meshes to obtain temporally coherent seeds, and grow hair strands according to a volumetric orientation field.*

Despite these considerable advantages, work on hair capture is comparatively scarce, in contrast with the development of hair design tools and physical simulation techniques. To ease the creation of highly accurate hair models, several static reconstruction techniques, e.g. [Paris et al. 2008; Jakob et al. 2009], were proposed, but are subject to a lengthy acquisition process, making them unsuitable for dynamic hair capture. The first work that strives at capturing both hair geometry and motion was only recently introduced by Yamaguchi and coworkers [2008]. This method produces temporally coherent dynamic hair strands by growing hair segments from the root of a scalp model and enforcing their orientations to follow the observed hair strands in multiple images. Although temporally coherent hair models can be obtained, the system is restricted to very low input resolution, a prohibitively smooth motion prior, and can only handle limited hair styles. In particular, it builds upon a long-held belief that passive stereo correspondence methods are inapplicable to the reconstruction of hair, and hence a rough global shape estimate can only be obtained from the visual hull [Wei et al. 2005; Yamaguchi et al. 2008]. As a result, complex hair styles that exhibit many concave regions and fine-scale details such as interpenetrating hair fibers cannot be accurately modeled.

Our key insight is to retain the flexibility of a passive stereo acquisition setup, but to build upon it a temporally-coherent reconstruction system that makes use of the unique appearance of hair. In particular, we begin by performing stereo matching not on windows of the raw color or intensity image, but rather on multiresolution ***orientation images*** (Section 4.3), which represent at each pixel the direction in which there is the strongest oriented image feature [Paris et al. 2004]. The key benefit of performing stereo matching using these features is their unique robustness against specularity and other sources of view-dependent appearance in hair, since they effectively act as high-pass filters that suppress overall differences in reflected intensity. They exploit the fact that hair geometry is inherently 1D (continuous along each strand) rather than 2D (piecewise continuous along the surface).

We enhance the performance of our orientation-based matching metric by introducing a ***bilateral aggregation*** step (Section 4.4), inspired by Yoon and Kweon [2006], that enforces smoothness along hair strands by exploiting the hair structures revealed in the orientation fields. We further leverage this approach to cover hair structures of all scales by using a multi-resolution pyramid of the orientation fields. The resulting multi-view stereo reconstruction gives us a plausible envelope of the hair geometry that captures many features such as clumps of hairs (though not individual hair strands). We apply temporal de-noising on the reconstructed mesh sequence using dense correspondences computed between adjacent frames. Following the directions of a volumetric orientation field, we ***grow hair strands*** with seeds placed on the reconstructed hair envelopes (Section 5). Our experiments show that the motion of the per-frame generated hair strands are temporally coherent when they lie on the surface and their orientation is sufficiently confident. To synthesize the remaining in-between and interior hair strands, we advect their seed points using a drift-reduced version of the non-rigid registration framework of Li et al. [2009].

The final output of our pipeline is a dense set of 3D hair strands for each frame of the initial video. Though we do not claim that these strands correspond directly to the real-world hair fibers, they provide a similar visual appearance and, most importantly, capture the complex dynamics exhibited by real-world hair performances. We demonstrate results (Section 6) on a variety of hair types and hair styles, ranging from smooth to chaotic, and including nontrivial motions which are difficult to model using a classical physical simulation approach (see supplemental video).

2 Related work

Hair is a critical aspect of digital characters and it has been studied from many angles [Ward et al. 2006]. Here we describe the work most related to ours.

Static hair capture Several approaches have been proposed to capture hair [Kong et al. 1997; Grabli et al. 2002; Paris et al. 2004; Wei et al. 2005], including complex hairstyles [Paris et al. 2008] and fiber-level geometry [Jakob et al. 2009]. These techniques require the hair to be static so that several photos under different lighting conditions or with varying focus settings can be taken. Because of this, the capture process cannot be repeated fast enough to handle hair in motion. Moreover, these methods lack robustness to motion, e.g., half the capture sessions with the hair photobooth of Paris et al. did not produce usable data because of the person's motion [Paris et al. 2011], while Jakob et al.'s setup is applicable only to a ponytail attached to a rig. Nonetheless, we build upon and extend some of the components proposed for the static case. We analyze the input images using oriented filters similarly to Paris et al. [2004] and compare the extracted orientations across viewpoints to locate the strands akin to Wei et al. [2005]. But instead of a space-carving scheme, we describe a robust Markov Random Field optimization that can recover fine details from only a few views. This enables the reconstruction at video rate of the visible hair surface, including the intricate arrangement typical of strands in motion (Figure 1). We also introduce a multiscale regularization scheme that enables the reconstruction of a thick volumetric layer of hair akin to Paris et al. [2008] and Jakob et al. [2009]. We then use a dedicated tracking algorithm and a temporally consistent strand-growing scheme to convert this surface into an animated hair model.

Dynamic hair capture While significant advances have been made in static hair modeling, much less work has been done to deal with dynamic hair. Ishikawa et al. [2007] use a motion capture system on hair by placing reflective markers on a few guide strands. The motion of the rest of the hair is interpolated from these captured the guides. Although real-time capture is possible with such a system, the reflective markers alter the dynamics of the hair and limit the resolution of the captured geometry. Yamaguchi et al. [2008] introduced an image-based system to generate temporally coherent hair motion. Their method extends the static capture method of Wei et al. [2005] with a temporally coherent hair growth algorithm. The system is only demonstrated on a medium-length straight-hair wig with limited motion. It is unclear how the method generalizes to more complex scenarios. In particular, the strong smoothness assumption is at odds with the curls and flying wisps observed on long hair during a head shake (Figure 1).

3 Overview

Our system (see Figure 2) begins by capturing video sequences of hair in motion using a set of high-speed video cameras. After keying, we compute a multi-resolution orientation field for each image. We solve the multi-view stereo correspondence problem in the Markov Random Field (MRF) framework [Szeliski et al. 2008] with graph cuts [Boykov et al. 2001], using a novel energy formulation that consistently integrates correspondence information at all resolution levels.

A bilateral filtering step is employed to aggregate the per-pixel data cost according to the local structures of the orientation field, improving the confidence of matches in areas of strong directional response (usually prominent clumps of hair). To refine the depth map from MRF, we perform a sub-pixel refinement similar to [Beeler et al. 2010], followed by an extra bilateral filtering step based on the orientation field on the depth map. We then compute a non-rigid alignment between consecutive frames of reconstructed hair envelope meshes to recover frame-to-frame correspondences. These correspondences are used to perform temporal filtering, as well as to advect seed points throughout the sequence. Finally, we compute a volumetric orientation field for each frame, and grow a hair strand from each seed point, following the orientation field.

4 Hair geometry reconstruction

4.1 Acquisition

We use four AVT Pike high speed cameras to capture monochromatic video of moving hair at 100 FPS in VGA (640x480) resolution. While more advanced equipment options with higher resolution exist, we find that high speed capture capability is particularly valuable in capturing interesting hair motions and the AVT Pike provides a good trade-off between resolution and speed. The cameras are arranged in a upside down T-pose placed at roughly 90 cm distance from the subject, close enough to minimize hair occlusions while maintaining sufficient stereo accuracy to faithfully capture the intricate geometry of hair (see Figure 3). 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, the horizontally positioned three cameras will have stereo failure for the hair strands in horizontal orientation. To address this problem, a top camera is added to extend the stereo baselines and prevent the blind point of any singular orientation.

We use strong lighting (3 light sources evenly placed behind the cameras) to ensure high contrast and a short exposure of 1 ms to prevent motion blur. We use aperture F/8 for all the cameras to have sufficient depth of field covering subject's head movement.

We achieve subpixel accurate camera calibration with a standard chessboard pattern [Zhang 2000] and bundle adjustment [Hartley and Zisserman 2004]. We redo calibration before capturing each subject and the checkerboard is positioned to cover the entire head volume of the subject to optimize the accuracy for the hair capture.

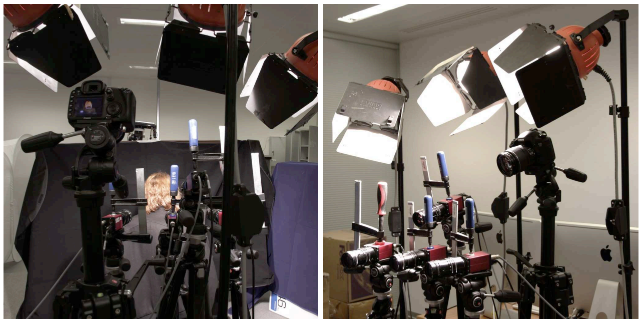

Figure 3: Acquisition setup of our dynamic hair capturing system.

4.2 Keying

Keying out the hair from the background is particularly challenging in our case because the captured images are monochromatic and hair may move quickly. We used the Roto Brush tool in Adobe After Effects CS5 to efficiently separate the hair from the background. This tool combines motion estimation with local models of the object's color and shape to produce accurate selections. The method runs at interactive rate and lets users specify corrections at any point, which are later propagated to the rest of the sequence. We refer to the original article by Bai et al. [2009] for details.

4.3 Multi-resolution 2D orientation field

Paris et al. [2004] first introduced the dense orientation field for hair modeling. We use the orientation field as the primary source of information in stereo matching because it is a distinctive and reliable feature of hair strands. Our orientation field definition differs from the prior definition in that we only consider highlighted hair strands (i.e., positive filter response), as we observed that the orientation in dark and shadowed regions is unreliable. Formally, given oriented filters K_θ, generated by rotating the 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 [Paris et al. 2004] by $\Phi(x, y) = \exp(2i\Theta(x, y))$. We also use the gamma corrected maximum response: $F(x, y) = \max_\theta |K_\theta * I(x, y)|^\gamma$ in our stereo algorithm, because it encodes the confidence of the orientation as well as the image intensity at the filter's characteristic scale. The gamma correction 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 $F(x, y)$:

$$O(x, y) = \begin{cases} F(x, y)\Phi(x, y) & K_\Theta * I(x, y) \geq 0 \\ 0 & K_\Theta * I(x, y) < 0 \end{cases} \quad (1)$$

Note that the orientation field at the region with negative maximum filter response is set to zero. We select a Difference-of-Gaussians (DoG) filter for K_0. Specifically, $K_0(x, y) = (G_\sigma(x) - G_{\sigma'}(x))\, G_{\sigma'}(y)$, where G_σ is 1D zero-mean Gaussian with standard deviation σ.

To generate the multi-resolution orientation field, we use a pyramid data structure to accelerate the computation: we recursively downsample the image for coarse levels in the pyramid and apply the oriented filter K_θ. We use a fixed sized K_θ with $\sigma = 1$ and $\sigma' = 2$ for all levels of orientation field. The multiresolution oriented pyramid is visualized in Figure 4.

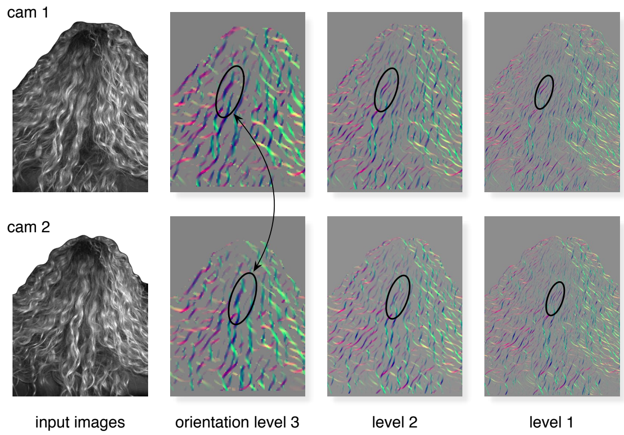

***Figure 4:** Multi-resolution orientation fields for a stereo pair of images. Color indicates the orientation angle while intensity represents the magnitude of the maximum filter response. Note the ease with which corresponding structures may be identified.*

4.4 Multi-view stereo

We reconstruct a depth map $D(p)$ for each pixel p of the center reference camera using the orientation fields computed from all four cameras. The reconstruction volume is bounded by the nearest depth d_{near} and the farthest depth d_{far} set to contain all possible hair depths in the reconstructed sequence.

Energy formulation We use the MRF 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 $\mathcal{N}(p)$:

$$\begin{aligned} E_d(D) &= \sum_p e_d(p,D) \\ E_s(D) &= \sum_p \sum_{p' \in \mathcal{N}(p)} w_s(p,p')|D(p) - D(p')|^2. \end{aligned} \tag{3}$$

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:

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

The parameter σ_o controls the constraint sensitivity and is set to $\sigma_o = 0.5$ for all our datasets.

Inspired by [Sasaki et al. 2006], we formulate the data term e_d based on the multi-resolution orientation field computed in Section 4.3. 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:

$$\begin{aligned} e_d(p,D) &= \sum_l e_d^{(l)}(p,D) \\ e_d^{(l)}(p,D) &= \sum_i c\big(O_{\text{ref}}^{(l)}(p), O_i^{(l)}(P_i(p,D))\big), \end{aligned} \tag{5}$$

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

$$c(O,O') = -\Re\{O^* O'\}, \tag{6}$$

where $\Re(z)$ denotes the real part of a complex number z. Intuitively, $c(O,O')$ measures the deviation of the orientation fields for corresponding pixels as the inverse correlation of the two orientation vectors O and O'.

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 $[d_{near}, d_{far}]$ (see illustration in Figure 5).

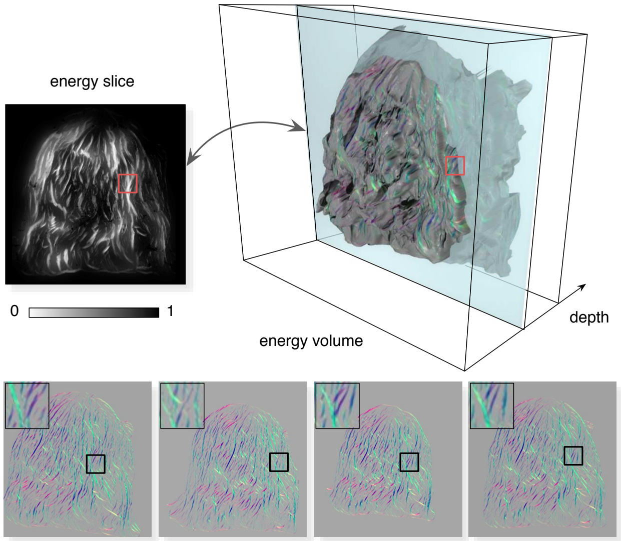

***Figure 5:** The energy volume defined by the data term $e_d(p,D)$ integrates the orientation fields of the four camera views.*

Bilateral aggregation To make the data energy more robust and adaptive to the local structure of the orientation field, we perform bilateral filtering on the data term energy on each level l based on the orientation field of the reference view on that level. The bilateral filter weights $w^{(l)}$ are computed as

$$w^{(l)}(p,p') = \exp\left(-\frac{|O_{\text{ref}}^{(l)}(p) - O_{\text{ref}}^{(l)}(p')|^2}{2\sigma_d^2} - \frac{\|p-p'\|^2}{2\sigma_p^2}\right), \tag{7}$$

where the parameters σ_d and σ_p control the aggregation by orientation similarity and proximity, respectively. The data energy $e_d^{(l)}$ in Equation 5 are aggregated as:

$$e_d^{(l)}(p,D) \leftarrow \frac{1}{Z} \sum_{p' \in \mathcal{K}(p)} w^{(l)}(p,p') e_d^{(l)}(p',D), \tag{8}$$

where $\mathcal{K}$ is a window centered at p with a size adaptive to σ_p and $Z = \sum_{p' \in \mathcal{K}(p)} w(p,p')$ is the normalization factor. Figure 12 illustrates the effect of our bilateral aggregation approach. The resulting energy in Equation (2) can be efficiently minimized by graph cuts [Boykov et al. 2001].

Depth map refinement We employ a similar sub-pixel refinement technique as [Beeler et al. 2010] 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 new depth $D'(p)$ is updated from $D(p)$ by:

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

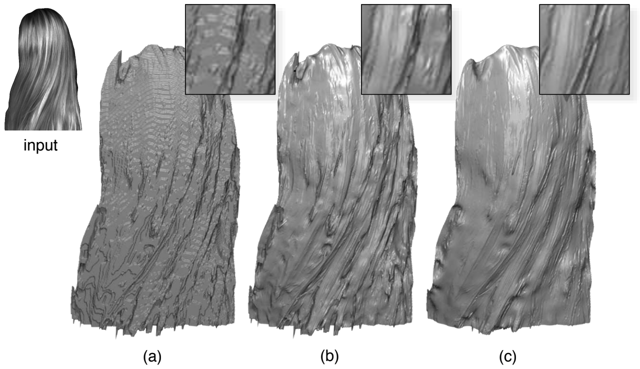

Figure 6: *The stages of depth map refinement improve reconstructed quality. The surface reconstructed from the initial MRF-optimized depth map (a) shows quantization artifacts that are removed by sub-pixel refinement (b). A post-reconstruction bilateral filtering step further improves quality (c).*

We then apply bilateral filtering once again on the depth map to further reduce the stereo noise with the same weights as in Equation 7:

$$D(p) \leftarrow \frac{1}{Z} \sum_{p' \in \mathcal{K}(p)} w(p,p') D(p') \tag{10}$$

Figure 6 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 bilateral filtering step increases surface quality modestly, but is not a replacement for pre-reconstruction bilateral aggregation.

Temporal De-noising Acquisition noise and inaccuracies in the per-frame stereo reconstruction can lead to high-frequency temporal artifacts in the captured hair envelope sequence. A common filtering strategy for static acquisition is to average a set of overlapping scans in order the reduce noise while preserving salient geometric features. Applying such a filtering method in our dynamic setting, i.e., averaging over multiple frames of the temporal sequence, requires unwarping the scans to compensate for the deformation between frames. For this purpose, we use the graph-based non-rigid registration algorithm described in Li et al. [2009] to perform coarse level deformations between partial scans of adjacent frames. Our goal here is to exploit the spatial and temporal coherence of motion at the level of salient geometric features such as curls and wisps of hair to reduce acquisition artifacts. We do not attempt to track individual hair strands over time, but rather estimate a smooth volumetric deformation field that captures the medium scale motion. Each envelope mesh is embedded in a *deformation graph* (see [Sumner et al. 2007] for details) with nodes that are uniformly sampled (15 mm spacing) on the captured hair geometry. The algorithm iteratively computes the optimal warp of the deformation graph by minimizing point-to-point and point-to-plane distances between closest points of the two scans and maximizing the local smoothness and rigidity in the node transformations. The spatial warps defined by the nodes are then transferred to the envelope mesh vertices through simple linear blend skinning using weights that are inversely proportional to Euclidean distance [Li et al. 2008]. Figure 7 illustrates the alignment of the hair envelope meshes.

We use the same parameters as in [Li et al. 2009]. As demonstrated in their work, the combination of deformation regularization and a low number of degrees of freedom enables accurate alignments between geometries affected by high-frequency noise. Hence, temporal de-noising can effectively separate noise from salient features by simply averaging the depth maps of unwarped scans. This separation recovers the medium-scale motion in a temporally coherent fashion, which is crucial for an accurate reconstruction of visually important features of the hair geometry. Figure 8 illustrates the effect of our temporal de-noising approach.

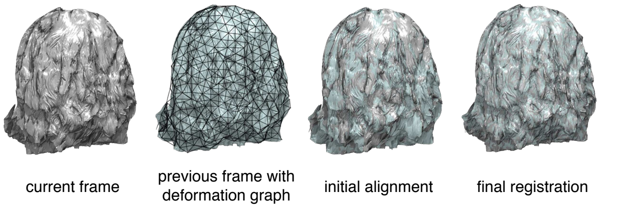

Figure 7: *Non rigid registration of consecutive frames based on embedded deformation for temporal de-noising.*

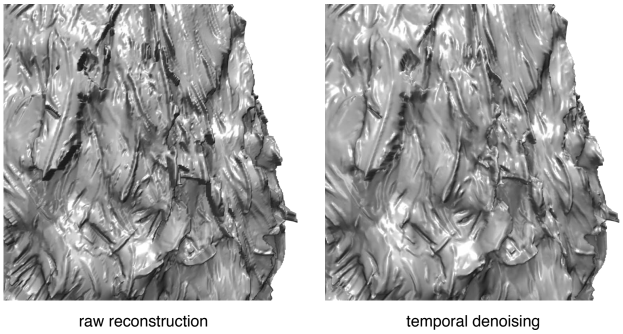

Figure 8: *Temporal filtering reduces acquisition noise while preserving the salient features of the acquired hair.*

5 Hair synthesis

We grow hair strands following a volumetric 3D orientation field, similar to the approach of Paris et al. [2008]. However, since we allow the captured subjects to move their heads freely for natural hair movements, it is difficult to reliably fit a scalp for the performing subject and use it as a prior for hair generation, as was done by previous methods. Therefore, we adopt a seed-and-grow scheme that employs two different sampling strategies for hair synthesis.

To properly recover the hair strands reconstructed from our system, we densely sample seeds on the high confidence regions of the envelope surface where the prominent hair strands are located. This seeding is followed by hair growth step that generates hair strands from the seeds according to the orientation field. In addition, we also sample seeds on an inward offset surface of the reconstructed hair envelope to fill the interior of the hair volume with strands. We show that this seed-and-grow scheme captures more fine-scale structures on the hair surface and avoids hair sparsity problems throughout the hair volume (see Figures 9 and 11).

5.1 3D orientation field

3D orientation computation We first compute the 3D orientation field on the envelope surface by combining the projected 2D orientation on the reference view and the surface normal. For every point p on the hair surface, we project back to the 2D orientation field of the reference view, in which we sample the 2D line l representing the 2D orientation of the projected point. Back-projecting l in space forms a plane π with plane normal n_l. The 3D orientation ℓ of p is computed as $\ell = n_l \times n_p$, where n_p is the normal of p on the hair surface. The magnitude of the 3D orientation $|\ell|$ is inherited from the 2D orientation, i.e., $|\ell| = F_{\text{ref}}(q)$ with F_{ref} the response amplitude of the orientation detector on the reference view. Note that because we use a relatively small stereo baseline, the surface

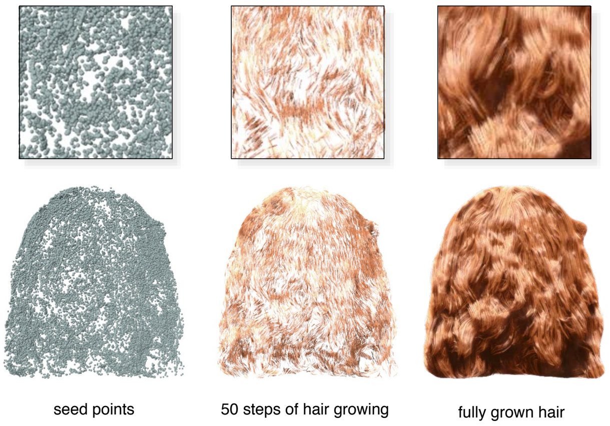

Figure 9: *Our seed-and-grow algorithm incrementally synthesizes hair strands from a temporally coherent set of seed points.*

normals provide more reliable information about the 3D orientation than the projected 2D orientation in other views.

Temporal filtering To ensure temporal coherence of the 3D orientation field, we perform temporal filtering on the 3D orientation field, using bilateral temporal filtering with the inter-frame correspondences computed in the temporal de-noising step (Section 4.4). For a point p_t at frame t, the bilateral filtering weights for the corresponding points p_{t-1} and p_{t+1} at frame $t-1$ and $t+1$ are computed as:

$$\begin{aligned} w_i &= 0.25\exp\left(-\frac{\|p_t - p_i\|^2}{2\sigma_t^2}\right), i = t-1, t+1 \\ w_t &= 1 - w_{t-1} - w_{t+1}, \end{aligned} \tag{11}$$

where σ_t controls the weight according to spatial proximity. We perform multiple passes of temporal filtering with these weights. For 3D orientation filtering, we use the structure tensor $T = \ell\ell^T$ defined in [Paris et al. 2008]. At each filtering step, the orientation T_t at frame t are updated as: $T_t \leftarrow T_t + w_{t-1}T_{t-1} + w_{t+1}T_{t+1}$. To ensure the orientation remains tangent to the hair surface, the new orientation ℓ_t extracted from T_t is projected to the tangent plane at p_t: $\ell_t \leftarrow \ell_t - (\ell_t \cdot n_p)n_p$, where n_p is the normal at p_t. We find that five iterations of orientation filtering yield satisfactory results for all the datasets.

Volumetric diffusion We propagate the 3D orientation field from the surface into a volumetric sampling of the hair volume by performing diffusion on the structure tensor. Note that because the magnitude of the 3D orientation ℓ encodes orientation confidence information, the structure tensor $T = \ell\ell^T$ is diffused with confidence.

5.2 Seeding

Hair strands are grown according to the 3D orientation field, beginning at 50-100k *seed points* on and beneath the hair surface. We distinguish between two types of seeds:

- **Surface seeds** are sampled randomly in high-confidence regions of the surface mesh (i.e., those having a strong oriented filter response — the highly-saturated regions in Figure 4). Strands grown according to these seeds faithfully reproduce the detail of actual groups of strands that we were able to capture. Because these are high-confidence regions, we have observed that they are relatively consistent from frame to frame, and we make no special effort to ensure temporal coherence for these seeds.
- **Interior seeds** are sampled within a volumetric hair layer between the hair surface and an inward offset surface. Hair strands grown from these seeds complement those grown from surface strands and fill in the hair volume, but they were not captured directly: they are inferred. For this reason, we have found it necessary to explicitly ensure temporal cnoherence for these seeds.

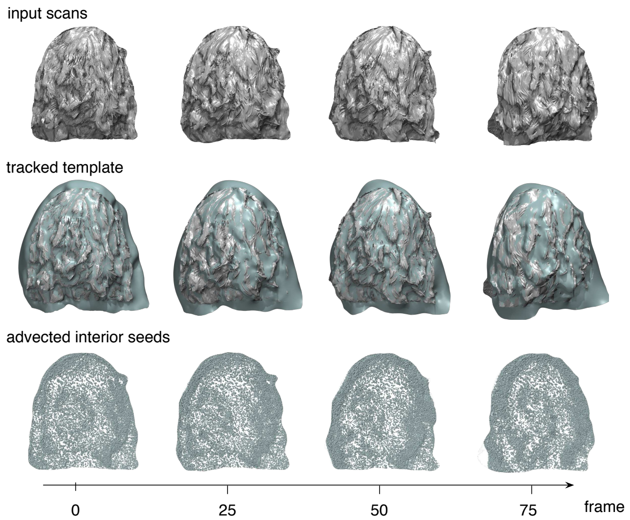

Figure 10: *We ensure smooth motion of interior seeds by advecting them using motion obtained from template-based tracking.*

5.3 Interior seed advection

Non-rigid surface tracking The interior seeds are not regenerated at every frame; they are generated once and then advected using the approximate motion of the hair surface, found using a template-based approach. Because the full hair surface is not observed on any individual frame, and indeed new parts of the surface are exposed throughout the performance, we use a tracking proxy ("template") that covers a larger area than the frames.

We initialize the template on the first frame by first building a planar surface with a somewhat larger extent than the visible portion of the hair, then warping this plane to conform to the surface mesh. The template mesh thus smoothly extends the partial hair surface visible on the first frame.

Next, we repeatedly warp the template to the hair surface on each frame. The warping uses two deformation models similar to the geometry and motion reconstruction scheme used in [Li et al. 2009]. A graph-based deformation model is used to capture the coarse level deformations (similar to the temporal de-noising stage) and a linear per-vertex deformation model is used to adapt the fine-scale details of the template to the current frame by displacing the vertices in normal direction. As opposed to the original formulation, the coarse level deformation is not reinitialized for each frame, but only once. The regularization weights (smoothness and local rigidity maximization) are kept constant during the entire reconstruction. In this way, we reduce the amount of drift for extended frame lengths. While the method becomes less robustness to large deformation since we disable the use of a regularization relaxation scheme as in [Li et al. 2009], we found that the deformations between consecutive frames of high-speed acquisitions were sufficiently small to be handled by multiple Gauss-Newton iterations. Another difference with the original method of [Li et al. 2009] is that detail synthesis using Laplacian deformations is no longer performed as a second step, but directly after the graph-based registration for each frame.

Interior seed generation We sample interior seeds on the first frame within a volume bounded by the (extended) first-frame template and an inner offset surface. We find the latter by computing a voxelized representation of the template's signed distance field and

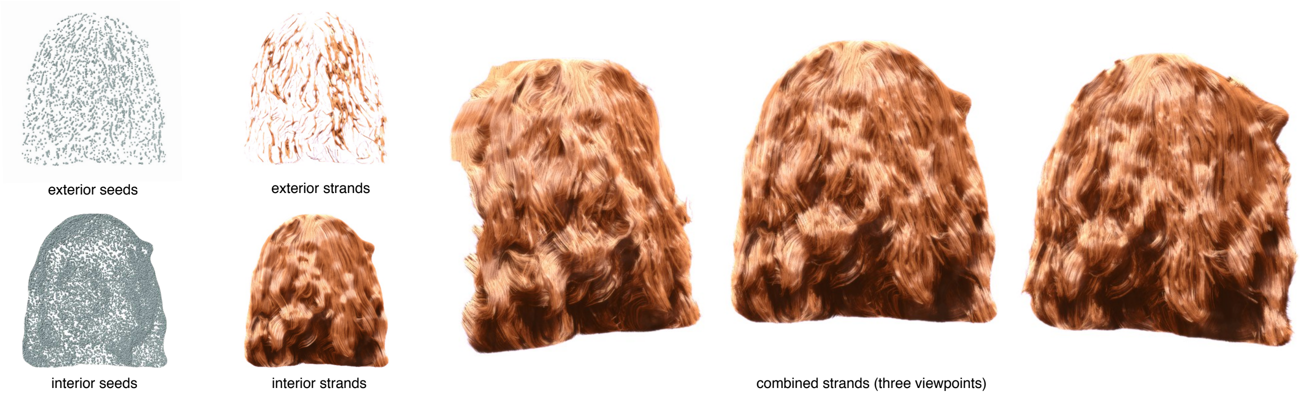

Figure 11: *The combination of interior and exterior strands is crucial for capturing the complex geometry of the most prominent hair strands while ensuring sufficient hair density in the entire volume as shown for different views on the right.*

extracting an isosurface. We have observed good results with a hair volume thickness of 2 to 3 cm, depending on the performance.

Advection We advect interior seeds via linear-blend skinning. We first uniformly resample the tracked template mesh to form nodes of a deformation cage. The displacement of each interior seed is then obtained as a linear combination of the $k = 8$ closest nodes, weighted proportionally to the inverse of the distance to the node [Li et al. 2008]. We adjust the resolution of the cage to minimize noise while maintaining motion detail: a coarser-resolution cage yields smoother motion, while a higher-resolution one matches the surface motion more closely. Figure 10 illustrates template tracking and interior seed advection.

Trimming The final step is to eliminate seeds that lie outside the visual hull at each frame. This is necessary because of our use of an extended template, which is noisy in parts of the surface that were not observed. We believe that with a camera setup having greater angular coverage, we would be able to have a full-head template observed completely at each frame, eliminating the need for this step.

5.4 Hair growing

Beginning with each seed, we grow hairs following the 3D orientation field using a simple marching method, i.e., forward Euler. This is essentially the hair strand growing algorithm of Paris et al. [2004; 2008], except that we grow the strand from the seeds inside the hair layer, in both directions of the 3D orientation, instead of growing outward from the scalp. The growth is terminated if any of the following conditions applies:

1. The strand exceeds a pre-defined length.
2. The strand grows out of the hair layer.
3. The strand grows to a new direction d', and the angle ψ between d' and the original direction d is larger than a threshold T_ψ.

Condition 3 prevents the creation of unrealistic sharp turns due to errors in our orientation measures.

6 Evaluation and discussion

To demonstrate our method we processed three hair performances captured from different subjects. The acquired hair styles vary from straight and orderly to curly and wavy and all involve significant motion. The statistics of the input and output data are shown in the table below (see also Figure 16). Note that the complexity of the reconstructed hair geometry determines the number of strands.

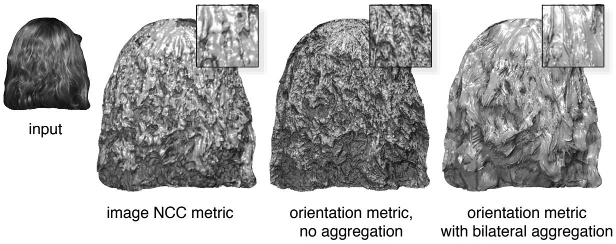

Figure 12: *The quality of our mesh reconstructions (right) is due to many design choices in our pipeline, including the use of orientation-based matching (as opposed to color-based matching, left) and bilateral aggregation (without which noise is significantly more evident, center).*

Dataset	#Frames	Style	Color	#Strands
Hair1	300	Straight	Blonde	80k
Hair2	130	Messy	Red	60k
Hair3	200	Curly	Blonde	120k

Our whole pipeline takes roughly 11 minutes of computation time per frame on a modern machine. Per-frame processing times are subdivided as follows: multi-resolution orientation field generation takes 30 seconds, multi-view stereo takes 5 minutes, temporal denoising adds 3 minutes, non-rigid surface tracking and seed advection takes 1 minute, hair growing for typical 80k strands in a 200x100x100 grid takes 2 minutes, hair rendering based on the hair shading model by Marschner et al. [2003] takes 10 seconds.

Orientation-based metric To evaluate the orientation-based metric for the reconstruction of the hair envelope surfaces, we also implemented our reconstruction pipeline with a conventional color-based metric, i.e., normalized cross correlation (NCC). Figure 12 shows the reconstruction results of the two approaches. While both reconstruct the rough geometry correctly, the orientation-based implementation resolves much finer and more detailed hair geometry.

Reconstruction evaluation In Figure 13 we evaluate the accuracy of our reconstructed hair envelope surface by applying our approach to a set of synthetic hair geometries rendered using the shading model of [Marschner et al. 2003]. Most of the envelope lies within 1cm of the nearest strands in the reference camera view. As the cut-out on the right illustrates, our reconstructed hair envelope optimally fits the intersecting points of the hair strands.

We also compare the quality of our reconstruction with the state-of-the-art multi-view stereo algorithms [Beeler et al. 2010] and [Furukawa and Ponce 2010] in Figure 14. Our method reveals

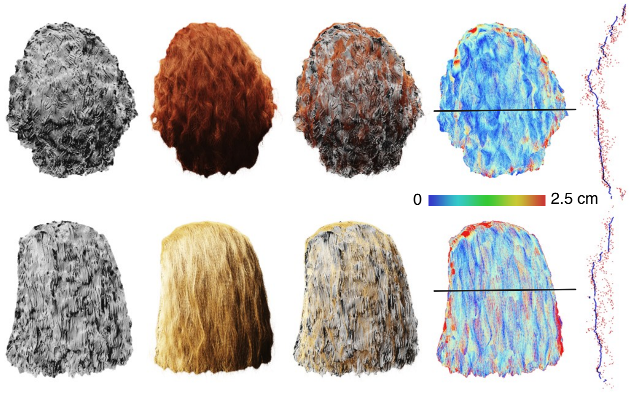

Figure 13: Reconstruction evaluation with two synthetic hair geometries. (a) our reconstructed hair envelope, (b) ground truth datasets, (c) overlay of both geometries, (d) depth difference, (e) cut-out depth map on a intersecting horizontal plane (indicated by black lines).

more hair geometry detail than the other two methods, thanks to the orientation-based metric and bilateral aggregation that imposes strong smoothness prior along the hair strands. In contrast, [Beeler et al. 2010] is designed for face geometry and assumes intensity invariance across views and has difficulties to adapt to intricate hair geometry and specular appearance. Both [Beeler et al. 2010] and [Furukawa and Ponce 2010] employ a patch-based matching metric, e.g., NCC for stereo reconstruction, and thus tend to blur the hair strand geometry detail within the matching window.

Number of cameras Our multi-view stereo system employs four cameras for reconstruction. The redundant information provided by the 4-camera system improves the overall accuracy and robustness of the reconstruction. We compare the results of 2-camera system and 4-camera system in Figure 15. As the results show using 4 cameras provides significant improvements in reconstruction quality and accuracy over the 2-camera system.

Limitations As can be seen in the accompanying video, our approach successfully captures hair in motion and we believe that it provides a useful complement to traditional modeling of hair animation by artists. However, as any technique dealing with complex real-world data, there are limits to what our method can cope with. Our system is not intended to recover individual real-world hair fibers, but rather to provide a set of strands that follow the orientation and motion of the real-world performance. With this (already limited) goal in mind, we have observed the following limitations on our results:

- Reconstruction error. In the results we do notice that a reconstruction error emerges consistently during the sequence due to the ambiguity of matching strands and less accuracy in the areas near grazing angles.
- Temporal incoherence at the fiber level. Although coherent seed point propagation and temporal filtering to the orientation field and hair envelope geometry improves the overall temporal coherence, we cannot ensure temporal coherence on the hair fiber level and we can still observe jittering in the sequence with highly dynamic wavy and curly hair patterns.
- Hair fiber topology. We currently use a hair synthesis method to generate hair fibers from sampled seeds within the hair volume, and it is not guaranteed to reflect the true orientation and connectivity of the captured hair strands. In particular, we do not generate full hair strands from the scalp.

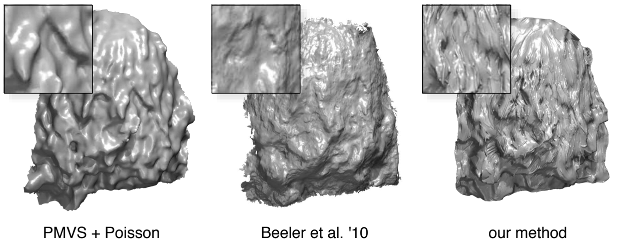

Figure 14: Comparison between our method and other state-of-the-art multi-view stereo methods: PMVS [Furukawa and Ponce 2010] + poisson surface reconstruction [Kazhdan et al. 2006] and [Beeler et al. 2010]

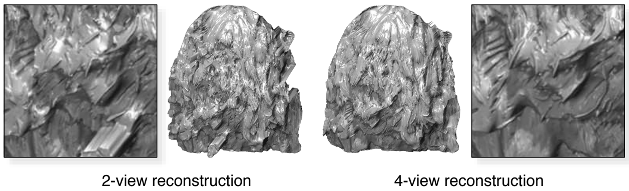

Figure 15: Comparison between 2-camera and 4-camera system.

Conclusion We have introduced a comprehensive system for acquisition and reconstruction of dynamic human hair performances, demonstrating that multi-resolution orientation filters enable multi-view stereo reconstruction of hair, which in turn allows for the synthesis of complex hair in motion. We see our approach as a first important step towards the goal of obtaining hole-free and error-free reconstructions of a complete moving hairstyle, a goal that will require substantial future investigations due to the immense complexity of the geometry, appearance, and motion of human hair.

Beyond improving the accuracy and efficiency of our system, and addressing the limitations discussed above, we envisage other interesting future work at the interface between acquisition and simulation. For example, our data could be used to drive a detailed hair simulation with realistic constraints and initial conditions, while the simulation would provide smoothness priors, allow the user to make edits, and most importantly fill in portions of the hair animation that cannot be observed in a given capture session. Integrating our framework with other performance capture systems for face or body acquisition is another interesting avenue for future work.

Acknowledgements We would like to thank Duygu Ceylan, Dorothée Baumann, and Anastasia Ailamaki for the hair performances. We also thank Thabo Beeler and his coworkers for providing the comparison with their multi-view stereo system and Cem Yuksel for the synthetic hair data for ground truth evaluation.

This work is supported by NSF grant CCF-1012147, SNF fellowship PBEZP2-134464, and SNF grant 20PA21L_129607.

References

BAI, X., WANG, J., SIMONS, D., AND SAPIRO, G. 2009. Video SnapCut: Robust video object cutout using localized classifiers. *ACM Trans. Graph. 28*, 3 (August), 70:1–70:11.

BEELER, T., BICKEL, B., BEARDSLEY, P., SUMNER, B., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graph. 29*, 4 (July), 40:1–40:9.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *PAMI 23*, 11 (November), 1222–1239.

DE AGUIAR, E., STOLL, C., THEOBALT, C., AHMED, N., SEIDEL, H.-P., AND THRUN, S. 2008. Performance capture from

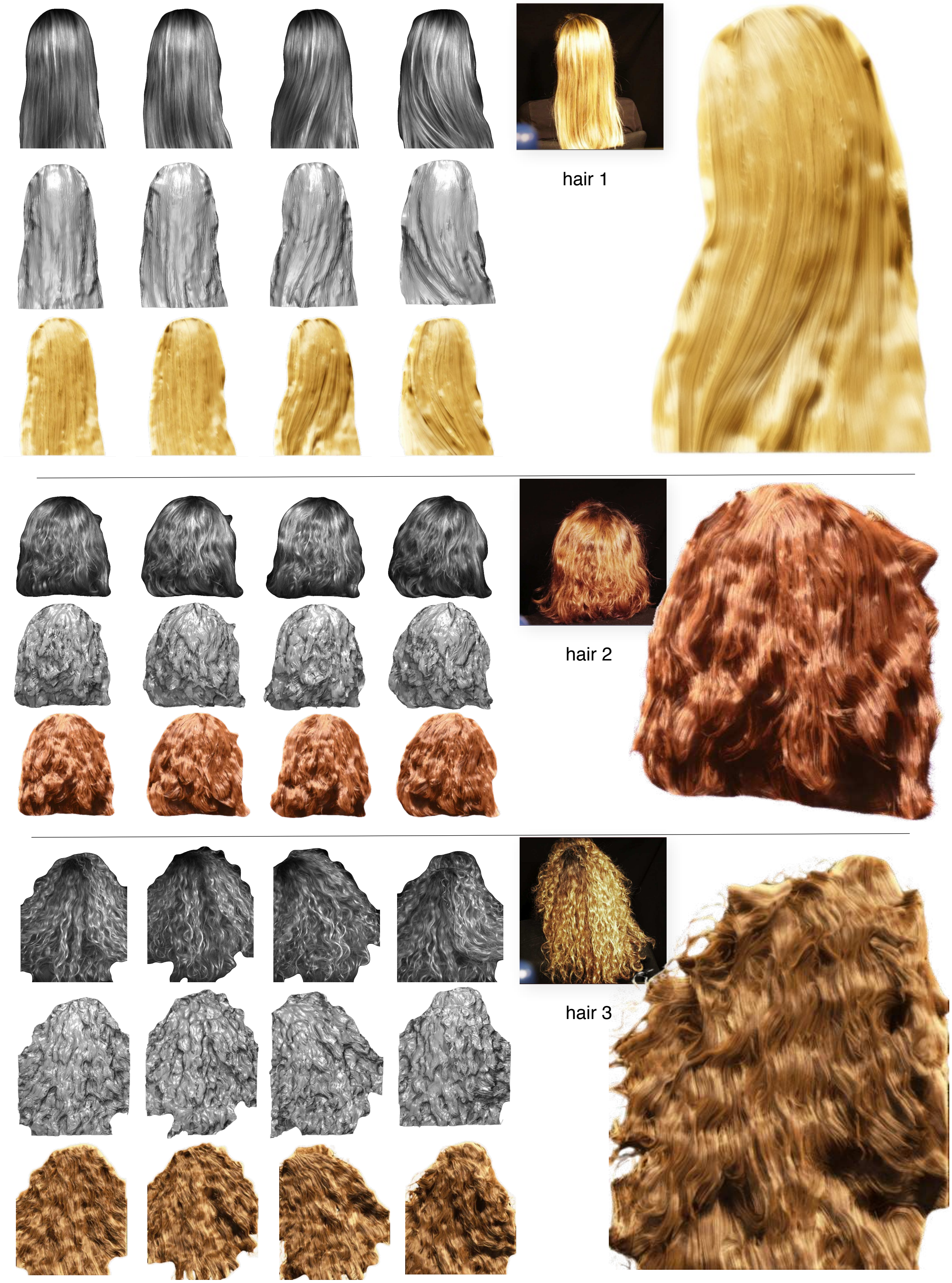

Figure 16: *Results for three datasets reconstructed using our pipeline. For each, we show four frames of photographs from the hair-capture session, still frames from the video sequence used for reconstruction, the corresponding reconstructed mesh, rendering of the synthesized hair and a close-up of the envelop (middle row) or of the final rendering (top and bottom rows).*

sparse multi-view video. *ACM Trans. Graph. 27*, 3 (July), 98:1–98:10.

FURUKAWA, Y., AND PONCE, J. 2010. Accurate, dense, and robust multiview stereopsis. *PAMI 32*, 1362–1376.

GRABLI, S., SILLION, F., MARSCHNER, S. R., AND LENGYEL, J. E. 2002. Image-based hair capture by inverse lighting. In *Proceedings of Graphics Interface*.

HARTLEY, R. I., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press.

ISHIKAWA, T., KAZAMA, Y., SUGISAKI, E., AND MORISHIMA, S. 2007. Hair motion reconstruction using motion capture system. In *ACM SIGGRAPH posters*.

JAKOB, W., MOON, J. T., AND MARSCHNER, S. 2009. Capturing hair assemblies fiber by fiber. *ACM Trans. Graph. 28*, 5 (December), 164:1–164:9.

KAZHDAN, M., BOLITHO, M., AND HOPPE, H. 2006. Poisson surface reconstruction. In *Proceedings of SGP*.

KONG, W., TAKAHASHI, H., AND NAKAJIMA, M. 1997. Generation of 3D hair model from multiple pictures. In *Proceedings of Multimedia Modeling*.

LI, H., SUMNER, R. W., AND PAULY, M. 2008. Global correspondence optimization for non-rigid registration of depth scans. In *Proceedings of SGP*.

LI, H., ADAMS, B., GUIBAS, L. J., AND PAULY, M. 2009. Robust single-view geometry and motion reconstruction. *ACM Trans. Graph. 28*, 5 (December), 175:1–175:10.

MARSCHNER, S., JENSEN, H. W., ANDS. WORLEY, M. C., AND HANRAHAN, P. 2003. Light scattering from human hair fibers. *ACM Trans. Graph. 22*, 3 (July), 780–791.

MCADAMS, A., SELLE, A., WARD, K., SIFAKIS, E., AND TERAN, J. 2009. Detail preserving continuum simulation of straight hair. *ACM Trans. Graph. 28*, 3 (July), 62:1–62:6.

MIHASHI, T., TEMPELAAR-LIETZ, C., AND BORSHUKOV, G. 2003. Generating realistic human hair for *The Matrix Reloaded*. In *ACM SIGGRAPH Sketches and Applications Program*.

PARIS ET AL., 2011. Personal communication with the authors of [Paris et al. 2008].

PARIS, S., BRICEÑO, H., AND SILLION, F. 2004. Capture of hair geometry from multiple images. *ACM Trans. Graph. 23*, 3 (August), 712–719.

PARIS, S., CHANG, W., KOZHUSHNYAN, O. I., JAROSZ, W., MATUSIK, W., ZWICKER, M., AND DURAND, F. 2008. Hair Photobooth: Geometric and photometric acquisition of real hairstyles. *ACM Trans. Graph. 27*, 3 (August), 30:1–30:9.

SASAKI, K., KAMEDA, S., ANDO, H., SASAKI, M., AND IWATA, A. 2006. Stereo matching algorithm using a weighted average of costs aggregated by various window sizes. In *Proceedings of ACCV*.

SELLE, A., LENTINE, M., AND FEDKIW, R. 2008. A mass spring model for hair simulation. *ACM Trans. Graph. 27*, 3 (August), 64:1–64:11.

SUMNER, R. W., SCHMID, J., AND PAULY, M. 2007. Embedded deformation for shape manipulation. *ACM Trans. Graph. 26*, 3 (July), 80:1–80:7.

SZELISKI, R., ZABIH, R., SCHARSTEIN, D., VEKSLER, O., KOLMOGOROV, V., AGARWALA, A., TAPPEN, M., AND ROTHER, C. 2008. A comparative study of energy minimization methods for Markov random fields with smoothness-based priors. *PAMI 30*, 6 (June), 1068–1080.

VLASIC, D., BARAN, I., MATUSIK, W., AND POPOVIĆ, J. 2008. Articulated mesh animation from multi-view silhouettes. *ACM Trans. Graph. 27*, 3 (August), 97:1–97:9.

VLASIC, D., PEERS, P., BARAN, I., DEBEVEC, P., POPOVIĆ, J., RUSINKIEWICZ, S., AND MATUSIK, W. 2009. Dynamic shape capture using multi-view photometric stereo. *ACM Trans. Graph. 28*, 5 (December), 174:1–174:11.

WARD, K., BERTAILS, F., YONG KIM, T., MARSCHNER, S. R., PAULE CANI, M., AND LIN, M. C. 2006. A survey on hair modeling: Styling, simulation, and rendering. *TVCG 13*, 2, 213–234.

WARD, K., SIMMONS, M., MILNE, A., YOSUMI, H., AND ZHAO, X. 2010. Simulating Rapunzel's hair in Disney's *Tangled*. In *ACM SIGGRAPH Talks*.

WEI, Y., OFEK, E., QUAN, L., AND SHUM, H.-Y. 2005. Modeling hair from multiple views. *ACM Trans. Graph. 24*, 3 (July), 816–820.

YAMAGUCHI, T., WILBURN, B., AND OFEK, E. 2008. Video-based modeling of dynamic hair. In *Proceedings of PSIVT*.

YOON, K.-J., AND KWEON, I. S. 2006. Adaptive support-weight approach for correspondence search. *PAMI 28*, 4 (April), 650–656.

ZHANG, Z. 2000. A flexible new technique for camera calibration. *PAMI 22*, 11 (November), 1330–1334.