Facial Hair Tracking for High Fidelity Performance Capture

SEBASTIAN WINBERG, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland
GASPARD ZOSS, DisneyResearch|Studios, Switzerland
PRASHANTH CHANDRAN, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland
PAULO GOTARDO, DisneyResearch|Studios, Switzerland
DEREK BRADLEY, DisneyResearch|Studios, Switzerland

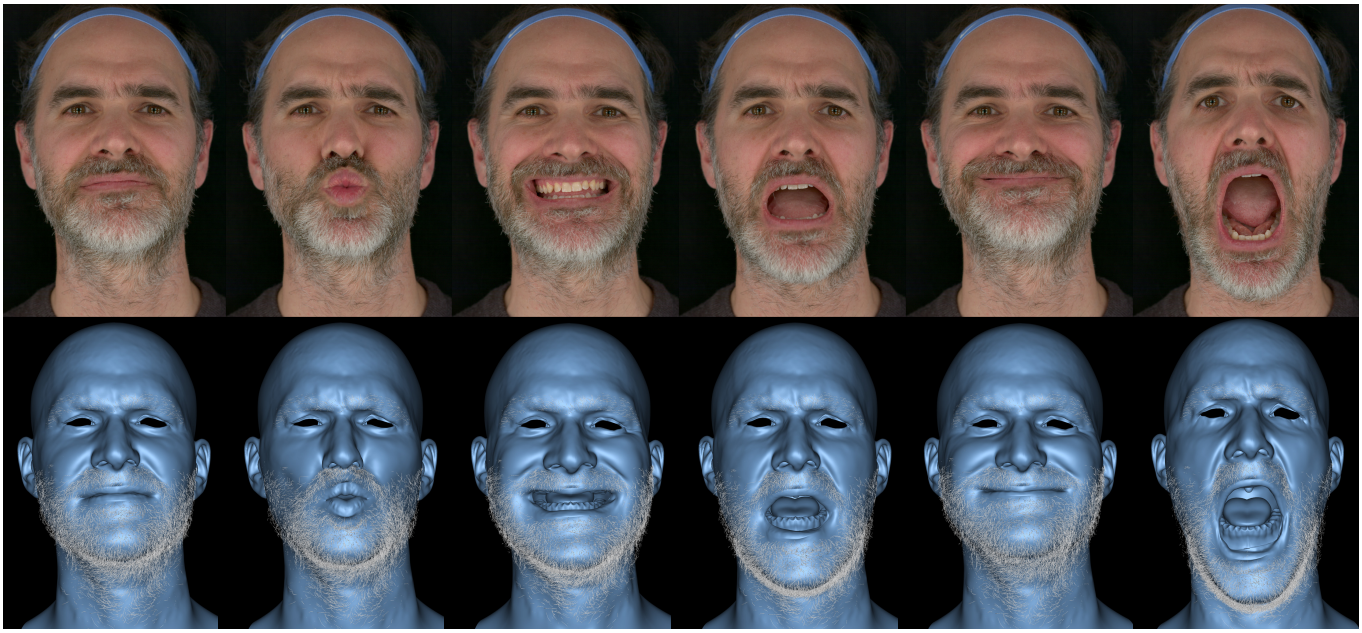

Fig. 1. We present the first method for reconstruction and tracking of dense facial hair along with the underlying skin surface for performance capture.

Facial hair is a largely overlooked topic in facial performance capture. Most production pipelines in the entertainment industry do not have a way to automatically capture facial hair or track the skin underneath it. Thus, actors are asked to shave clean before face capture, which is very often undesirable. Capturing the geometry of individual facial hairs is very challenging, and their presence makes it harder to capture the deforming shape of the underlying skin surface. Some attempts have already been made at automating this task, but only for static faces with relatively sparse 3D hair reconstructions. In particular, current methods lack the temporal correspondence needed when capturing a sequence of video frames depicting facial performance. The problem of robustly tracking the skin underneath also remains unaddressed. In this paper, we propose the first multiview reconstruction pipeline that tracks both the dense 3D facial hair, as well as the underlying 3D skin for entire performances. Our method operates with standard setups for face photogrammetry, without requiring dense camera arrays. For a given capture subject, our algorithm first reconstructs a dense, high-quality neutral 3D facial hairstyle by registering sparser hair reconstructions over multiple frames that depict a neutral face under quasi-rigid motion. This custom-built, reference facial hairstyle is then tracked throughout a variety of changing facial expressions in a captured performance, and the result is used to constrain the tracking of the 3D skin surface underneath. We demonstrate the proposed capture pipeline on a variety of different facial hairstyles and lengths, ranging from sparse and short to dense full-beards.

CCS Concepts: • **Computing methodologies** → **Motion capture**; **Motion processing**.

Additional Key Words and Phrases: Facial Hair Capture, Performance Capture, Hair Tracking.

ACM Reference Format:
Sebastian Winberg, Gaspard Zoss, Prashanth Chandran, Paulo Gotardo, and Derek Bradley. 2022. Facial Hair Tracking for High Fidelity Performance Capture. *ACM Trans. Graph.* 41, 4, Article 165 (July 2022), 12 pages. https://doi.org/10.1145/3528223.3530116

Authors' addresses: Sebastian Winberg, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland, winbergs@student.ethz.ch; Gaspard Zoss, DisneyResearch|Studios, Switzerland, gaspard.zoss@disneyresearch.com; Prashanth Chandran, ETH Zurich, Switzerland and DisneyResearch|Studios, Switzerland, prashanth.chandran@disneyresearch.com; Paulo Gotardo, DisneyResearch|Studios, Switzerland, paulo.gotardo@disneyresearch.com; Derek Bradley, DisneyResearch|Studios, Switzerland, derek.bradley@disneyresearch.com.

0730-0301/2022/7-ART165 $15.00
https://doi.org/10.1145/3528223.3530116

1 INTRODUCTION

The creation of realistic animated performances for digital characters has relied for decades on motion capture of real actors, in particular for high-quality visual effects in feature films. More and more, actors are performing in front of multi-camera reconstruction systems that accurately recover the actors' movements digitally, which is especially popular for facial animation. Such facial performances can then be used for face replacement, digital enhancement, aging or de-aging, or in order to drive fantasy characters.

While the technology for 3D reconstruction and tracking of the facial skin surface has advanced immensely, a major limitation is that current facial performance tracking methods do not operate well in the presence of facial hair. As multiview stereo reconstruction has been predominantly targeted at 3D surfaces, the presence of a multitude of small and thin structures such as facial hair, with a complex mutual occlusion pattern, most often results in largely incorrect, shrink-wrapped (water-tight) 3D surfaces. Therefore, actors are consistently asked to shave their beards and mustaches just before their facial performances are captured. This can pose many problems, ultimately increasing the cost and complexity of the capture pipeline. For example, it might be that the digital character is in fact *meant* to have the facial hair of the actor, in which case the clean-shaved skin of the actor is first reconstructed in 3D, and then artists must digitally re-add the facial hair on top of the skin geometry — which is obviously inconvenient and time consuming (expensive). An additional issue that arises very often is that many actors are simply unwilling to shave, as it causes a large and undesired change in their appearance. Shaving can also be a huge commitment, as it can take multiple months until facial hair grows to reach the same length and density again. Thus, capturing an actor in a role with a beard or mustache also imposes scheduling challenges, as all principal photography shooting must be done before the actor shaves for facial capture work, or only several months after face capture, leaving enough time for the beard to re-grow. Furthermore, eyebrows exhibit similar challenges as beards and mustaches, so a universal solution for facial hair capture would be very valuable.

In this work, we present the first method (to our knowledge) that enables facial performance capture in the presence of facial hair (*e.g.* beards, mustaches and eyebrows). Our method focuses on detailed 3D facial hair reconstruction and tracking over time, yielding a dense digital 3D facial hairstyle and its deformation over an entire performance. The resulting hair geometry sequence is in full correspondence over time and can be easily edited by artists. Our method also estimates a time-varying face surface under the hair, in order to provide a plausible clean-shaven facial performance reconstruction. Combining the face and the hair captured by our method, a high fidelity digital double can be obtained. Applications include reconstructing faces with the facial hair for more faithful digital actors, or focusing on the underlying facial surface and discarding the recovered hair, to model clean-shaven digital doubles without requiring the actor to shave.

2 RELATED WORK

This section reviews related work on facial performance capture and hair capture, focusing on facial and scalp hair.

Face Capture. The last decades have witnessed significant progress on research focused on capturing the 3D geometry of unoccluded facial skin, going beyond static faces to also capture facial deformation over time in a markerless way. In the domain of static reconstruction, there have been both active lighting methods [Ghosh et al. 2011; Ma et al. 2007], as well as passive, stereo approaches [Beeler et al. 2010] that can capture fine geometric detail at the level of skin pores, with sub-millimeter accuracy. In the area of dynamic performance capture, multiview techniques [Beeler et al. 2011; Bradley et al. 2010; Fyffe et al. 2017] allow to capture high-quality sequences of 3D facial geometry with dense temporal correspondences in a markerless way. Recently, Wu et al. [2016b] have also enabled accurate tracking of 3D faces for performance capture using few camera views, by leveraging stronger face priors in the form of an anatomical local model of facial skin deformation. More recently, Deep Learning based methods [Laine et al. 2017; Li et al. 2021; Tewari et al. 2017] have helped lower the processing time. In line with tracking the skin surface, there have also been works that focus on capturing and tracking secondary facial features like eyes [Bérard et al. 2016], eyelids [Bermano et al. 2015], teeth [Wu et al. 2016a], as well as the skull and jaw [Zoss et al. 2019]. These works all contribute to the goal of automatically capturing a complete human face with all its inherent details. Most of these works assume that skin patches are mostly hair free; when this assumption does not hold, the resulting geometry tends to exhibit artifacts like spurious noise and a shrink-wrapping effect around the hair.

Static Facial Hair Capture. In contrast to the large body of work on face capture, less than a handful of papers address the problem of capturing facial hair. Early work by Herrera et al. [2010] investigated image-based techniques for modeling static facial hair in texture space, growing 3D hairs by shooting particles, and achieving limited fidelity and realism. Most closely related to our work is the method by Beeler et al. [2012] for multiview 3D reconstruction of facial hair on a static face, while also estimating the underlying skin position. Their work uses a passive, multi-camera setup with uniform lighting and a custom sequence of 2D and 3D algorithms for hair detection, reconstruction of 3D hair segments, and growing of individual facial hair strands. As an additional step, hair root positions are used to provide estimates of the underlying skin surface positions. While the resulting epi-surface works well for sparse and short facial hairstyles, it creates anatomically implausible skin geometry for subjects with long and dense beards, whose hair roots are not visible. Parallel work by Fyffe [2012] captures facial hair using a specialized photo-consistency loss and particle-based hair primitives. Furthermore, LeGendre et al. [2017] investigated modeling the very fine, low contrast vellus hair using asperity scattering along a subject's backlit silhouette. More recently, the ideas in Beeler et al. [2012] were extended for monocular facial hair capture by Rotger et al. [2019], but with more limited fidelity due to the lack of multiple viewpoints. All of these methods target individual static faces in isolation and do not address the problem of performance capture with dense temporal correspondence. In this work, we build upon Beeler et al. [2012] to add these missing capabilities and further improve quality and density of the recovered hairstyles, also

providing more plausible estimates of the underlying skin surface as it deforms during a facial performance.

Static Scalp Hair Capture. Scalp hair growing on top and around a person's head is one of the most distinctive identity features that is external to the face. As such, significantly more research work has been devoted to modeling and capturing scalp hair. Early work by Nakajima et al. [1997] first explored the idea of reconstructing 3D scalp hair with multiple images, creating and then growing hairs inside a 3D volume. Subsequently, Grabli et al. [2002] leveraged specular reflections over an image sequence under moving light sources to acquire the 3D orientation of hair strands and ultimately grow hair from such orientation field. Paris et al. [2004] further improved on these ideas and, in their well-known work *Hair Photobooth* [Paris et al. 2008], used projectors and multiple cameras to capture both the geometry and appearance of a diverse set of hairstyles. Jakob et al. [2009] captured single hair fibers using a sequence of images and a camera with very shallow depth of field. Herrera et al. [2012] proposed another innovative approach for hair capture using thermal imaging techniques. The work by Luo et al. [2013] is another very important contribution towards hair capture, based on graph data structures and global optimization of complex hairstyles. Zhang et al. [2017] investigated the use of only four sparse views (front, back, left, right) of the head to reconstruct the hair, allowing for the creation of different combinations of hairstyles. Hu et al. [2014] leveraged a dataset of physically simulated hairs, achieving robust results on hairstyles with large occlusions, multiple layers, and complex structures. The state-of-the-art work by Nam et al. [2019] redefined hair capturing as a multiview line-based patch match problem but relies on a dense array of 70 or more cameras. Recently, Sun et al. [2021] proposed a method to both capture scalp hair geometry and appearance, reformulating multiview triangulation of oriented hair segments, and using controllable lighting to provide robust features for multiview stereo matching and appearance computation. Their method arguably works with as few as 24 cameras, but they only demonstrate results on synthetic datasets. In contrast to all the above hair static capture methods, we focus on high-quality *dynamic facial hair* capture using only 14 frontal cameras in a standard photogrammetry setup with passive, uniform lighting, without requiring large and expensive capture volumes.

Dynamic Scalp Hair Capture. In the entertainment industry, scalp hair geometry is significantly more useful when combined with a motion model. While the predominant technique is still manual hair grooming and simulation, some work has investigated the challenging task of capturing dynamic scalp hair. Ishikawa et al. [2007] approached hair motion capture by attaching physical markers on long hair strands to track them. Yamaguchi et al. [2009] proposed the first passive approach for dynamic hair capture, extending static capture ideas with temporal smoothing of hair strands, but remaining limited to mostly straight hairstyles. Xu et al. [2014] analyzed space-time slices through stacks of hair images and traced a motion path for each hair pixel, obtaining impressive results for long hair, but also remaining limited to fairly straight hairstyles. Hu et al. [2017] combined dynamic scalp hair capturing with physical simulation to facilitate artistic control. Other works have focused on monocular scalp hair capture using learning-based methods to derive stronger data-driven priors, but achieving low-fidelity results that mainly capture the overall shape of the hair and predominant strand orientation [Chai et al. 2013; Liang et al. 2018; Yang et al. 2019]. Despite the larger body of work on scalp hair capture, such methods are not directly applicable in the scenario of facial performance capture due to differences in the hardware and accuracy requirements and also the nature of facial hair, which is shorter, with less uniform density and orientation, and undergoes a unique type of self-collision, occlusion and deformation as the underlying skin surface deforms during facial expressions. These challenges are exactly the ones we target in this work.

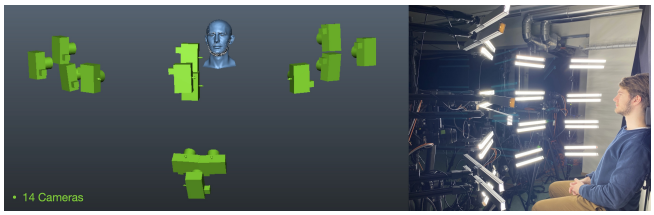

Fig. 2. We adopt a standard setup for facial performance capture based on Riviere et al. [2020], with 14 video cameras organized into four triplets (each including a stereo pair) and two additional cameras for hair capture.

Neural Rendering. The recent explosion of neural rendering approaches [Gafni et al. 2021; Lombardi et al. 2019; Park et al. 2021; Tewari et al. 2020] has enabled the synthesis of photorealistic images of complete human head models, including scalp and facial hair, even under arbitrary viewpoints, head poses and expressions. This is achieved in a data-driven way, without the explicit estimation of high-quality 3D geometry assets. Although impressive, these methods do not yet allow for the level of artistic control and editability (e.g. relighting, beard grooming, removal, animation, etc.) that is expected in familiar, high-quality production pipelines in the entertainment industry and related fields.

3 PERFORMANCE CAPTURE WITH FACIAL HAIR

This section presents our novel method for dynamic 3D facial hair capture and tracking that is designed for use with standard photogrammetry systems for facial performance capture, such as [Beeler et al. 2011; Gotardo et al. 2018; Riviere et al. 2020]. We use a common capture setup comprised of 16 banks of LED lights, for constant and uniform lighting, and 14 synchronized, color video cameras with resolutions ranging from 12 to 50 megapixels. All cameras and lights are placed around the frontal hemisphere of the captured subject, who can deliver facial performances in a fixed seated position, Fig. 2. Twelve cameras are organized into four triplets (front-up, front-bellow, left, right), where each triplet contains a stereo pair and a third central camera that is cross-polarized relative to the lighting; this central camera does not capture specular highlights and cross-polarization is only required to capture the appearance parameters of facial skin [Riviere et al. 2020]. Two additional cameras are placed in between the triplets above, one on each side of the captured subject, to assist in facial hair capture. Note that our facial hair capture method can work with additional cameras, in different

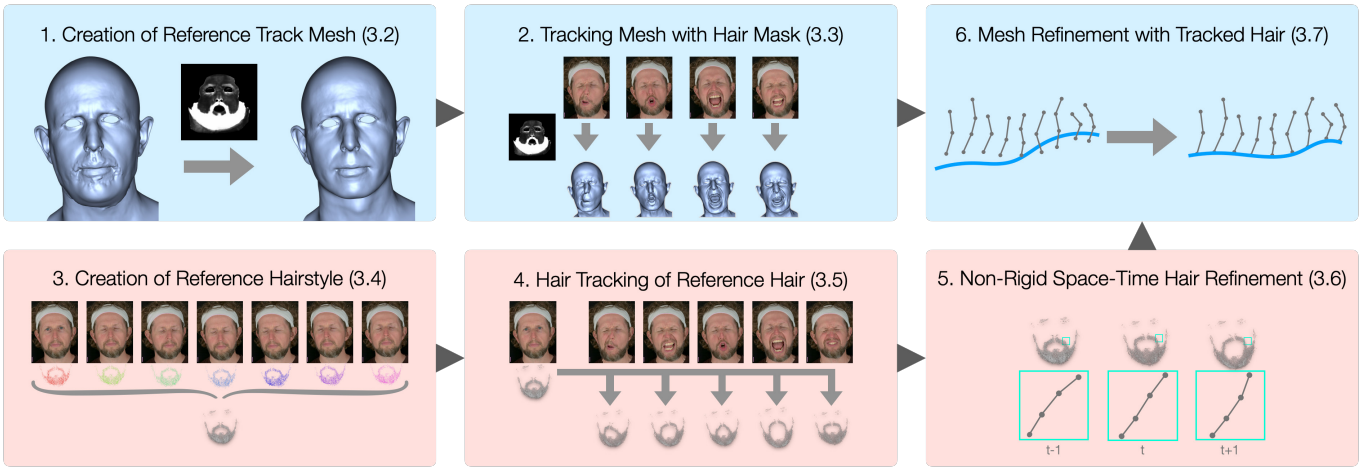

Fig. 3. Overview of the proposed dynamic facial hair capture pipeline comprising two main stages: (top) facial skin reconstruction and tracking, and (bottom) facial hair reconstruction and tracking, whose solutions are coupled and computed in alternation, via multiple refinement steps (Section 3.*x*, as indicated).

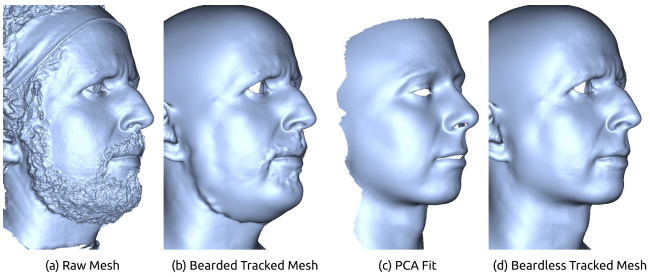

Fig. 4. Steps for creating a hair-free reference mesh with well-defined canonical topology for tracking throughout a facial performance.

arrangements and without polarized lighting, as long as the stereo pairs provide sufficient coverage of the face.

We approach the problem of dynamic facial hair capture by splitting it into two parts: (*i*) facial skin reconstruction and tracking, and (*ii*) facial hair fiber reconstruction and tracking. Our pipeline alternates between these two tasks during refinement steps, converging to a well-tracked underlying skin surface with individually-tracked 3D facial hairs. These two main stages are further split into a sequence of steps as outlined in Fig. 3 and described below.

3.1 Initialization

Our facial performance capture pipeline begins with per-frame multiview stereo [Beeler et al. 2010], yielding a high quality face shape that we refer to as the *raw mesh* for each frame. This mesh contains detailed 3D geometric reconstructions for skin areas, but shrink-wrapped surfaces over facial hair regions, Fig. 4 (a). At this stage, the raw meshes of different frames have different topologies and vertex counts and are not yet in temporal correspondence.

We then borrow from the static facial hair capture method of Beeler et al. [2012] and compute initial and sparse per-frame hair reconstructions, adapting the method to work well with our specific camera layout and image resolution. In particular, images were resized such that individual hairs had a width of approximately 3 pixels, allowing us to use the original hyper-parameters of Beeler et al. [2012], with the exception of the threshold for 2D hair growing, which we reduced from 0.5 to 0.3 in order to obtain slightly more density in the static hairstyle reconstructions given that we had fewer camera angles. Note, however, that such static hair reconstructions are not directly suitable for performance capture as they are too sparse and noisy, with significant temporal jitter due to the inconsistent reconstructions and lack of common topology across frames. These inaccuracies get amplified with more expressive facial performances and with added motion and depth-of-field blur.

3.2 Creating a Hair-Free Reference Mesh for Tracking

We now begin the novel contributions of our work, starting with the creation of a hair-free neutral reference mesh for tracking. Ultimately, our face performance capture goal is to track a standard mesh topology over all the per-frame raw meshes obtained above, to model the dense, temporal 3D deformation of all facial skin areas, regardless of whether or not they are covered by hair. This task requires that we first create an identity-specific *track mesh*, without facial hair, for the captured subject. To this end, we first select a neutral frame of the performance and register a template 3D face mesh with a well-defined canonical topology to the corresponding raw mesh. In this process, we need to discard and replace incorrectly reconstructed areas of the raw mesh that show shrink-wrap artifacts due to facial hair, or we end up with a bearded track mesh as shown in Fig. 4 (b). Fortunately, the facial hair areas can be identified based on inconsistencies in the multiview reconstruction, denoted by a surface confidence value, which is computed by examining photometric consistency as evaluated via normalized cross-correlation across the multiview images [Beeler et al. 2010]. We encode this confidence into a *facial hair mask* in UV texture space, as illustrated in the top-left block of Fig. 3. Using this mask, our goal is to predict a plausible skin surface for areas occluded by facial hair. Unlike previous work [Beeler et al. 2012], we take a data-driven approach

to the surface prediction task. Specifically, using a dense 3D face database [Chandran et al. 2020], we create a PCA face model and fit it to the hair-free skin surface region, while additionally ensuring that the surface remains behind any reconstructed hair strands, Fig. 4 (c). At this point we aim to obtain only an approximate skin surface under the hair regions, and so our approach does not depend on the particular PCA face model we constructed and we envision that alternative models would work equally well. As a final step, we combine the occluded skin surface predicted by PCA with the accurate multiview stereo geometry of non-hair regions via Laplacian mesh deformation [Sorkine et al. 2004], yielding our hair-free reference face mesh with a well-defined canonical topology for tracking throughout the facial performance, Fig. 4 (d).

3.3 Tracking the Reference Mesh with Facial Hair Mask

We can now apply the facial tracking method of Beeler et al. [2011], which aims to propagate the hair-less track mesh with template topology to all frames in the performance. Due to the presence of the facial hair, we adapt the tracking method by again masking all pixels that correspond to hair regions, and instead of tracking we allow the surface to deform as-rigidly-as-possible in these regions, while properly tracking the non-hair regions. While not a perfect solution, this process yields initial hair-free tracked meshes that have a consistent topology, dense temporal correspondences, and are devoid of the shrink-wrap artifacts of the original per-frame raw meshes. The resulting facial surface under hair regions will be further refined in the subsequent steps, but is already plausible and accurate enough to proceed with 3D hair fiber reconstruction.

3.4 Creating the Reference Facial Hairstyle

We now turn to our other goal: capturing a temporally-consistent, high-quality, dense facial hairstyle for each video frame. We start with our reference frame, showing a frontal neutral face expression, and build a reference hairstyle for our reference track mesh. Empirically, we found that using the static method of Beeler et al. [2012] on the reference frame resulted in sparse and inaccurate facial hair due to hair occlusions, defocus from shallow depth of field, and the few stereo cameras in our setup. Instead of requiring the use of a denser array of cameras, we opted to design our method so that it builds a denser, high-quality reference hairstyle by accumulating hair reconstructions over multiple frames in time. To achieve this goal, we capture additional viewpoints of the target subject by asking them to slowly rotate their face while maintaining a neutral facial expression, Fig. 5. As a result, we effectively increase the number of viewpoints by at least an order of magnitude. For each video frame showing this circular motion, we compute initial, static hair reconstruction following Beeler et al. [2012], as described above.

For creating a reference facial hairstyle, for subsequent tracking, we also make use of cumulative optical flow relative to the reference frame, which was already precomputed for each camera view during the previous mesh tracking step. Here, the optical flow vectors are used to constrain a non-linear optimization for 3D hairstyle registration throughout the rotating neutral face video, bringing all per-frame 3D hair reconstructions into alignment with the coordinate system of the reference frame. This process allows our

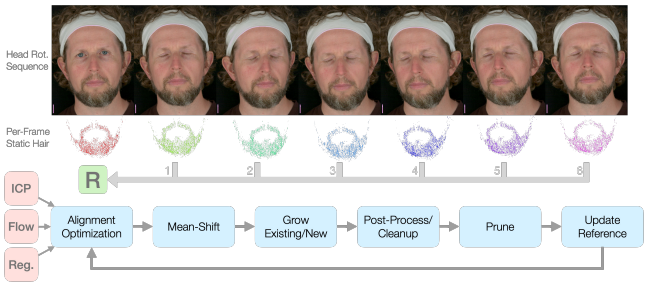

Fig. 5. Reference hairstyle reconstruction from rotating neutral face video: static hair reconstructions (1-6) [Beeler et al. 2012] are brought into alignment with the reference frame and merged by our 3D hair registration algorithm to build up a single, denser reference hairstyle, **R**. The resulting, high-quality reference hairstyle, together with the hair-free reference track mesh, can then be used together for performance capture in other videos.

reference hairstyle to grow denser and more accurate by accumulating more 3D hairs, in comparison to the initial sparse and noisy reconstructions based on Beeler et al. [2012] in Fig. 5.

3.4.1 Hair Alignment Optimization. More formally, 3D hair registration is done as an iterative closest point (ICP) optimization that considers optical flow and both the 3D position and orientation of hair points. We optimize for a locally rigid deformation field with per-frame and per-hair 3D translations $\mathbf{t}$ and rotations $\mathbf{q}$, encoded as unit quaternions,

$$\min_{\mathbf{t},\mathbf{q}} \quad \lambda_I E_{ICP}(\mathbf{t},\mathbf{q}) + \lambda_f E_{flow}(\mathbf{t},\mathbf{q}) + \lambda_n E_{neigh}(\mathbf{t},\mathbf{q}), \tag{1}$$

where the rightmost neighborhood energy enforces spatial regulation to constrain the hairstyle alignment to be as rigid as possible. The per-frame deformation fields $(\mathbf{t},\mathbf{q})$ are initialized from rigid transformations of small skin patches around each hair, relative to the reference track mesh. We implement this hair registration ICP method using the well-known, auto-differentiable Ceres Solver package [Agarwal et al. 2016].

ICP Energy. Iterative closest point alignment is a natural solution for the registration of our 3D hair point clouds. First, we establish correspondences between each hair point p_{ij} (point j of i-th hair), at a given video frame, and the closest hair point p' in the reference frame, with the reference hairstyle. If these hairs are in close proximity of 5mm and have orientations matching within a threshold of 30 degrees, then the match (p_{ij}, p') is considered in the optimization. The ICP energy is defined as:

$$E_{ICP}(\mathbf{t},\mathbf{q}) = \sum_{(p_{ij},p')} \left\| R(\mathbf{q}_i)p_{ij} + \mathbf{t}_i - p' \right\|^2, \tag{2}$$

where $R(\mathbf{q}_i)$ is the rotation matrix of the i-th quaternion (hair). As ICP is iterative, we repeat the optimization after each update on the matched point set (p_{ij}, p') and $\lambda_I = 0.001$. In practice, we found that 3 repetitions are enough, due to the good initialization obtained from the local rotation of skin patches in the tracked face mesh.

Flow Energy. Instead of point matches, this term considers individual 3D hair points p_{ij} and enforces that their alignment $(\mathbf{t}_i, \mathbf{q}_i)$ is consistent with the motion estimate given by the optical flow in each

view. Let x_{ijk} denote the projection of p_{ij} onto camera view k, using the projection matrix Q_k. We then use the optical flow F_k to propagate this point and derive a target position $\hat{x}_{ijk} = x_{ijk} + F_k(x_{ijk})$ at the reference frame. The flow energy is then given by,

$$E_{flow}(\mathbf{t}, \mathbf{q}) = \sum_k \sum_{p_{ij}} \psi\left(\left\|Q_k(R(\mathbf{q}_i)p_{ij} + \mathbf{t}_i) - \hat{x}_{ijk}\right\|^2\right) V_k(p_{ij}), \quad (3)$$

where $\psi(\cdot)$ is the robust Huber loss function, $V_k(p_{ij})$ is a cosine-weighted per-camera visibility term, and $\lambda_f = 0.0001$.

Neighborhood Regularizer. The third energy for hair alignment optimization is a neighborhood regularization term that seeks to preserve the local structure of the hairstyle by favoring similar transformations $(\mathbf{t}_i, \mathbf{q}_i)$ for aligning neighboring hair follicles at the current frame towards the reference frame. For the i-th hair follicle, let N_{ri} be the set of its neighboring hairs within a distance of $r = 5$mm. The spatial smoothness energy is defined as,

$$E_{neigh}(\mathbf{t}, \mathbf{q}) = \sum_i \sum_{i' \in N_{ri}} W(i, i') \left(\begin{bmatrix} \lambda_{trans} \\ \lambda_{rot} \end{bmatrix}^T \begin{bmatrix} \|\mathbf{t}_i - \mathbf{t}_{i'}\|^2 \\ \|\mathbf{q}_i - \mathbf{q}_{i'}\|^2 \end{bmatrix} \right), \quad (4)$$

where $W(\cdot, \cdot)$ is a Gaussian weighting factor with $\sigma = 2.5$mm. Note that for this term we chose $\lambda_{trans} = 0.01$ and $\lambda_{rot} = 1$ for the translation and rotation regularization weights, as well as $\lambda_n = 1$.

3.4.2 Hair Growth and Refinement. Once aligned to the reference frame, the reference hairstyle needs to be updated by merging the newly aligned static hairs. To this end, we first use the line mean-shift algorithm to filter out the noise. We then subsequently grow first the existing hairs and then the new ones with a Forward Euler approach, in line with techniques described in Nam et al. [2019]. Next, we further apply the post-processing, cleanup and pruning methods described in Section 4.3 of Beeler et al. [2012], and finally update the new reference hairstyle. We apply this entire pipeline iteratively for each time step in the head rotation sequence and thus slowly build up a more and more dense hairstyle over time, as in Fig. 5. After iterating through the entire sequence, we apply one last post-processing step before fixing the subject specific hairstyle topology which we will use to track novel performances.

3.5 Tracking the Reference Facial hairstyle

At this stage, we have already built (*i*) a hair-free reference mesh that can be tracked to any frame of a captured facial performance, and (*ii*) a dense, reference 3D hairstyle with a fixed topology, to be tracked with the reference skin mesh. We now present the method for tracking this hairstyle as it deforms over time due to the changes in facial expression and head pose, while also maintaining the internal structure of the reference hairstyle to some extent. We assume that our hair-free reference mesh has already been tracked to each video frame of a captured performance, using the method in Section 3.3.

Tracking the reference hairstyle to a particular video frame of a performance is also formulated as a non-linear optimization problem, similar to the one used above for building the hairstyle. The goal now is to deform the reference hairs and bring them into alignment with the current video frame. This time, however, we do not need to compute static hairs for the current frame using Beeler et al. [2012] because these single-frame results are noisy and incomplete,

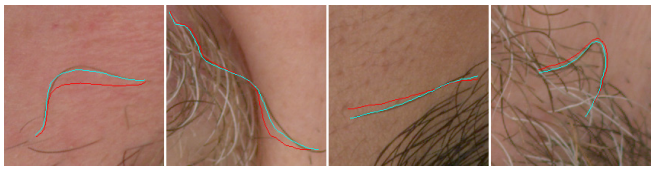

Fig. 6. Example rigid (red) and non-rigid (cyan) hair tracking: the non-rigid hair refinement improves alignment towards the input images by deforming the tracked hair geometry.

for use in isolation. Thus, we drop the ICP constraints in our optimization problem above and rely solely on optical flow (already computed before mesh tracking) and the spatial regularizer,

$$\min_{\mathbf{t},\mathbf{q}} \quad \lambda_f E_{flow}(\mathbf{t}, \mathbf{q}) + \lambda_n E_{neigh}(\mathbf{t}, \mathbf{q}). \quad (5)$$

Note that, now, both the optical flow and the solution $(\mathbf{t}, \mathbf{q})$ describe a motion field in a different direction, from the reference hairstyle onto the facial pose and expression of the current video frame.

The result of this processing stage is an initial, complete solution with dynamic facial hairstyle tracked throughout a complete facial performance. The subsequent steps in the following are then used to further refine the tracked hair and the skin surface underneath.

3.6 Non-Rigid Space-Time Hair Refinement

The tracked facial hairs already follow the underlying facial performance motion quite well after the step above, but may not perfectly align to the individual hair strands in the images of each frame. This is largely because (a) the flow vector computation was perhaps inaccurate, (b) the regularization prevented hitting the perfect result, and/or (c) we have so far only solved for rigid per-hair transformations, while hairs do in fact deform slightly due to facial expressions. Thus, starting from the results above, this final non-rigid hair refinement step deforms each tracked hair, in each frame, such that it best aligns with the hair images from the multiple cameras, Fig. 6.

This step is also implemented as a non-linear optimization, but now the inputs and outputs are individual points along an individual hair. The refinement is formulated over time, taking multiple frames into consideration at once, such that we can add a temporal regularization term to promote smoothness and avoid temporal jitter. This non-rigid optimization has the following terms:

$$\min_{\mathbf{p}} \sum_t \sum_i \quad \lambda_H E_{HDF}(p_i^t) + \lambda_p E_{pos}(p_i^t) + \lambda_{len} E_{len}(p_i^t) + \lambda_{lap} E_{lap}(p_i^t) + \lambda_t E_{temp}(p_i^t), \quad (6)$$

where t is the frame (time) index, with the energy terms as detailed next. The final result is an accurate, temporally-tracked, deforming hairstyle that more accurately matches the real actor's facial hair throughout the captured performance.

Hair Distance Field (HDF) Energy. To better align 3D hairs with their projects on the input images, we repurpose the hair distance fields (HDFs) computed during static hair reconstruction with Beeler et al. [2012] — an HDF encodes the distance from each image pixel to the closest hair follicle. The intended effect of this energy term is to encourage the optimization to move hair points

p_{ij}^t towards the detected hair lines on the image plane:

$$E_{HDF}(p_i^t) = \sum_k \sum_j \psi \left(\left\| \frac{1}{H_k^t(Q_k p_{ij}^t) + \epsilon} - 1 \right\|_2^2 \right) V_k(p_{ij}), \quad (7)$$

where H_k^t denotes the HDF for camera k at time t, and the other terms are as in Eq. 3. The energy weight is set to $\lambda_H = 0.01$.

Position, Length and Laplacian Regularizers. To balance the deformation term represented by the HDF energy, we add regularizers that seek to preserve the position and geometry of each hair where the HDF is noisy or ill defined. We use simple regularizers in the form of a position energy E_{pos} towards the rigid initialization, a segment length energy E_{len}, as well as 1D Laplacian E_{lap}. Note that for position regularization, we hereby increase the root weight by a factor of 100, to prevent the tip from drifting. Furthermore, while E_{len} is regularizing the current hair points towards the previous rigid distance between points with a simple L2 norm, E_{lap} does the same for the 1D Laplacian vector of the rigid solution. We set energy weights as $\lambda_p = 0.01$, $\lambda_{len} = 100$ and $\lambda_{lap} = 1000$.

Temporal Regularizer. Due to the potentially noisy nature of the driving deformation term E_{HDF}, we add temporal regularization to preserve consistency in 3D hair geometry over time. We use the common second-order central difference formula to allow for a smooth trajectory through time between hair points p_{ij}^{t-1}, p_{ij}^t and p_{ij}^{t+1}. However, to penalize only non-rigid deformation, we first rigidly align the hair geometry within this 3 frame temporal window, using the previously computed, per-hair transformations $(\mathbf{t}, \mathbf{q})$ from Section 3.5. Let $\hat{T}_i^{t-1}$, $\hat{T}_i^t$ and $\hat{T}_i^{t+1}$ denote these rigid transformations that bring the hair shapes into alignment in the coordinate system of the reference frame. Then, the temporal energy is:

$$E_{temp}(p_i^t) = \sum_j \left\| \hat{T}_i^{t-1} p_{ij}^{t-1} - 2\hat{T}_i^t p_{ij}^t + \hat{T}_i^{t+1} p_{ij}^{t+1} \right\|_2^2,$$

with weight $\lambda_t = 1$.

We solve this entire optimization for all hairs one frame at a time, alternating between frames while keeping the results for neighboring frames fixed. This provides an efficient optimization strategy and temporally smooth solution within 5 iterations on each frame.

3.7 Mesh Refinement with Tracked hairstyle

Now that we have more accurate hairs for each video frame, we return to the problem of refining our estimates of the underlying skin surface, improving upon the results from the previous step in Section 3.4. To accomplish this, we place deformation handles on a sparse set of 3D points on the skin areas covered by facial hair, Fig. 7 (left). For each handle, we determine a subset of the facial hairs that are close to the handle in the reference frame, and assign the set of root hair points to the handle. Then, for every video frame, we compute the least-squares fit of a rigid transformation for that set of root points between the reference frame and the deformed frame, and then apply the same transformation to the handle point. Finally, given all handle deformations, we solve for the final facial surface which best tries to match those deformations, while also remaining smooth (via Laplacian regularization) and fixing in space the facial surface that was not underneath facial hair. The final result is a temporally tracked facial skin surface under the accurately tracked facial hair, Fig. 7 (right).

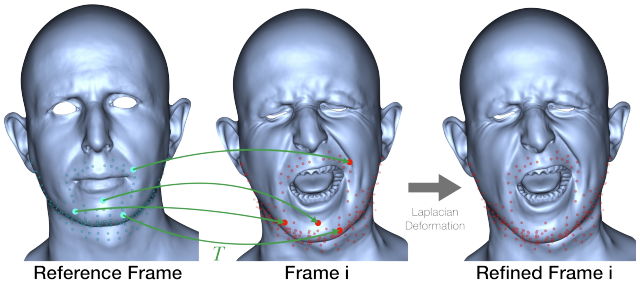

Fig. 7. We sample a sparse set of points in the facial hair region of the reference surface (cyan) and for each point, use the surrounding hair point correspondences between reference and frame i to find a local transformation T. We use T to get the target surface points (red) and apply Laplacian deformation [Sorkine et al. 2004] to arrive at the refined result on the right.

4 RESULTS

We now show the results of our facial hair tracking method for high-fidelity performance capture. To evaluate our method, we captured 5 different actors with varying facial hair length, density and color. Each actor performed a small quasi-rigid rotation with neutral expression in order to build the dense hairstyle, as described in Section 3.4, followed by additional performances that included either dialog or a random set of facial expressions. For all of the following results, please refer to the supplemental video in order to better appreciate the result quality in motion.

4.1 Qualitative Evaluation

We begin with showing the variety of reconstruction results as a whole, in Fig. 1, Fig. 8, and Fig. 9, where several frames from dynamic performances are illustrated. For each frame we show one of the multiview input images, and corresponding 3D reconstruction results of both the face and the hair geometry. Fig. 8 and Fig. 9 additionally show the accuracy of the recovered hairs, by overlaying them (in cyan) on the input image. All of these results were generated automatically, without any artist intervention or manual cleanup. Note how our approach is able to reconstruct beards, mustaches, and even partial eyebrows. To better examine the individual hair reconstructions, we show a cropped-zoom result in Fig. 10, for two different frames of two of the actors, also shown from two different views from the multiview set. Although not every single hair is perfectly aligned with the images, our reconstruction is overall quite accurate and represents a very faithful reconstruction of the particular facial hairstyle.

After reconstructing the facial hair performance, our method further refines the underlying skin geometry using the recovered hairs (Section 3.7). The benefit of doing this is illustrated for one frame in Fig. 11. Here you see that the initial estimated face surface can be inaccurate, occluding many of the reconstructed hairs. After refinement, we obtain a more faithful reconstruction.

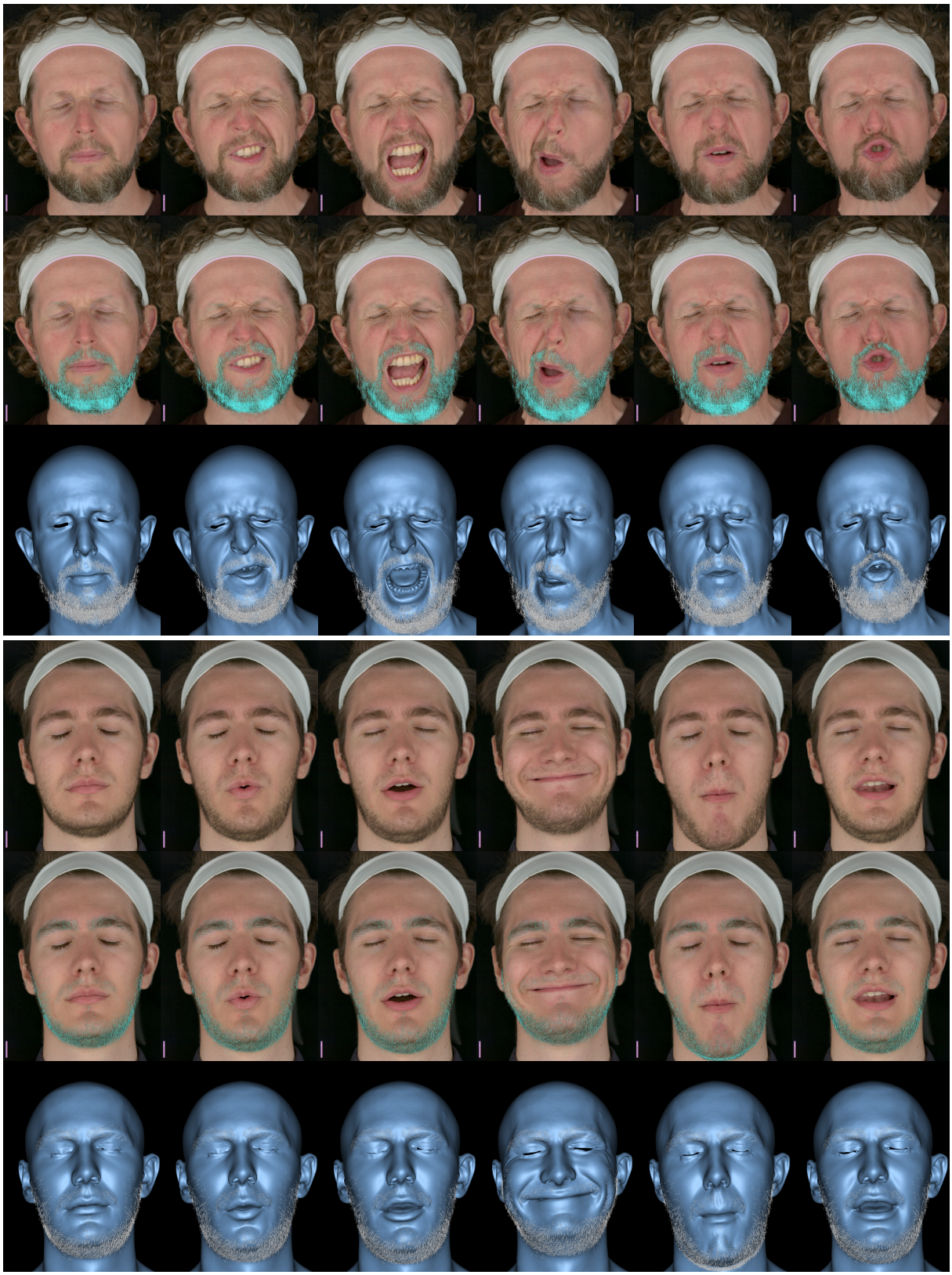

Fig. 8. Several frames of performance with 3D face and facial hair capture for two actors, showing the input images, hair overlay in cyan, and 3D geometry.

Fig. 9. Several frames of performance with 3D face and facial hair capture for two actors, showing the input images, hair overlay in cyan, and 3D geometry.

Fig. 10. Here we show close-up views of our hair reconstruction results for two different frames of actors (shown from different views), including the reconstructed hairs overlaid in cyan and the full 3D geometry results.

Finally, we highlight that our temporally-consistent reconstructions with a common topology for the facial hair allows to artistically edit captured performances in a straightforward way. For example, the neutral frame hairstyle can be groomed or trimmed to any desired style, and then these edits are automatically propagated to the full sequence, as shown in Fig. 12, where the full beard was trimmed on the sides to form a large goatee instead. The performance naturally reflects the change in facial hairstyle. Further artistic examples are provided in the supplemental video.

4.2 Comparisons and Ablations

We start by showing an ablation on the impact of the different terms in Eq. 1 (Section 3.4), used to create the reference hairstyle, in Fig. 13. Here, we zoom in on a small patch where three hairs from one of the source frames (green) are to be aligned with the current estimate of the reference hairstyle in the target reference frame

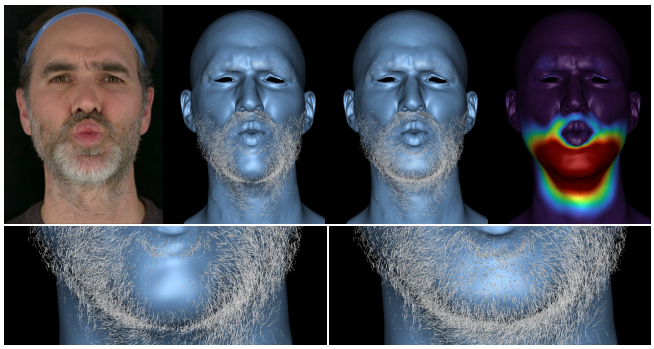

Fig. 11. Our per-frame face mesh refinement step corrects the initial surface estimate using the reconstructed facial hairs. Here we see 1 frame of correction, from left to right: one input image, initial surface estimate with reconstructed hairs, our refined surface with reconstructed hairs, difference between original and refined surface (blue: 0mm to red: 5mm). A zoom-in on the improvement is shown in the bottom row.

Fig. 12. As the reconstructed facial hair has a temporally-consistent connectivity, we can allow artistic edits to automatically propagate to full sequences, like trimming the sides of the beard.

(red). Note that the center hair does not yet exist in the reference frame, which includes other hairs that were not recovered in the source frame. Without the flow term E_{flow} (yellow) the center hair tries to align incorrectly to existing hairs due to the ICP term E_{ICP}. Yet without the ICP term (pink), the flow term alone cannot always match all hairs, as seen by the third hair. Using both terms (cyan) allows our optimization to align all hairs reasonably well. In Fig. 14, we show the necessity for adding a neighborhood regularizer to the rigid tracking in Eq. 5 (Section 3.5). Due to inaccuracies in the optical flow, some hairs fly off when only relying on E_{flow}. By also adding the neighborhood regularizer E_{neigh}, the local structure of the hairstyle is preserved. As a final ablation, we demonstrate why E_{ICP} was not used during rigid hair tracking. Since the per-frame hair reconstructions from Beeler et al. [2012] are inconsistent and noisy, introducing an ICP term into Eq. 5 results in jittery hair reconstructions, as shown in the supplemental video.

Related to our method is the sparse facial hair and skin reconstruction method of Beeler et al. [2012]. As described earlier, their approach targets single shot reconstructions only and is thus not applicable to performances, which is our goal. However, there exists yet an opportunity for comparison, both in terms of the facial hair recovered as well as the estimation of the underlying surface. As illustrated in Fig. 15 (top), our data-driven approach for creating a

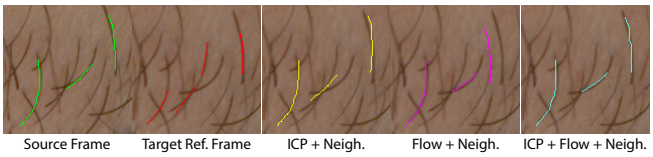

Fig. 13. An ablation on the impact of each term in the creation of the reference hairstyle (Eq. 1). Three hairs in one of the source frames are being aligned to the current hairstyle in the reference target frame. Both the Flow term and the ICP term are required to achieve high-quality alignments.

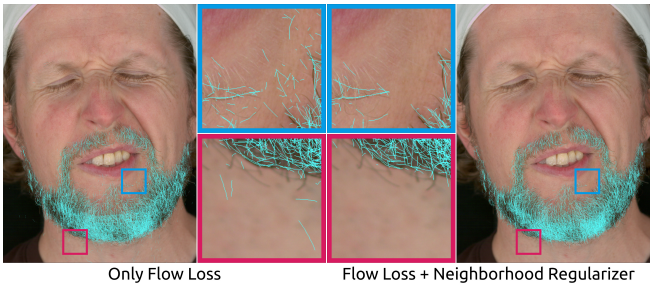

Fig. 14. An ablation on the impact of adding the neighborhood regularizer to the rigid hair tracking optimization.

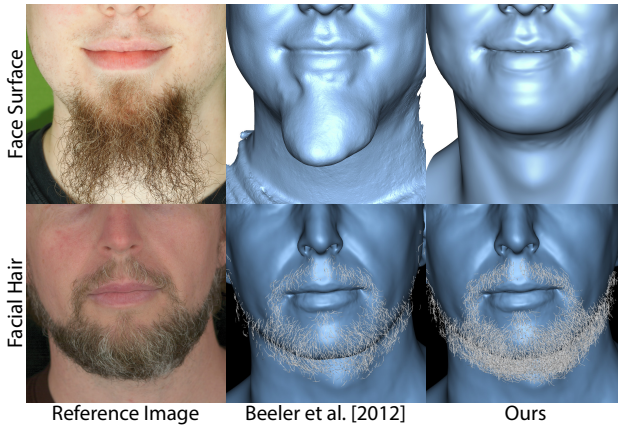

Fig. 15. We compare our face surface and facial hair reconstruction methods to the static capture method of Beeler et al. [2012].

hair-free reference mesh for tracking (Section 3.2) performs better than the surface estimation method of Beeler et al. [2012] in the case of longer beards. Furthermore, we can justify our approach to combine information from many frames to build the reference facial hairstyle (Section 3.4) by comparing to the single-shot hair reconstruction of Beeler et al. [2012]. As shown in Fig. 15 (bottom), combining data from several frames with our novel registration process results in a more dense and faithful hairstyle.

We further compare our non-rigid space-time facial hair tracking method to a recent point-based non-rigid registration technique known as BCPD [Hirose 2021], which is a Bayesian coherent point drift method designed for point set registration. In our context, we

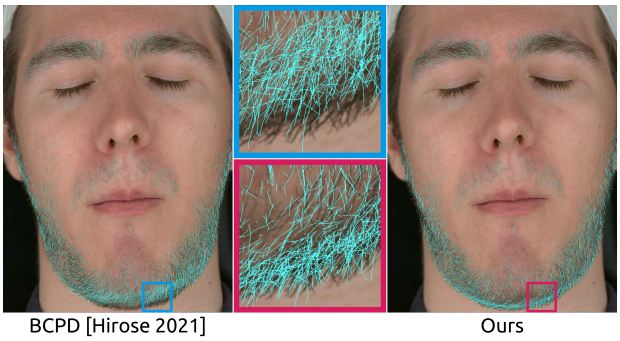

Fig. 16. We compare our facial hair tracking approach to the non-rigid point registration algorithm of Hirose [2021].

evaluate the use of BCPD to track the hair points as compared to the proposed method in Section 3.5. For both methods we start with the dense reference facial hairstyle and propagate it through time. When using BCPD, we track the reference hairstyle by aligning it non-rigidly to individually-reconstructed per-frame hairstyle using the method of Beeler et al. [2012]. The result is shown in Fig. 16, which illustrates a limitation of BCPD in that it cannot always align the hairstyle to the correct position, where our proposed method is more accurate. BCPD also produces more noisy temporal results, as illustrated in the accompanying video. Furthermore, BCPD requires the additional time-consuming step of reconstructing the per-frame facial hair as a pre-process, which our method does not require.

5 CONCLUSION

In this work we tackle the complex problem of facial hair tracking for high-fidelity performance capture. Our method is able to reconstruct and track individual facial hairs over complex performance sequences in a traditional multiview reconstruction scenario. We additionally create a realistic approximation of the dynamic clean-shaven facial surface, as if the actor had been captured without facial hair, thus removing the need to actually shave. Both the face surface and the facial hairs are in topological correspondence over time, and so the result fits naturally into production-level facial animation pipelines. As such, our approach is directly applicable to industry-based performance capture methods.

To the best of our knowledge, this work represents the first method for facial performance capture with facial hair. As such, there are some aspects that are open for further improvement. For example, while our results very closely match the hairstyles of the captured actors, even under complex motions, the recovered hairs may not perfectly match the multiview imagery at a strand-accurate level. We have also not tested extremely long facial hair, which may pose additional problems due to large occlusions. As we rely on the static hair reconstruction method of Beeler et al. [2012], we also share their limitations regarding image quality requirements and we acknowledge the challenge of hair color that too closely matches the skin appearance. Also, our recovered hairs are not rooted into the facial surface and thus a method to connect the hairs naturally to the face might be desirable. Finally, in addition to the geometry, facial performance capture methods often reconstruct texture maps for

attributes like color, displacements and normals. While we propose a method to estimate an approximate skin surface underneath hair regions, we have not yet attempted to recover also the texture maps for these regions, nor hair appearance parameters. We consider all of these topics to be great avenues for follow up research.

ACKNOWLEDGMENTS

We wish to thank Steven Poulakos for posing as a capture subject, and Max Grosse for his continuous support of our capture hardware.

REFERENCES

Sameer Agarwal, Keir Mierle, and Others. 2016. Ceres Solver. http://ceres-solver.org.

Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-Quality Single-Shot Capture of Facial Geometry. *ACM Trans. Graphics (Proc. SIGGRAPH)* 29, 4, Article 40 (2010).

Thabo Beeler, Bernd Bickel, Gioacchino Noris, Paul Beardsley, Steve Marschner, Robert W. Sumner, and Markus Gross. 2012. Coupled 3D Reconstruction of Sparse Facial Hair and Skin. *ACM Trans. Graphics (Proc. SIGGRAPH)* 31, 4, Article 117 (2012).

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-Quality Passive Facial Performance Capture Using Anchor Frames. *ACM Trans. Graphics (Proc. SIGGRAPH)* 30, 4, Article 75 (2011).

Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. Lightweight Eye Capture Using a Parametric Model. *ACM Trans. Graphics (Proc. SIGGRAPH)* 35, 4, Article 117 (2016).

Amit Bermano, Thabo Beeler, Yeara Kozlov, Derek Bradley, Bernd Bickel, and Markus Gross. 2015. Detailed Spatio-Temporal Reconstruction of Eyelids. *ACM Trans. Graphics (Proc. SIGGRAPH)* 34, 4, Article 44 (2015).

Derek Bradley, Wolfgang Heidrich, Tiberiu Popa, and Alla Sheffer. 2010. High Resolution Passive Facial Performance Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 29, 4, Article 41 (2010).

Menglei Chai, Lvdi Wang, Yanlin Weng, Xiaogang Jin, and Kun Zhou. 2013. Dynamic Hair Manipulation in Images and Videos. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 75 (2013).

Prashanth Chandran, Derek Bradley, Markus Gross, and Thabo Beeler. 2020. Semantic Deep Face Models. In *Int. Conf. on 3D Vision*. 345–354.

Graham Fyffe. 2012. High Fidelity Facial Hair Capture. In *ACM SIGGRAPH 2012 Talks*. Article 23.

Graham Fyffe, Koki Nagano, Loc Huynh, Shunsuke Saito, Jay Busch, Andrew Jones, Hao Li, and Paul Debevec. 2017. Multi-View Stereo on Consistent Face Topology. *Comp. Graphics Forum (Proc. Eurographics)* 36, 2 (2017), 295–309.

Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Niessner. 2021. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. In *IEEE Computer Vision and Pattern Recognition (CVPR)*. 8649–8658.

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 30, 6 (2011), 1–10.

Paulo Gotardo, Jérémy Riviere, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2018. Practical Dynamic Facial Appearance Modeling and Acquisition. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 37, 6, Article 232 (2018).

Stéphane Grabli, François X. Sillion, Stephen R. Marschner, and Jerome E. Lengyel. 2002. Image-Based Hair Capture by Inverse Lighting. In *Proc. of Graphics Interface (GI)*. 51–58.

Tomas Lay Herrera, Arno Zinke, and Andreas Weber. 2012. Lighting Hair from the inside: A Thermal Approach to Hair Reconstruction. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 31, 6, Article 146 (2012).

Tomas Lay Herrera, Arno Zinke, Andreas Weber, and Thomas Vetter. 2010. Toward Image-Based Facial Hair Modeling. In *Proc. of the 26th Spring Conf. on Computer Graphics*. 93–100.

Osamu Hirose. 2021. A Bayesian Formulation of Coherent Point Drift. *IEEE TPAMI* 43, 7 (2021), 2269–2286.

Liwen Hu, Derek Bradley, Hao Li, and Thabo Beeler. 2017. Simulation-Ready Hair Capture. *Comp. Graphics Forum (Proc. Eurographics)* 36, 2 (2017), 281–294.

Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2014. Robust Hair Capture Using Simulated Examples. *ACM Trans. Graphics (Proc. SIGGRAPH)* 33, 4, Article 126 (2014).

Takahito Ishikawa, Yosuke Kazama, Eiji Sugisaki, and Shigeo Morishima. 2007. Hair Motion Reconstruction Using Motion Capture System. In *ACM SIGGRAPH 2007 Posters*. 78–es.

Wenzel Jakob, Jonathan T. Moon, and Steve Marschner. 2009. Capturing Hair Assemblies Fiber by Fiber. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 28, 5 (2009), 1–9.

Samuli Laine, Tero Karras, Timo Aila, Antti Herva, Shunsuke Saito, Ronald Yu, Hao Li, and Jaakko Lehtinen. 2017. Production-Level Facial Performance Capture Using Deep Convolutional Neural Networks. In *Proc. of Eurographics Symposium on Computer Animation*. Article 10.

Chloe LeGendre, Loc Hyunh, Shanhe Wang, and Paul Debevec. 2017. Modeling Vellus Facial Hair from Asperity Scattering Silhouettes. In *ACM SIGGRAPH 2017 Talks*.

Tianye Li, Shichen Liu, Timo Bolkart, Jiayi Liu, Hao Li, and Yajie Zhao. 2021. Topologically Consistent Multi-View Face Inference Using Volumetric Sampling. In *IEEE Int. Conf. on Computer Vision (ICCV)*. 3824–3834.

Shu Liang, Xiufeng Huang, Xianyu Meng, Kunyao Chen, Linda G. Shapiro, and Ira Kemelmacher-Shlizerman. 2018. Video to Fully Automatic 3D Hair Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 37, 6, Article 206 (2018).

Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. 2019. Neural Volumes: Learning Dynamic Renderable Volumes from Images. *ACM Trans. Graphics (Proc. SIGGRAPH)* 38, 4, Article 65 (2019).

Linjie Luo, Hao Li, and Szymon Rusinkiewicz. 2013. Structure-Aware Hair Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 32, 4, Article 76 (2013).

Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. 2007. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. In *Proc. Eurographics Conf. on Rendering Techniques*. 183–194.

Masayuki Nakajima, Kong Wai Ming, and Hiroki Takashi. 1997. Generation of 3d hair model from multiple pictures. *IEEE Comp. Graphics and Applications* (1997).

Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheikh. 2019. Strand-Accurate Multi-View Hair Capture. In *IEEE Computer Vision and Pattern Recognition (CVPR)*. 155–164.

Sylvain Paris, Hector M. Briceño, and François X. Sillion. 2004. Capture of Hair Geometry from Multiple Images. *ACM Trans. Graphics (Proc. SIGGRAPH)* 23, 3 (2004), 712–719.

Sylvain Paris, Will Chang, Oleg I. Kozhushnyan, Wojciech Jarosz, Wojciech Matusik, Matthias Zwicker, and Frédo Durand. 2008. Hair Photobooth: Geometric and Photometric Acquisition of Real Hairstyles. *ACM Trans. Graphics (Proc. SIGGRAPH)* 27, 3 (2008), 1–9.

Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. 2021. HyperNeRF: A Higher-Dimensional Representation for Topologically Varying Neural Radiance Fields. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 40, 6, Article 238 (2021).

Jérémy Riviere, Paulo Gotardo, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2020. Single-Shot High-Quality Facial Geometry and Skin Appearance Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 39, 4, Article 81 (2020).

Gemma Rotger, Francesc Moreno-Noguer, Felipe Lumbreras, and Antonio Agudo. 2019. Single View Facial Hair 3D Reconstruction. In *Pattern Rec. and Image Anal.* 423–436.

Olga Sorkine, Daniel Cohen-Or, Yaron Lipman, Marc Alexa, Christian Rössl, and Hans-Peter Seidel. 2004. Laplacian Surface Editing. In *Proc. of the Symposium on Geometry Processing*. 175–184.

Tiancheng Sun, Giljoo Nam, Carlos Aliaga, Christophe Hery, and Ravi Ramamoorthi. 2021. Human Hair Inverse Rendering using Multi-View Photometric data. In *Eurographics Symposium on Rendering*.

Ayush Tewari, Mohamed Elgharib, Mallikarjun B R, Florian Bernard, Hans-Peter Seidel, Patrick Pérez, Michael Zollhöfer, and Christian Theobalt. 2020. PIE: Portrait Image Embedding for Semantic Control. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 39, 6, Article 223 (2020).

Ayush Tewari, Michael Zollhofer, Hyeongwoo Kim, Pablo Garrido, Florian Bernard, Patrick Perez, and Christian Theobalt. 2017. MoFA: Model-Based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction. In *Proc. ICCV Workshops*.

Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus Gross, and Thabo Beeler. 2016a. Model-Based Teeth Reconstruction. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 35, 6, Article 220 (2016).

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016b. An Anatomically-Constrained Local Deformation Model for Monocular Face Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 35, 4, Article 115 (2016).

Zexiang Xu, Hsiang-Tao Wu, Lvdi Wang, Changxi Zheng, Xin Tong, and Yue Qi. 2014. Dynamic Hair Capture Using Spacetime Optimization. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 33, 6, Article 224 (2014).

Tatsuhisa Yamaguchi, Bennett Wilburn, and Eyal Ofek. 2009. Video-Based Modeling of Dynamic Hair. In *Adv. in Image and Video Technology*. 585–596.

Lingchen Yang, Zefeng Shi, Youyi Zheng, and Kun Zhou. 2019. Dynamic Hair Modeling from Monocular Videos Using Deep Neural Networks. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 38, 6, Article 235 (2019).

Meng Zhang, Menglei Chai, Hongzhi Wu, Hao Yang, and Kun Zhou. 2017. A Data-Driven Approach to Four-View Image-Based Hair Modeling. *ACM Trans. Graphics (Proc. SIGGRAPH)* 36, 4, Article 156 (2017).

Gaspard Zoss, Thabo Beeler, Markus Gross, and Derek Bradley. 2019. Accurate Markerless Jaw Tracking for Facial Performance Capture. *ACM Trans. Graphics (Proc. SIGGRAPH)* 38, 4, Article 50 (2019).