Coupled 3D Reconstruction of Sparse Facial Hair and Skin

Thabo Beeler[1,2]   Bernd Bickel[1]   Gioacchino Noris[1,2]

Paul Beardsley[1]   Steve Marschner[1,3]   Robert W. Sumner[1]   Markus Gross[1,2]

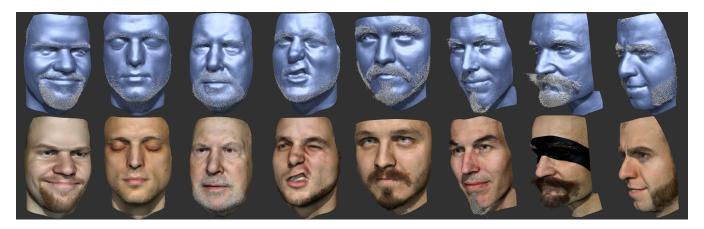[1]Disney Research Zurich   [2]ETH Zurich   [3]Cornell University

Figure 1: *Coupled 3D reconstruction of sparse facial hair and skin applied to various different facial hair styles.*

Abstract

Although facial hair plays an important role in individual expression, facial-hair reconstruction is not addressed by current face-capture systems. Our research addresses this limitation with an algorithm that treats hair and skin surface capture together in a coupled fashion so that a high-quality representation of hair fibers as well as the underlying skin surface can be reconstructed. We propose a passive, camera-based system that is robust against arbitrary motion since all data is acquired within the time period of a single exposure. Our reconstruction algorithm detects and traces hairs in the captured images and reconstructs them in 3D using a multi-view stereo approach. Our coupled skin-reconstruction algorithm uses information about the detected hairs to deliver a skin surface that lies underneath all hairs irrespective of occlusions. In dense regions like eyebrows, we employ a hair-synthesis method to create hair fibers that plausibly match the image data. We demonstrate our scanning system on a number of individuals and show that it can successfully reconstruct a variety of facial-hair styles together with the underlying skin surface.

CR Categories: I.3.3 [COMPUTER GRAPHICS]: Picture/Image Generation—Digitizing and scanning;

Keywords: Facial Hair, Skin Surface, Reconstruction.
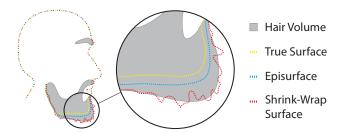
Links: ◆DL ▯PDF ▣WEB ▶VIDEO

1 Introduction

Since facial appearance plays such an important role in human communication, mastering the human face has long been a central goal of computer graphics. The characteristics of someone's face are a core component of their individuality and help make their physical appearance unique from every other person. While many facial characteristics are difficult to change, facial hair is one feature that is easily adapted. Some individuals meticulously sculpt their eyebrows hair-by-hair to ensure that the overall shape is perfectly formed and symmetric. A clean-shaven male face can look boyish and innocent. Many men instead choose to convey a more rugged, masculine appearance through a nearly unlimited variety of facial-hair styles, including beards, mustaches, and sideburns of all shapes and sizes. The popularity of these different facial-hair styles can fluctuate just as rapidly as fashion trends and varies dramatically from region to region, making facial hair a core piece of popular culture.

Despite the important role that facial hair plays in individual expression, existing face-capture technology does not easily accommodate facial-hair features. The problem setting is especially difficult because the hair and the underlying skin are often both visible, and a faithful face reconstruction must deliver accurate geometry for both the hair fibers and the underlying skin geometry. An accurate scan of a person with a scruffy beard should include each visible beard hair together with a high-fidelity representation of the skin underneath. Capture algorithms that focus on skin only will often deliver a skin surface that is "shrink-wrapped" around the facial-hair features, rather than reconstructing them as individual fibers on top of a skin surface. On the flip side, algorithms specialized for hair reconstruction typically focus on statistical properties of a thick head of hair that fully obscures the underlying scalp. These assumptions obviate the need for skin reconstruction or the consideration of individual fibers in isolation, making the algorithms unsuited for facial-hair reconstruction. As a result, existing scanning systems quietly ignore the huge variety of facial hair styles, treating them as error cases rather than as unique forms of human expression that deserve accurate reconstruction.

Our research improves the state-of-the-art of face capture with an algorithm that treats hair and skin surface capture together in a coupled fashion so that a high-quality representation of hair fibers as well as the underlying skin surface can be reconstructed. Since individual hair features are extremely fine and can vary greatly with head movement, we propose a single-shot capture system that uses consumer digital cameras without the need of multiple exposures or active illumination. Our hardware setup supports a variable number of cameras, so that additional face coverage is achieved simply by adding more cameras. All imagery needed for accurate reconstruction is captured within the time period of a single exposure.

Our reconstruction algorithm processes the captured images using a steerable filter kernel for explicit hair detection and produces a hair map for each image. Within these hair maps individual hairs are traced and then reconstructed and refined in 3D using multiview stereo. We then employ a skin reconstruction algorithm that uses information about detected hair pixels and the reconstructed hair fibers to deliver a skin surface that lies underneath all hairs irrespective of hair occlusions. In sparse regions where individual hairs are clearly visible in the captured images, our algorithm reconstructs each and every detected hair as a collection of line segments. In dense regions, such as the eyebrows, where many hairs overlap and obscure one another, our algorithm employs a hair-synthesis method to create hair fibers that plausibly match the image data. Likewise, when skin is visible through sparse hair, our system accurately reconstructs it, and when skin is obscured our system proposes a plausible solution. We demonstrate this algorithm with a collection of scans of individuals exhibiting a variety of different facial-hair styles.

As a useful geometric construct for the system, we introduce the concept of the *skin episurface* (see Figure 2). In the case where there is no hair or low density hair over visible skin, the skin episurface is a close approximation to the true skin surface. In the case where the hair is dense and no skin is visible, the skin episurface is a postulated 3D surface below the top layer of visible hair. While it is not a true surface, the motivation for the construct is that it enables a unified approach to the processing, across areas of clear skin and dense beard. Note that it should not be thought of as a dilation of the skin surface - in the case of protruding facial hair, something like a goatee, the episurface will also form a protruding shape. In addition to the technical building blocks of our solution, we also show that the fidelity of face scans is enhanced when facial hair is accurately reconstructed and provides a more faithful representation of an individual's unique look.

Figure 2: *The skin episurface is a pseudo-surface that closely matches the real skin surface when the skin is clearly visible and passes below the visible hairs when the hair is dense.*

2 Related Work

Because hair and faces are very important parts of virtual human characters, a good deal of effort has gone into methods for capturing models of the surface of the face, and of volumes of hair.

Capturing face geometry was one of the first applications of laser triangulation scanners in computer graphics. For example, a Cyberware face scanner was used to construct models for the groundbreaking computer-generated visual effects in the 1989 film *The Abyss*. The technology for scanning faces has improved since, and faces have been captured by faster and higher resolution laser scanners [Blanz and Vetter 1999]; structured white light scanners [Debevec et al. 2000; Weyrich et al. 2006; Alexander et al. 2009], gradient-based illumination [Ma et al. 2007; Ghosh et al. 2011], and more recently, multi-view stereo systems [Furukawa and Ponce 2009; Bradley et al. 2010; Beeler et al. 2010; Beeler et al. 2011]. To capture the skin surface in this paper, we use the method of Beeler et al. [2010].

Driven by the difficulty of modeling hairstyles realistically by hand, various research has worked on capturing models of scalp hair. Grabli et al. [Grabli et al. 2002] introduced the idea of identifying 3D hair orientation from moving-light image sequences by the specular reflection peak and image-space orientation. Several other projects have since addressed the problem of modeling an entire head of hair from many views, with or without structured or controlled lighting [Paris et al. 2004; Wei et al. 2005; Paris et al. 2008]. All these methods have focused on creating a model of the large-scale geometry, normally by constructing a smooth vector field in 3D that represents fiber orientation and then growing random fibers through this field. Because they focus on capturing whole-head hairstyles, capturing each hair would be prohibitive at the required scale. Recently Jakob et al. [2009] took a different approach, using depth-from-focus and triangulation to capture the geometry of individual fibers in a smaller area of hair. In this paper we take the latter approach, since we are capturing smaller hair volumes than the traditional hair capture papers with their goal of capturing whole heads of long hair.

3 Overview

This section contains a system overview for the 3D reconstruction of the skin episurface plus individual hairs. Many traditional 3D reconstruction algorithms treat the whole of a scene in an undifferentiated way. In contrast, the algorithm described here works by explicitly differentiating the skin episurface and the hair, and processing each of them in a distinct way.

The pipeline is -

(a) separate hairs and skin in the captured images, extract 2D hair fibers using a growing algorithm and remove the detected hairs from the images using inpainting;

(b) reconstruct, filter, refine and grow 3D fibers in 3-space based on the extracted 2D fibers using multi-view stereo (MVS);

(c) compute the skin episurface combining traditional MVS and the estimated roots of the 3D fibers;

(d) synthesize hairs in areas where image data indicates the presence of hair, but individual hairs are indistinct and cannot be reconstructed.

Figure 3 illustrates the pipeline. Section 4 will describe stages (a) and (b), and Section 5 will describe stages (c) and (d).

A preliminary step to the processing below is to obtain a first estimate of the 3D model of the face using the method in [Beeler et al. 2010], without any use of special algorithms for facial hair. We will use the terminology *shrink-wrap surface* below to refer to this first estimate of the 3D model, because it has the effect of wrapping the surface around the hairs. The shrink-wrap model will be used in Sections 4 and 5.
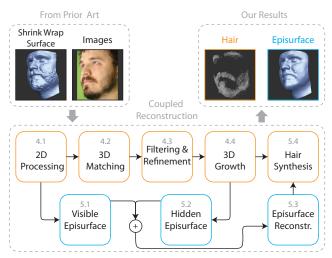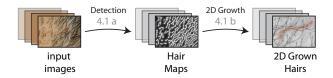
Figure 3: *Main stages of the algorithm. A preliminary step to the processing is to obtain a first estimate of the 3D model using the method in [Beeler et al. 2010]. The algorithm then explicitly detects and reconstructs hair fibers (top flow) and uses this information to provide a better estimate of the underlying surface (bottom flow). This surface, called the skin episurface, is in return used by the hair reconstruction to synthesize new hairs.*

4 Computing 3D Hair

This section describes the computation of hair maps in the image, stereo matching to compute 3D hairs, and a final cleanup step to remove outliers and refine the recovered 3D hairs. We perform the reconstruction independently for left and right sides of the face and merge the hairs prior to the refinement. To increase robustness we start by reconstructing long hair in a first pass (> 5 mm) and shorter hair in a second.

4.1 2D Processing

The first stage of the processing is to detect hair in the images, and seek piecewise linear segments as the first step towards obtaining long fibers of hair. An overview of this process is shown in Figure 4.
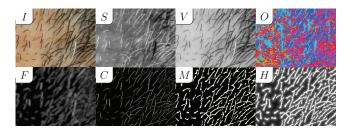

Figure 4: *2D Processing. Given a set of input images, each image is filtered in order to produce a hair map H (4.1a) used to grow 2D-hair fibers (4.1b).*

(a) Computing a 'Hair Map' for each Image

Our experience is that neither hair color nor hair diameter are uniform across a subject's face. These properties moreover vary along an individual hair, which precludes approaches that assume uniformity. We observed that hair and skin exhibit larger contrast in saturation and value than in hue. Thus images are converted from RGB to HSV space, and the S and V channels are used to discriminate hair. Figure 5 shows the original image I in RGB as well as the S and V channels.

Paris et al. [2004] show that oriented filters are well suited to estimating the local orientation of hair. They employ different filters at multiple scales and determine the best score based on the variance of the filters. This provides an efficient way to estimate a dense orientation field for a dense hair volume. For sparse hair, the situation is different as the hair fibers cover only parts of the image and we need to identify which parts. As we have a good prior on the size of the structural element (the hair thickness) we use only a single oriented filter kernel. An oriented filter kernel K_θ is a kernel that is designed to produce a high response for structures that are oriented along the direction θ when it is convolved with an image. We tried several different filters, such as Gabor and Second Order Gaussians, and found their performance to be very similar in practice. In the following, we use the real part of a Gabor filter. The wavelength λ and standard deviation σ of the filter are set according to the expected hair thickness ($\lambda = 4$, $\sigma = 3$ pixels).

Figure 5: *Images used during reconstruction: I: input image, S: saturation channel, V: value channel, O: orientation map, F: Gabor filter response, C: confidence map, M: binary mask, H: hair map.*
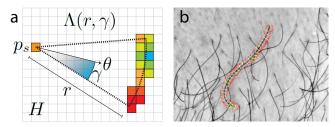
The S and V channels are both convolved with the filter kernel K_θ for different θ (we use 18 different orientations, one every 10 degrees) and the orientation $\tilde{\theta}$ that produces the highest score $F(x, y) = |K_{\tilde{\theta}} * V|_{(x,y)} + |K_{\tilde{\theta}} * S|_{(x,y)}$ at a pixel (x, y) is stored in the orientation map $O(x, y) = \tilde{\theta}$. As can be seen in Figure 5 (F) the filter generates ringing artifacts around the true location of the hair. To suppress these artifacts we propose a non-maximum suppression strategy [Canny 1986]. The pixel is suppressed unless its score is the maximum score in direction orthogonal to θ over the extent of the filter. The resulting confidence map C is shown in Figure 5 (C).

The confidence map C is thresholded using hysteresis [Canny 1986]. The upper and lower thresholds applied are 0.07 and 0.05. From the binary mask M a hair map H is computed according to $H(x, y) = 1/(1 + d(x, y))$ where $d(x, y)$ is the euclidean distance at (x, y) to the closest foreground pixel in M. The hair map H shown in Figure 5 (H) has a value of 1 where hairs are suspected. The value decays quickly when moving away from suspected hair pixels allowing for accurate and robust matching. The hair map is of central importance in matching and growing and is a key difference as opposed to other multi-view stereo systems which usually rely solely on intensity variation. The appearance of hair fibers in different viewpoints can vary substantially due to specularity, translucency, camera focus and of course occlusion through other fibers. The hair map proves to be more reliable under these conditions and permits robust matching of the hair fibers.

(b) Growing Hair in 2D

At the scale we are capturing hair, fibers are essentially one-dimensional structures. Thus the only reasonable neighborhood suited for matching is a one-dimensional neighborhood along the hair itself. We start reconstruction by identifying this neighborhood in an image using a line growing algorithm.

Hair growing in 2D produces a chain of 2D hair segments. A hair segment $s(p_s, \theta, \ell)$ is a linear segment of length ℓ starting at p_s

a b

$\Lambda(r,\gamma)$

p_s

H

Figure 6: *2D hair growth. Hair growing in 2D makes use of the hair map H. (a) Starting point p_s and an initial estimate of the growth direction θ are given by the previous segment. The apex of a growth cone $\Lambda(r,\gamma)$ with growth resolution r and opening angle 2γ is placed at p_s and oriented along θ. For all possible target pixels a score is computed and the pixel with highest score is added to the hair. This process is repeated until the matching score drops below a threshold. (b) An example of a traced hair overlaid the input image.*

in direction of θ. $\mathcal{P}(s)$ denotes the set of pixels covered by the segment s. Growing is based on the hair map H and works as follows. First, a pixel $p_s = (x_s, y_s)$ in H with $H(p_s) = 1$ that is not yet part of a hair is selected as seed and the growing direction is determined from the orientation map O. We define a growth cone $\Lambda(r,\gamma)$ with growth resolution r and opening angle 2γ. The cone defines a set of possible next segments that form an angle of less than γ with the axis of the cone and whose distance to the apex is r. See Figure 6 for a schematic. The parameters chosen for the growth cone are $r = 10$ pixels and $\gamma = 60$ degrees. When growing a segment, the axis of the growth cone is oriented along the direction θ of the last segment and the apex of the cone is placed at p_s. For each potential growth direction in the growth cone a score is computed as

$$\xi(d\theta) = \left(1 - \frac{|d\theta|}{2\gamma}\right)\psi(\mathcal{P}(s_{d\theta})), \qquad (1)$$

where $d\theta$ is the angular direction relative to the axis of the cone and $\mathcal{P}(s_{d\theta})$ denotes the set of pixels covered by the segment $s_{d\theta}$. The function ψ is given as

$$\psi(\mathcal{P}) = \frac{1}{\|\mathcal{P}\|} \sum_{p_i \in \mathcal{P}} \frac{H(p_i) - \nu}{1 - \nu} \qquad (2)$$

and will also be used in later sections when operating on the hair map. The parameter ν is defined within $[0, 1[$ and controls how tolerant the score is. The higher ν the more restrictive the score is regarding deviation from the detected hair — but at the same time it will also become less robust. We set $\nu = 0.4$. The pixel $\tilde{p}$ that produces the highest score is kept as the next segment of the hair and the process is repeated until $\xi(\tilde{p})$ falls below a given threshold. The threshold was set to 0.5. See Figure 6(b) for an example of a grown 2D hair.

4.2 Matching Hair Segments in 3D

In Section 4.1 hair segments were computed separately in each image. This section describes the use of the detected segments in matching across images.
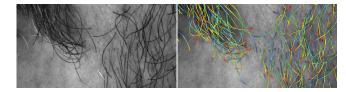
Once the neighborhood has been established, the hair segment is matched in 3D using multi-view stereo. The view in which the 2D hair was grown will be referred to as the reference view $\tilde{c}$. The cameras used for matching are denoted as $\mathcal{C}$. A point p in the reference view $\tilde{c}$ describes a ray $r(p)$ in space. Given a constrained search

space in depth, either computed from the shrink-wrap surface or given by the calibration, the ray $r(p)$ is constrained to a line segment. The projection of this 3D line segment generates an epipolar line segment in every other view $c \in \mathcal{C}$. From the view containing the longest epipolar line segment we find the set of potential 3D positions by creating rays through every pixel on the epipolar line segment and intersecting them with $r(p)$. These potential 3D positions are converted into potential depths d_j along $r(p)$. We then sample the hair segment and compute a matching matrix that contains for every sample p_i at every depth d_j the highest matching score ζ_{ij}. The matching score is computed from the hair map H using Equation 2 and the V-channel of the image as

$$\zeta_{ij} = \left(\frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\psi\left(P_{ij}^c\right)\right)\left(1 - \frac{1}{|\mathcal{C}|}\sum_{c\in\mathcal{C}}\left\|V(P_{ij}^c) - V(P_{ij}^{\tilde{c}})\right\|\right),$$

$$(3)$$

where P_{ij}^c is shorthand for the projection of the point P_{ij} into camera c. P_{ij} is the point on the ray $r(p_i)$ at depth d_j.
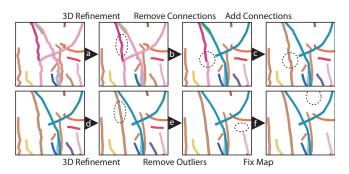
The longest contiguous ridge in the matching matrix is detected and kept as 3D hair consisting of a piecewise linear chain of 3D hair segments. A 3D hair segment is defined either via start and end points as $S(P_0, P_1)$ or via length ℓ from P_0 in direction ω as $S(P_0, \omega, \ell)$. Both notations will be used in the following. Figure 7 shows the projections of the reconstructed hair into one of the views.

Figure 7: *The reconstructed hairs projected into one of the views.*

4.3 Refinement and Outlier Removal

The processing so far has generated 3D piecewise linear segments for the hairs, with connectivity between the hair segments. Because of the discrete nature of the matching as well as noise and other imperfections, these hair segments are jagged and contain outliers. This section describes a refinement process, including a refinement of the computed 3D data, and a reanalysis of the 3D connectivity. The final step is the removal of 3D outliers. An overview of the individual steps is given in Figure 8.

Figure 8: *Overview of the refinement and outlier removal steps described in Section 4.3.*

The following steps are carried out in sequence:

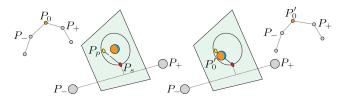**(a) 3D Refinement of Computed Hair Segments.**

This step does a more careful computation of the 3D data for the hair segments, taking the existing 3D hair segments as the start point for the refinement. The computation is a minimization involving a data term that seeks consistency between the images, and a smoothness term that seeks low curvature of the reconstructed hair segment. See Figure 9 for a schematic. A point P_0 on the hair with neighbors P_- and P_+ is refined on the plane normal to $P_+ - P_-$. An update is computed as

$$P_0' = \frac{wP_p + \lambda P_s}{w + \lambda}, \tag{4}$$

where P_p denotes the position that highest data fidelity in a local neighborhood defined by the resolution τ and P_s denotes the position that has highest smoothness given by the projection of $0.5(P_- + P_+)$ into the neighborhood. λ is a regularization parameter and the weight w is computed as

$$w = \frac{\Xi(S(P_-, P_p)) + \Xi(S(P_p, P_+))}{\Xi(S(P_-, P_s)) + \Xi(S(P_s, P_+))} - 1, \tag{5}$$

where Ξ is the matching score defined in Equation 6. The refinement is run for 20 iterations with parameters $\tau = 0.1$mm and $\lambda = 0.01$.

Figure 9: *The refinement stage. The point P_0 with neighbors P_+ and P_- is refined on the plane normal to $P_+ - P_-$. The refinement computes within a local neighborhood the point P_p that has highest data fidelity and the point P_s that has highest smoothness. The refined position P_0' is computed as the weighted average of these two points as described in Section 4.3(a).*

(b) Removal of Low-Confidence Connectivity.

Due to the projection of the hairs into the image plane, it may happen that the 2D hair tracing algorithm (Section 4.1) traces multiple hairs as a single one. If these hairs differ in depth, the 3D matching (Section 4.2) will only match and reconstruct one of the hairs. If they do not differ in depth but only in direction, then there will be a point of direction change where one touches the other. This step removes the connectivity of two connected hair segments if the difference in orientation computed by the scalar product is above a fixed threshold ($45°$).

(c) Addition of New Connectivity. This step involves an explicit search for additional connectivity among the hair segments. Figure 10 illustrates the three cases. Firstly, two segments are marked as connected if they satisfy these conditions: the segments have unconnected endpoints with the segments on opposite sides of those endpoints, and the unconnected endpoints are close in space, and the segments have consistent direction. Secondly, two segments are marked as connected if they satisfy the following conditions: the segments have unconnected endpoints with the segments on opposite sides of those endpoints, and the segments are overlapping and have consistent direction. Thirdly, shorter hairs that are completely enclosed by longer hairs are merged into the longer ones.

We allow linking hairs whose tips are closer than 1mm and enclose an angle of $< 20°$. We allow merging segments that are closer than 0.1 mm and enclose an angle of $< 20°$.

Figure 10: *Identification of connectivity between hair segments. (a) Hairs with tips that are spatially close and enclose a small angle are linked. (b) Hairs that have overlapping parts are merged. (c) Hairs that fall completely into an other hair are removed.*

(d) Repeat Refinement. Step (a) is repeated.

(e) Removal of 3D Outliers. Outlier removal is done by creating a grid for the 3D workspace, counting the number of hairs in each voxel, and deleting hairs which are distant from the surface and in voxels with a low count (distant from other hairs). This is a basic approach but sufficient for the kind of outliers that are observed— single isolated hairs away from the true surface.

(f) Fix Map. The image structure in the iris or the lips is locally very similar to hair and so the system might reconstruct outlier hairs in these areas. We remove these outliers by drawing two 'Fix Maps', one for a center camera and one for a camera from below. These Fix Maps are very sketchy binary masks that can be created rapidly. Figure 11 shows example Fix Maps. Hairs that get projected mostly into the masked areas are removed. This is the only manual step in our pipeline and could be automated using feature detectors.
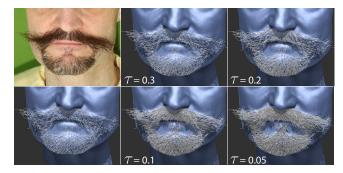

Figure 11: *Fix Maps used to identify areas like eyes or lips where no hairs should be detected. We provide Fix Maps for the two shown viewpoints for all subjects. This is the only manual step in our pipeline and could be automated using feature detectors.*

4.4 Growing Hair in 3D

Hair growing in 3D is performed in a similar way to hair growing in 2D (Section 4.1). A three-dimensional growing cone $\Lambda(r, \gamma)$ is used to determine potential segments. We set the growth resolution to $r = 1$mm and $\gamma = 30°$. The apex P_s of the cone is placed at the tip of the last segment and the axis is aligned with its direction $\omega = (\theta, \phi)$. For every potential segment $S_\omega = S(P_s, \omega, r)$ of the growth cone with direction $\omega = (d\theta, d\phi)$ a score is computed using all cameras $\mathcal{C}$ where the segment is expected to be visible

$$\Xi(S_\omega) = \frac{\sum_{c \in \mathcal{C}} \rho_\omega^c \psi(\mathcal{P}(S_\omega^c))}{\sum_{c \in \mathcal{C}} \rho_\omega^c} \tag{6}$$
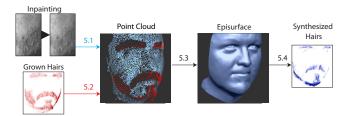
where S_ω^c is shorthand for the projection of the segment S_ω into camera c, $\mathcal{P}$ denotes the pixels spanned by this projection and ρ_ω^c denotes the angle the direction ω encloses with the optical axis of the camera c. The growth is terminated when the score value drops below a user defined threshold τ. The effect of varying τ can be seen in Figure 12. We used values from 0.1-0.3 for the examples in this paper.

Figure 12: *The effect of varying the threshold τ which terminates the hair growth in 3D. The left column shows the input image (top) and reconstructed hair fibers (bottom). The center and right columns show different growth results obtained by varying the threshold τ.*

5 Computing Skin Episurface

The concept of the skin episurface was introduced in Section 1. This section describes its computation, as depicted in Figure 13.

Figure 13: *Episurface reconstruction overview. The episurface is reconstructed from the inpainted images (blue) where the surface is visible and from the grown hairs (red) where it is covered with hair. From the episurface, additional hairs can be synthesized.*
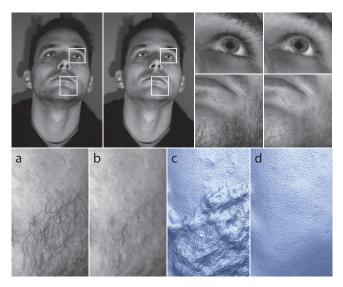
5.1 Computing Visible Skin Episurface

The part of the skin episurface exposed to the cameras is recomputed using the first part of the method proposed in [Beeler et al. 2010]. This part performs pair-wise stereo reconstruction and produces a 3D point cloud. The difference from reconstructing the shrink-wrap surface is that we now know where hair is to be expected. Using this knowledge we prepare the data as follows:

- Masking - Areas that contain denser hair are masked and therefore excluded from reconstruction. We employ an opening operation followed by a closing operation on the mask to only exclude areas that have considerable amounts of hair. Individual hairs will not be affected as they are removed by the inpainting step.

- Inpainting - We employ a Gaussian filter to the image data with spatially varying, anisotropic kernel. The orientation of the filter is given by the orientation map O and the spatial extent σ is computed depending on the hair map H. In areas where H is low, the spatial extent will be low as well ($< 1px$) and no filtering occurs. In areas where H is high the image is blurred, effectively inpainting individual hairs. This reduces the strong image gradients produced by individual hairs and prevents matching.

The reason for using inpainting instead of masking out individual hairs is that the method proposed in [Beeler et al. 2010] uses image

pyramids to compute the stereo reconstruction. While the images are subsampled linearly, the masks are subsampled using nearest neighbor, which leads to a discrepancy of mask and image at higher layers of the pyramid. Inpainting circumvents this. The effect of inpainting is demonstrated in Figure 14.

Figure 14: *Effects of inpainting. Top row: inpainting reduces the presence of hair while maintaining other facial features. Bottom row: inpainting of sparse hair prevents the stereo reconstruction from creating artifacts in these areas. (a) The captured image; (b) the effect of inpainting as described in Section 5.1. (c,d) The reconstructions based on the original and inpainted images. Note how the areas of visible skin are the same, while (d) does not exhibit the shrink-wrap artifacts in areas that contain hair.*

5.2 Estimating Hidden Skin Episurface

The part of the skin episurface that is not exposed to the cameras is estimated in the following way:

- Take each 3D hair computed in Section 4. The root of the hair is designated as the end which is furthest below the surface of the shrink-wrap model described in Section 3. For short whiskers, both ends of the hair may be close to the surface. In this case, it is arbitrary which end will be chosen as the root, but there is no adverse effect on the subsequent computation. Hairs that are distant from the surface are not considered.

- For each hair root, find neighboring roots within a pre-defined search radius. Do a least-squares fit of a plane to the root and its neighbors, to estimate the surface normal at the root.

- Collect, for all roots, the 3D coordinates and estimated surface normals.

This step produces a point cloud which is a sampling of the underlying hidden episurface.

5.3 Computing Skin Episurface

The previous two steps produce two sets of points for the visible and the hidden skin episurface, resp. See Figure 15(b). These are combined and serve as input for the second part of [Beeler et al. 2010]. This part performs a Poisson reconstruction [Kazhdan et al. 2006] followed by a refinement step. We change the refinement step to incorporate the hair map as regularization prior, preferring

smoother solutions in areas where hair is expected. Figure 15(d) shows the final episurface compared with the initial shrink-wrap surface.
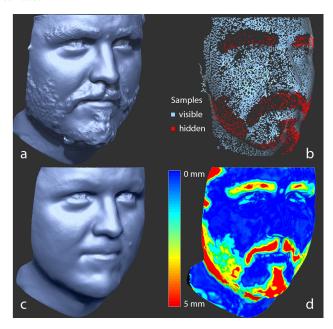

Figure 15: *The construction of the episurface. (a) Shrink-Wrap surface produced by the method of Beeler et al.; (b) the point cloud of the visible episurface in blue (Section 5.1) and the points from the hidden episurface in red (Section 5.2); (c) the final reconstructed episurface (Section 5.3); and (d) visualizes the difference between the two meshes. The two surfaces are almost identical in areas of clear skin but the episurface provides a much smoother hypothesis in areas with hair coverage.*
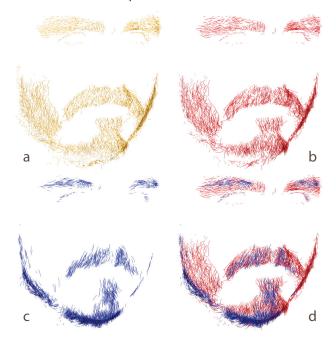
5.4 Synthesizing Hair

This section describes how the 3D hairs computed in Section 4 can be augmented in a physically plausible way. The main goal of hair synthesis is to achieve greater density of 3D hair in areas where image data indicates the presence of hair, but individual hairs are indistinct and cannot be reconstructed. Hair synthesis happens in two steps: first, seeds are found on the episurface, and second, the hairs are grown starting from these seeds.

5.4.1 Finding Seed points

Finding seed points starts by projecting the reconstructed 3D hairs into the cameras to prevent finding seeds in areas where hair has already been reconstructed. Next every vertex of the episurface is projected into the images. Values $\bar{H}^c$ are computed for all cameras c by averaging the hair map H within a window. The vertex is chosen as hair seed if $\bar{H}^{\tilde{c}} > \alpha$ and $\frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \bar{H}^c > \beta$, where $\tilde{c}$ is the camera that has the least foreshortening with the vertex. The parameters α and β are set to 0.3 and 0.2. If a vertex has been selected as seed it will prevent other vertices in its neighborhood from becoming seed points.

5.4.2 Growing Hair

The seeds found in the previous step serve as starting points for hair growth. The default growth direction is normal to the surface and the maximal growth length is the average length of all reconstructed sample hairs. These default properties are blended with

Figure 16: *The individual stages of the hair reconstruction. (a) The 3D reconstructed hair after the matching stage (Section 4.2); (b) the hair after the growing stage (Section 4.4), (c) shows the synthesized hair (Section 5.4), and (d) the final composition of (b) and (c).*

properties sampled from neighboring hairs that were successfully reconstructed. The properties of the sample hairs and the default growth properties are interpolated using Gaussian RBFs. This leads to a smooth interpolation of growth properties that is faithful to the sample hairs where they could be reconstructed and plausible in areas of dense hair. Using these properties a hair is grown as described in Section 4.4 with the difference that the orientation maps are also consulted to give a sense of directionality for the growth. The growth properties are updated as a hair grows to ensure that it remains faithful to the style of the reconstructed fibers within its current neighborhood. The effect of hair synthesis based on reconstructed sample hairs can be seen in Figure 16(c). Hair whiskers in the eyelashes are short and point directly away from the surface, while the hair fibers in the mustache are longer and follow the overall direction in this area.

6 Results

Setup We captured all examples using the setup shown in Figure 17. The setup consists of 8 Canon Rebel T1i cameras designated to capture the full face plus 6 Canon Rebel T2i cameras that aim at capturing the chin area with higher resolution. The T1is are equipped with 85mm lenses and are arranged in pairs of two around the subject — one pair on each side, one straight on and one from below. The T2is are equipped with 100mm macro lenses and are arranged in tuples of three on both sides of the subject. The skin reconstruction only uses the T1is while the hair reconstruction makes use of all cameras. The subjects are illuminated with 4 Canon flashes (430EX,580EX) and we use cross polarization to remove specularities. Only a single frame per camera is required and thus the setup can acquire the data within fractions of a second. Reconstruction with our unoptimized prototype pipeline takes approximately 45 minutes on a Mac Pro Desktop computer (8 cores).
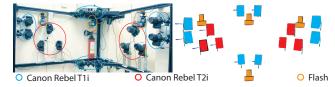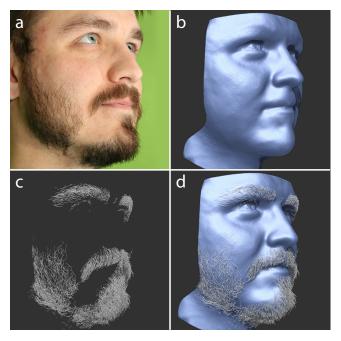
Figure 17: *Setup used to capture the data. The setup consists of 8 Canon Rebel T1i cameras equipped with 85mm lenses to capture the full face and 6 Canon Rebel T2i cameras equipped with 100mm lenses to capture the chin area. The subject is illuminated with 4 Canon Flashes (430EX,580EX).*

Recon.	4160	5003	4193	5811	4565	3552	6851	2779
Synth.	104	179	694	973	4963	1831	494	654

Table 1: *Number of hair fibers reconstructed (top row) and synthesized (bottom row) for the subjects displayed in the teaser. The amount of synthesized hair is typically low, except for subjects that have dense hair coverage.*
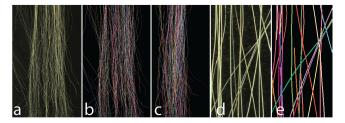

Figure 18: *Individual steps of the reconstruction pipeline. (a) Raw image; (b) reconstructed skin episurface; (c) reconstructed hair; (d) skin episurface plus hair.*

Figure 21: *Reconstruction of the skin episurface for a tufty beard. Note that the skin episurface is a pseudo-surface underneath the uppermost layer of hair and is not expected to approximate the shape of the underlying skin in the case of protruding facial hair.*

Figure 18 shows the individual steps of the pipeline for a selected subject. Figure 19 shows results for various different styles of facial hair demonstrating the robustness of the approach and Figure 20 shows selected details and demonstrates the quality of the reconstruction. Table 1 lists the number of hair fibers reconstructed and synthesized for the subjects shown in the teaser.

Finally Figure 21 shows the reconstructed episurface for a shaped tufty beard. The shape approximates the beard rather than the underlying skin surface in this case. This is not an unexpected result because the episurface is not formulated with the idea of recovering underlying face shape. It would be an interesting line of future research to provide episurfaces that have the goal to be anatomically plausible.

7   Conclusion

We have presented an algorithm for face capture that successfully reconstructs facial hair fibers as well as the face's underlying skin surface. We show that treating skin and hair in a coupled fashion delivers accurate reconstruction in areas of high visibility, and gives plausible results in areas of dense occlusion. We demonstrate reconstructions of a number of individuals exhibiting a variety of

facial-hair styles. The impact of our work is reflected by the significance of facial hair in our cultural heritage. Our image of many historic figures is dominated by their facial-hair features, including Albert Einstein's bushy white whiskers, Abraham Lincoln's characteristic beard, and Salvador Dali's distinctive mustache [Dalí and Halsman 1954] (now eponymously known as a "dali"). Today, facial hair remains at the core of individual expression, as evidenced by the ever changing popularity of different facial-hair styles. Our work provides a means to capture this piece of popular culture for use in contemporary applications as well as accurate preservation for future generations.

Limitations of our system direct us to areas of future work. Our algorithm delivers the best results in the presence of short, sparse hairs. In areas of dense hair, such as a long, thick beard, the degree of occlusion can be so great that accurate hair and skin reconstruction is not feasible. Future work could combine algorithms that target long, whole-head hair styles [Paris et al. 2004; Wei et al. 2005; Paris et al. 2008] with our facial hair-system to deliver high-quality results even for long, thick beards, mustaches, or side burns. Preliminary tests on synthetic hair (Figure 22) show that the presented method works well with long hair fibers, which would allow reconstructing the outermost hair layer of whole-head hair styles. There could be a tremendous benefit in terms of overall realism in reconstructing the outer hair layer exactly using our method, and augmenting this result with the powerful estimation methods of whole-head hair capture systems. Our hair-detection algorithm is limited by the amount of contrast in the camera images, and skin-colored hairs may be missed. Likewise, dynamic range can be an issue in cases where the skin and hair vary greatly in brightness (a white beard on black skin, or vice versa) due to exposure limitations. Extending our current algorithm to be robust against contrast and dynamic range limitations is another area of future work. We focus on geometry capture, and only include a limited amount of color information in our results. A thorough treatment of skin and hair appearance capture under varying lighting conditions offers many challenges for future work.

Other hair features such as shape or thickness could also be reconstructed and incorporated in our example-based hair-synthesis algorithm.

Figure 22: *Synthetic hair test. Given a sample of long synthetic hair (a), the system reconstructs the 3D hairs shown in (b) and (c). A close-up of the input image and the corresponding reconstructed hairs are shown in (d) and (e). Notice how most fibers are correctly reconstructed.*

Extending the work to capture hair dynamics would be a challenging and very interesting line of future research. Spatio-temporal capture and reconstruction would on the one hand increase the complexity of the problem, but on the other hand also introduce additional data and constraints that could be leveraged. A further topic of future research is given by the episurface concept as it would be interesting to investigate ways to provide episurfaces which are anatomically correct. This could be achieved by incorporating prior knowledge, e.g. in the form of a morphable model [Blanz and Vetter 1999].

Perhaps the most exciting area of future work lies in extending hair and skin surface capture beyond the face, to include the entire human body. Doing so will permit the capture of the human form at a level of fidelity not yet possible. An even more far-reaching goal lies in moving beyond humans to other species. A characteristic feature of all mammals is the presence of hair. This huge range—from the soft coat of a cat, to the wiry bristles of an elephant, to the dense, waterproof fur of a sea otter—provides an exciting and compelling long-term target for future work in hair and skin surface capture.

Acknowledgements

References

ALEXANDER, O., ROGERS, M., LAMBETH, W., CHIANG, M., AND DEBEVEC, P. 2009. The digital Emily project: photoreal facial modeling and animation. In *ACM SIGGRAPH 2009 Courses*, ACM, New York, NY, USA, SIGGRAPH '09, 12:1–12:15.

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

BEELER, T., HAHN, F., BRADLEY, D., BICKEL, B., BEARDSLEY, P., GOTSMAN, C., SUMNER, R. W., AND GROSS, M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (Aug.), 75:1–75:10.

BLANZ, V., AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '99, 187–194.

BRADLEY, D., HEIDRICH, W., POPA, T., AND SHEFFER, A. 2010. High resolution passive facial performance capture. *ACM Trans. Graph. (Proc. SIGGRPAH) 29* (July), 41:1–41:10.

CANNY, J. 1986. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Trans. PAMI-8*, 6 (nov.), 679 –698.

DALÍ, S., AND HALSMAN, P. 1954. *Dali's mustache: a photographic interview*. Flammarion.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, SIGGRAPH '00, 145–156.

FURUKAWA, Y., AND PONCE, J. 2009. Dense 3D motion capture for human faces. In *Computer Vision and Pattern Recognition*, 1674 –1681.

GHOSH, A., FYFFE, G., TUNWATTANAPONG, B., BUSCH, J., YU, X., AND DEBEVEC, P. 2011. Multiview face capture using polarized spherical gradient illumination. ACM, New York, NY, USA, vol. 30, 129:1–129:10.

GRABLI, S., SILLION, F. X., MARSCHNER, S. R., AND LENGYEL, J. E. 2002. Image-based hair capture by inverse lighting. In *Proceedings of Graphics Interface (GI)*, 51–58.

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

KAZHDAN, M., BOLITHO, M., AND HOPPE, H. 2006. Poisson surface reconstruction. *Proceedings of the fourth Eurographics symposium on Geometry processing*, 61–70.

MA, W.-C., HAWKINS, T., PEERS, P., CHABERT, C.-F., WEISS, M., AND DEBEVEC, P. E. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. *Rendering Techniques 9*.

PARIS, S., BRICEÑO, H. M., AND SILLION, F. X. 2004. Capture of hair geometry from multiple images. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3 (Aug.), 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. (Proc. SIGGRAPH) 27*, 3 (Aug.), 30:1–30:9.

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

WEYRICH, T., MATUSIK, W., PFISTER, H., BICKEL, B., DONNER, C., TU, C., MCANDLESS, J., LEE, J., NGAN, A., JENSEN, H. W., AND GROSS, M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3 (July), 1013–1024.
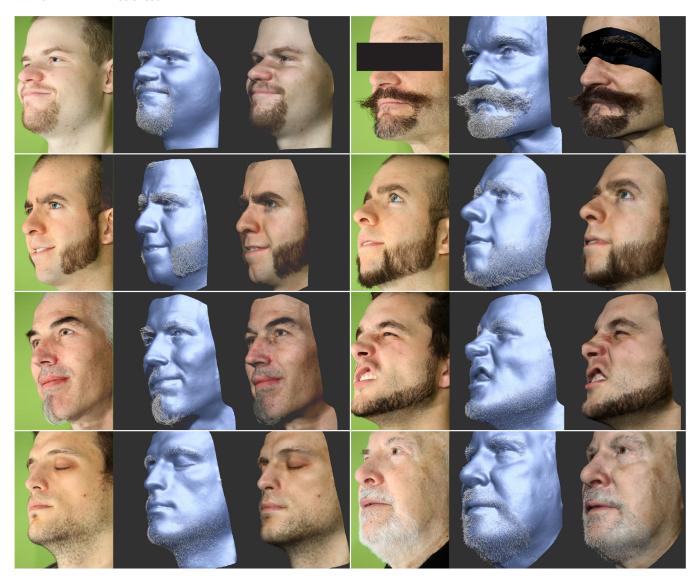
Figure 19: *Reconstructed models for a variety of subjects demonstrating robust performance for different facial hair stylings.*

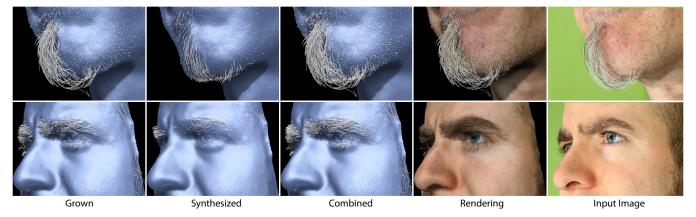| Grown | Synthesized | Combined | Rendering | Input Image |

Figure 20: *Close-up comparison between different stages of the reconstructed geometry and the real photographs. From left to right: reconstructed hair fibers, synthesized hair fibers, final hair fibers, realistic rendering and real photograph. Note how the overall structure is captured well by the algorithm. Where individual hairs are visible, the algorithm correctly reconstructs them and in areas with dense hair coverage the synthesis provides a plausible volume of hair fibers. Likewise the surface is reconstructed with high quality in areas with no or little hair coverage and provides a plausible substrate in areas with dense hair.*