Single Photo Estimation of Hair Appearance

Nicolas Bonneel[1], Sylvain Paris[2], Michiel van de Panne[3,1], Frédo Durand[4], George Drettakis[1]

[1]REVES/INRIA Sophia-Antipolis, [2]Adobe Systems, Inc.
[3]University of British Columbia, [4]MIT CSAIL

Abstract

Significant progress has been made in high-quality hair rendering, but it remains difficult to choose parameter values that reproduce a given real hair appearance. In particular, for applications such as games where naive users want to create their own avatars, tuning complex parameters is not practical. Our approach analyses a single flash photograph and estimates model parameters that reproduce the visual likeness of the observed hair. The estimated parameters include color absorptions, three reflectance lobe parameters of a multiple-scattering rendering model, and a geometric noise parameter. We use a novel melanin-based model to capture the natural subspace of hair absorption parameters. At its core, the method assumes that images of hair with similar color distributions are also similar in appearance. This allows us to recast the issue as an image retrieval problem where the photo is matched with a dataset of rendered images; we thus also match the model parameters used to generate these images. An earth-mover's distance is used between luminance-weighted color distributions to gauge similarity. We conduct a perceptual experiment to evaluate this metric in the context of hair appearance and demonstrate the method on 64 photographs, showing that it can achieve a visual likeness for a large variety of input photos.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [I.3.7]: Three-Dimensional Graphics and Realism—

1. Introduction

Recent advances improve the rendering of realistic hair using advanced models of scattering [MWM08, ZYWK08, Zin07, MM06, MJC*03]. However, they involve numerous parameters, and matching the hair appearance of a given person is difficult even when the geometry is given. This process is time-consuming and error-prone even for trained artists [MTLB03]. An alternative is to use image-based rendering, but current hair-capture methods rely on complex hardware with many cameras and tens of lights, and produce models with very large storage requirements [P*08].

Our primary goal is to develop a simple, low-cost way to 'close the loop' between modern hair rendering and hair as it is readily observed in the real world or in photos. We abandon the notion of pixel-accurate reconstruction of a given photo and instead take a statistical and perceptual modeling approach. We introduce a method that enables the estimation of hair appearance with a very lightweight structured lighting setup: a single photograph taken with a flash on the camera. We use only approximate knowledge of the hair geometry. The simplicity of the setup is of particular importance for the user-driven creation of avatars and characters in games and interactive simulations, which we envisage as a prime application area. Recently, hairstyle and hair color have been reported to be among the most important features for avatar personalization [DWYW09].

The acquisition of hair geometry is beyond the scope of this paper and we assume the existence of a small set of macroscopic models that define 3D hair-strand geometries for distinct classes of hairstyle, e.g., long curly hair, straight layered hair, etc. These models (see Fig. 2) are modeled by artists or reuse existing captured hair geometry [P*08]. We distinguish this *hairstyle model* from the remaining parameters, which form an independently-defined *hair appearance model*. Our hair appearance model consists of absorption, reflectance [MJC*03], and geometric noise [Yu01] parameters, which we seek to estimate. A pair of melanin absorption
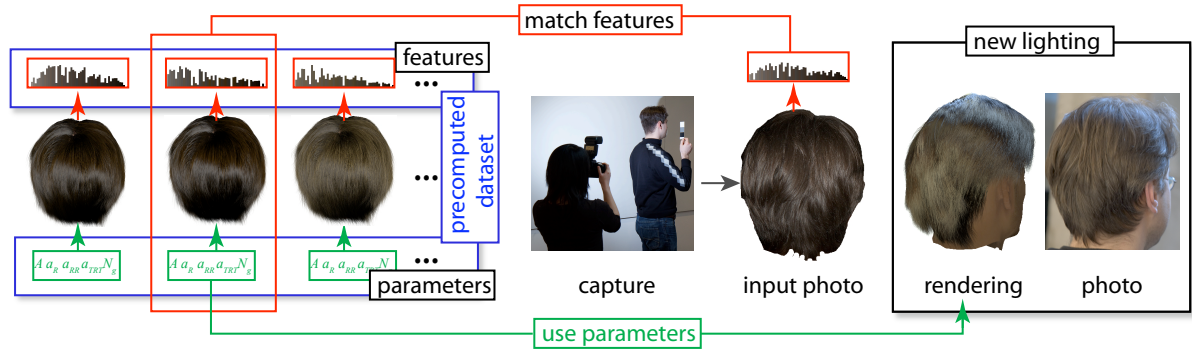
Figure 1: *Appearance Capture. Using a flash photograph as input, we estimate a set of hair appearance parameters that can be used to reproduce the overall likeness of the observed hair. We computed image features and a distance on this feature to match the input photo with a dataset of prerendered images with known appearance parameters. The estimated parameters are taken from the best-matching image, and can now be used to render new images in new conditions.*

parameters determines the overall hair color, and three lobe width parameters define the visual appearance of the specular highlights, glints, and transmissivity effects.

In this work the best-matching hairstyle model is manually selected; we shall see that this choice can impact parameter estimation (§6). It is also possible to assign the same hair appearance to different hairstyle models, although best likeness of a photo is still achieved by rendering with an appropriate match in both the hairstyle and appearance. We also define a *geometric noise* parameter that can have a significant impact on the appearance of hair. Taken together, these parameters define a six dimensional hair appearance parameter, $\mathcal{P}$. For rendering, we use the fast multiple-scattering technique of Zinke et al. [ZYWK08]. We chose this method for efficiency; others could be used instead.

Given an approximate geometry, we wish to match the appearance of the rendering to the appearance of the input flash photo, using the available free parameters in the appearance model (see Fig. 1). We first define a feature that captures the aggregate visual properties of hair, together with a metric on this feature. We propose the use of a luminosity-weighted color distribution in *Lab* color-space as an appropriate image feature, $\mathcal{F} = f(I)$, and an earth-mover's distance, $\delta(f(I_A), f(I_B))$, as a metric between two images in this feature space. Although our method does not produce physically validated parameter estimations, it generally achieves visually plausible results by working within the degrees of freedom available in the rendering model. We validate our image feature and the related metric using a perceptual evaluation.

To match hair appearance, we compute a reference dataset by sampling the parameter space and rendering images of our representative geometric models. For each reference rendering, we compute our feature, which provides us with a large dataset of $(\mathcal{P}, \mathcal{F})$ tuples, i.e., images for which we

know both the parameters and the feature. Given the computed feature for an input photo, we search for the closest feature match in the dataset and return the associated model parameters. The entire process is illustrated in Fig. 1.

Contributions: We present a new approach to estimate hair appearance using a single photo; we achieve this by recasting the process as an image retrieval problem. To do this, we first introduce a hair appearance model with two new components: a melanin-based hair pigmentation model that reproduces the natural subspace of hair absorptions, and a geometry noise parameter. We then introduce an image feature and distance metric for matching hair appearance, and do a perceptual evaluation of these. Lastly, we test the single-photo hair appearance estimation on a set of 64 photos and do robustness evaluations.

2. Related Work

Modeling the geometry and appearance of human hair is challenging. Overviews of the progress that has been made on these problems over the past decade can be found in [WBK*07] and [Ber06]. There is also considerable knowledge about the biology of human hair, including its microstructure, density, growth, response to humidity, and pigmentation [HS01, L'O08]. Hair rendering models have evolved considerably, with increasingly sophisticated modeling of the complex light paths that occur in hair [KK89, MJC*03, MM06, Zin07, ZYWK08, MWM08]. However, setting the various required model parameters remains an unaddressed problem and motivates our work. The intricate geometry of hair and complex light diffusion precludes the use of general appearance analysis methods such as those proposed for BRDF estimation.

There have been several notable efforts to model aspects of hair from images. The work of Grabli et al [GSML02] infers the ambient, diffuse, and specular colors of a wig. However,

real hair fibers are known to have a more complex reflectance than synthetic materials [MJC*03] and thus it is unclear how well this method generalizes. Paris et al. [PBS04] capture the hair geometry from a single, controlled light source and a video camera but the reflectance is not retrieved in this process. Wei et al. [WOQS05] improve this method with an approach that works under ambient lighting. They also map photographic data on the hair geometry to model the hair appearance. These mapped colors are fixed and do not vary with the lighting environment nor the view direction, thereby limiting the rendering to reproducing the captured conditions. Paris et al [P*08] describe an image-based rendering method that yields faithful matches to photographs. This technique requires a large amount of data which would be unwieldy for many applications and requires a complex capture setup.

More broadly related are interfaces that help guide users in parameter selection [M*97], the determination of specular lobe shape from image statistics for linear light sources [GTHD03], skin BRDF models based on pigment concentrations [DWD*08], and the use of multiscale statistics for estimating BTF parameters [ND06].

3. Synthetic Appearance Model

We first discuss the rendering model used and then describe a novel color absorption model based on human hair pigmentation. Lastly, fine-scale geometric distributions can also have a significant visual impact on the appearance of hair. To this end, we define a global *geometry noise* parameter.

3.1. Rendering

For rendering, we use the forward scattering map of Zinke et al. [ZYWK08] because both efficiency and quality of the rendering are of concern in our setting. In this paragraph we summarize the rendering details for completeness.

We use the scattering model presented by Marschner et al [MJC*03]. We split the scattering parameters into two groups: free model parameters and fixed parameters. For the latter, we found that they have the least influence on the overall rendered hair appearance or can be assumed to have typical values [MJC*03]. Our fixed-value parameters are: cuticle angle $(-3°)$; eccentricity (0.9); caustic power (0.4); caustic width $(1.5°)$; fading range for caustic (0.3); index of refraction (1.55); density [ZYWK08] (0.7); and hair radius $(120 \, \mu m)$. The free parameters are the absorption coefficients, A_r, A_g, A_b, and the R, TT, and TRT lobe widths, $\alpha_R, \alpha_{TT}, \alpha_{TRT}$, which intuitively correspond to the "color" and "shininess" of the hair. In our tests, hair radius has been increased compared to values given in the literature to account for the sparsity of our hair models.

To efficiently render images using [ZYWK08], we use a

Figure 2: *Hairstyle geometries used for datasets (without noise), from left to right: straight short hair, straight long hair, straight clumpy hair, wavy hair, tangled hair, long curly hair, very long straight hair*

CPU implementation of the raytracing-based forward scattering map method and parallelize it to exploit multicore architectures. We perform the final rendering step directly on the GPU. With this, 640×480 images are rendered at a rate of 1.2–1.9 images per minute. We choose this over the faster GPU-based algorithm because of the higher quality we obtained from the forward scattering map.

Figure 2 shows the hairstyle geometries that we use: three are modeled using Maya, and four are captured data [P*08] (http://graphics.ucsd.edu/~will/download/HairPhotobooth/). The hairstyles each have 80,000–120,000 hairs.

3.2. Melanin Model

In order to reduce the number of parameters for hair appearance, we exploit knowledge of natural human hair pigment absorption to reduce the number of parameters in the original Marschner model [MJC*03]. Our reparameterization can also greatly facilitate manual hair appearance specification by reducing the number of degrees of freedom and ensuring that the absorption is realistic.

Natural hair color is largely due to wavelength-dependent absorption within hair fibers. Hair pigmentation is composed of two kinds of melanin pigments: *eumelanin* and *pheomelanin* [L'O08, Tob08], and the spectral absorption of these two pigments is known [SS88]. The absorption can thus be modeled as a linear combination of the concentration of these melanins i.e., $A(\lambda) = a_1(\lambda)m_1 + a_2(\lambda)m_2$, where m_1, m_2 are the eumelanin and pheomelanin concentrations and $a_1(\lambda)$, $a_2(\lambda)$ are their spectral absorption curves. We point sample the absorption curves for red, green and blue wavelengths, (575, 535, and 445 *nm*, respectively), and normalize with respect to the r concentration of the melanins, yielding: $A_r = m_1 + m_2$, $A_g = 1.3036m_1 + 1.6390m_2$, $A_b = 2.2272m_1 + 3.8370m_2$. This defines a 2D subspace in the 3D absorption parameter space without reducing the desired expressivity of the model. Because $m_1, m_2 \in [0, \infty]$, it will be convenient to instead represent the melanin concentrations by $\hat{m}_1 = e^{-km_1}$ and $\hat{m}_2 = e^{-km_2}$, where k is experimentally determined (§4.2). This gives finite ranges, $\hat{m}_1, \hat{m}_2 \in [0, 1]$. We shall exploit the melanin model during appearance estimation, where it will help achieve better sampling of the absorption parameters by eliminating unnatural hair colors from consideration, such as green or blue hair.

3.3. Geometry Noise

We distinguish between two different levels of geometry variation. The macroscopic geometry or hairstyle is chosen by the user from a fixed set of models (Figure 2) and captures aspects such as the curliness and length of the hair. However, a direct application of rendering techniques [MJC*03, ZYWK08] to many modeled or captured geometries can yield unrealistic results. Figure 3 (left) shows an example of this for captured geometry [P*08]. This can be rectified using an additional geometry variation based on small-scale noise. We add noise in a manner similar to Yu et al [Yu01]. The perturbation applied to a given vertex v on a hairstrand is given by

$$\Delta v = N_g A V (\alpha \sin(2\pi\alpha p) + \frac{1}{2}e^{3\alpha} - 1) \qquad (1)$$

where $\alpha \in [0,1]$ is the normalized curvilinear abscissa , V is a unit-length random direction, $p \in [0,8]$ is a random frequency, N_g is a noise amplitude, and A is a hairstyle-specific scaling factor that is required to deal with the given modeling scale of any hairstyle. Fig. 3(right) shows the changed appearance with the noise model. Without any geometric noise, the hair can have a noticeable unrealistic plastic-like appearance.


Figure 3: *Impact of geometric noise on hair appearance. Left: Without geometry noise. Right: With geometry noise.*

The final set of free parameters in our appearance model is given by $\mathcal{P} : \{\hat{m}_1, \hat{m}_2, \alpha_R, \alpha_{TT}, \alpha_{TRT}, N_g\}$.

4. Appearance Estimation

Our method relies on a reference dataset of rendered images sampling the parameter space and we seek to find the reference image whose hair appearance is most similar to the input photograph. A schematic illustration of the appearance estimation problem is shown in Figure 4. Both the true hair appearance and the rendered hair appearance can be seen as forming manifolds in a suitably-chosen feature space, $\mathcal{F}$, although we only know the underlying parameterization of the rendered appearance. An input photo may not in general lie on the manifold of images achievable by the rendering, which may in part be attributed to a limited expressivity of the rendering model. Given an input photo, we propose a novel method to estimate its underlying parameters, $\mathcal{P}$, by
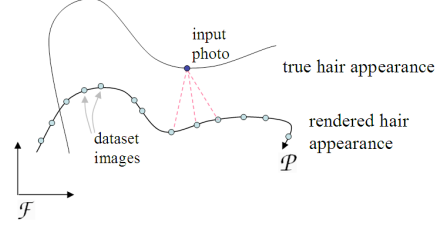
Figure 4: *An abstract view of the appearance estimation problem. $\mathcal{F}$ defines the space of features used to represent the appearance. Rendered images lie on a manifold parameterized by the model parameters, $\mathcal{P}$.*

posing this as a search problem. We search for the point on the rendered manifold which is most similar in appearance to the input photo, and then return its associated parameters.

4.1. Feature Selection and Distance Metric

Image features for matching hair appearance should ideally be easy to compute, induce a perceptually-meaningful metric, work across a large variety of hair appearances, and be somewhat invariant to the specific geometry of the hair. We propose the use of luminance-reweighted color distributions in *Lab* color space as our feature, and the use of Earthmover's distance (EMD) as a distance metric between features. These choices are inspired by the use of color distributions and EMD for Image Retrieval [RTG00]. A modification we propose in the context of the hair appearance estimation is to use a luminance-based reweighting of the color distributions, which we found to yield improved estimates.

The image color distributions are represented by color-clusters in the *Lab* color space. We use k-means clustering with 50 clusters and initialized using k-means++ [AV07]. To achieve fast clustering, a subsampled set of 1 in every 50 pixels is used during the k-means iterations. The final nearest-neighbor assignment is done for all pixels in the image to obtain a pixel count for each cluster. Each cluster's pixel count n_i is reweighted by the cluster luminance L_i, $n_i' = L n_i$. Finally, the resulting cluster pixel counts $\{n_i'\}$ are normalized, i.e., $\hat{n}_i' = n_i'/N$, where $N = \sum_i n_i'$. The final image feature is then a set of tuples, $F = \{(C_i, \hat{n}_i')\}$, where C_i is the *Lab* color of cluster i, i.e., the cluster center. Computing the color clusters for an image takes ~600 *ms*.

The EMD metric minimizes the work required to turn one distribution into another, and is symmetric, i.e., $\delta(\mathcal{F}_a, \mathcal{F}_b) = \delta(\mathcal{F}_b, \mathcal{F}_a)$. The metric is computed by solving a transport problem, where the 'mass' in the source distribution bins needs to be transported with minimal cost to the target distribution bins to exactly fill them. In our problem, the source and target bins correspond to the clusters of $\mathcal{F}_a$ and $\mathcal{F}_b$, and mass corresponds to normalized cluster counts, $\hat{n}'$. As a distance, the EMD returns the minimal total work required,

as defined by $\delta = \sum_{ij} d_{ij} f_{ij}$, where d_{ij} is the *ground distance* between source bin i and target bin j and f_{ij} is the flow (mass) carried between these two bins. We use Euclidean distance in the *Lab* space as our ground distance. The EMD optimization is posed as a linear programming problem and is solved using a streamlined version of the simplex method [RTG00].

4.2. Synthetic Dataset

The synthetic dataset consists of a large set of precomputed tuples, $D : \{(\mathcal{F}_i, \mathcal{P}_i)\}$, with one tuple per rendered image. The final datasets are compact in practice because there is no need to retain the images once its features have been computed. We include a copy of the datasets in the supplementary materials.

We compute multiple datasets, D^k, one for each modeled hairstyle, k, and using 4000–5000 sample points per dataset. We use a simple sampling strategy given that the real distributions of the model parameters for human hair are unknown. We draw random samples using uniform distributions U. The Marschner lobe parameters are sampled using $\alpha_R, \alpha_{TT}, \alpha_{TRT} \sim U[2°, 20°]$. This represents an extension of the typical values given in [MJC*03]. However, we note that these typical value ranges are not necessarily strictly respected in prior art. For example, [MJC*03] uses $\alpha_R = 8°$ and $\alpha_{TT} = 6°$, yielding an α_R to α_{TT} ratio of 1.33 instead of the recommended ratio of 2.0, and [ZYWK08] uses $\alpha_R = 8°$ and $\alpha_{TT} = 10°$ for a ratio of 0.8. Establishing accurate ranges or prior likelihoods for these parameters would require physically-validated measurements for a large set of examples. Noise is sampled using $N_g \sim U[0.3, 1]$.

For sampling melanin concentrations, we use an informed strategy that samples the space of final observed hair colors in an approximately uniform fashion. We build on our observation that the mean color of rendered hair is approximately linearly correlated with $exp(-kA_i)$, with $k \approx 6.3$. We sample m_1 and m_2 so as to maintain uniform sampling of the dominant red component as much as possible. We use $\hat{m}_1 \sim U[0, 1]$, where $\hat{m}_1 = exp(-km_1)$ and $\hat{m}_2 \sim U[0, 1]$, where $\hat{m}_2 = exp(-km_2)$. From this, the absorptions can be expressed as $A_i = -\ln(\hat{m}_1^{p_i} \hat{m}_2^{q_i})$. To keep the correlation as much as possible for the dominant red component, we set $p_r = 1$, $q_r = 1$, $p_g = 1.3036$, $q_g = 1.6390$, $p_b = 2.2272$ and $q_b = 3.8370$, as determined by the linear melanin combination model described earlier (§3.2). Our sampling for absorptions allows for a broad range of $[0, \infty]$. This results in values outside the typical range of $[0.2, \infty]$ given in [MJC*03] for the absorptions; however values as small as 0.03 are found in [MM06].

Our acquisition configuration is only loosely specified, namely a photo of the back of the head with a flash near to the lens. To increase the robustness of our technique to variation in light and camera direction, we include random perturbations of these factors in our dataset. For each dataset image, we add random offsets in $[-2.5°, 2.5°]$ to the viewing direction and lighting direction.

The rendering illumination is the same across all dataset images and all hairstyles and thus gives a self-consistent default exposure for the renderings. In order to achieve compatible exposures with the gray-card calibrated photos, we apply an illumination scaling parameter, s, to the rendered images. The value of s is determined using a one-time cross-validation. Specifically, we sample s in the set of bracket of exposures $\{0.6, 0.8, 1.0, 1.2, 1.4\}$ and we keep the value yielding the best overall match results ($s = 1.4$ in our case) for a small set of test photos (we used a minimum of 10 photos). We noted that values of $s \leq 1.0$ were unable to reproduce lighter hair colors. Lastly, we also compute an image mask so that pixels that are alpha-blended with the background can be excluded from the feature computation. All images are rendered against a blue background. All pixels having color components $b > g$ and $b > r$ are excluded from the mask, and this is further followed by a 3×3 image erosion operation.

4.3. Photo Preprocessing

We take a flash photo of the back of the head together with a reference 18% gray card, taken indoors with a short exposure, e.g., $1/200s$, in order to minimize the impact of indirect lighting. Input photographs are first downsampled to 640×480 using filtering based on bicubic interpolants to match the resolution of the database. They are then processed for white balance using the color of the imaged gray card. We currently use Photoshop for this step, as well as for segmenting the hair in the photograph. Lastly, we scale image luminance to achieve a 12% on-screen luminance for the imaged gray card as commonly done by photographers.

4.4. Parameter Estimation

Given a preprocessed input photograph, its features $\mathcal{F}$ are computed and the best-matching hairstyle, i.e., choice of dataset, is manually selected. A simple linear search is then used to select the nearest neighbor, i.e., $j^* = \arg\min_j \delta(\mathcal{F}, \mathcal{F}_j)$, and the estimated parameters are given by $\mathcal{P}_{j^*}$. This requires approximately 20 seconds for a dataset of 5000 images. More sophisticated forms of non-parametric regression could also be used, i.e., applying kernel regression to $\mathcal{P} = f(\mathcal{F})$. However, these did not improve the resulting estimates (see discussion in §7), likely because the distance from the photo to the manifold of rendered images is generally larger than the distances between neighboring samples on the manifold. As a result, a large kernel spans too many neighboring samples while a small kernel effectively results in nearest-neighbor selection.

5. Perceptual Evaluation

We performed an experiment that provides a perceptual evaluation of the EMD-based metric used by the estimation procedure. Our goal is to verify that if the metric predicts that renderings (and thus the appearance parameters) are similar to photographs, human observers also find them similar; while if the metric predicts a large difference, humans agree with this prediction.

Subjects are asked to make relative assessments as to which of two renderings they find to be most similar in appearance to a given photo. The two renderings use the same hairstyle, which avoids confusing hairstyles with hair appearance. The experimental setup thus exactly mimics decisions of the type that need to be made during the parameter estimation (§4.4), and it fits naturally into a two-alternative forced choice (2AFC) protocol. The experiment does not provide a measure of *absolute similarity* of a rendering to a photograph: This is a very hard problem, and there is no established way to do this. In addition, such a hypothetical test would also involve the evaluation of the quality of the rendering and lighting model, which are beyond the scope of this work.

Given a photo, P, and two renderings, A and B, we expect that our metric will be weakly predictive of the choice of closest image for cases where $\delta_{AB}^P = |\delta(\mathcal{F}_P, \mathcal{F}_A) - \delta(\mathcal{F}_P, \mathcal{F}_B)|$ is small, i.e., the metric finds that images A and B are roughly equidistant to the photo, P. Similarly, we expect the metric to be strongly predictive as δ_{AB}^P becomes large. We thus define three categories producing almost equidistant, quite different and very different renderings : $\Delta_1 : 0.2 \leq \delta_{AB}^P \leq 0.5$, $\Delta_2 : 2 \leq \delta_{AB}^P \leq 3$, and $\Delta_3 : 4 \leq \delta_{AB}^P \leq 10$.

We choose ten photographs from our results data set that approximately span the space of hair appearances. For any given photo, we first build a base bin, B, of three images, where each image i is chosen to be close to the photograph, P by satisfying $\delta_{NN} \leq \delta_{Pi} \leq \delta_{NN} + b_w$, where δ_{NN} is the distance of the nearest neighbor (best match) in the dataset to the photo, and b_w is the base-bin width. We also always include the nearest neighbor as one of the three images in this bin; the other two are randomly chosen. For all image pairs shown in the test, one of the images will come from this base bin. This helps ensure that distance differences are only compared for a similar reference distance. For each photograph, we then select 9 pairs of renderings such that there are 3 pairs for each of the categories, Δ_1, Δ_2, and Δ_3. This experimental setup results in 90 comparisons per test session, where the subject must choose the "rendering with hair appearance most similar to the photograph". The test has a duration of approximately 15 minutes. Example screenshots of questions for each category are included in the supplementary materials.

We used an internet-based survey, distributed to three university or research institutions. Participants were instructed not to complete the survey if they were color-blind and were recommended to use a bright, good quality monitor with sufficient resolution so that the photo and the image-pair could all appear on-screen simultaneously. A total of 47 participants completed the survey, giving a total of 1410 evaluations for each distance category (47 participants $\times$ 10 photos $\times$ 3 tests per category). We consider separate hypotheses for each Δ_1, Δ_2, and Δ_3. The null hypothesis for each case is that users will be at chance in having their selection of the closest image match that given by the metric. The alternate hypothesis is that the metric helps predict the participant's choice of most similar appearance.

For category Δ_1, the subject's choice agreed with the metric in 71.4% of all tests, with a standard deviation across subjects of σ_s=8%. The null hypothesis can be rejected: $\chi^2(1, n = 1410) = 276, p < 10^{-5}$. For category Δ_2, the agreement rises to 75.3% (σ_s=7), and the null hypothesis can be rejected: $\chi^2(1, n = 1410) = 360, p < 10^{-5}$. Lastly, for the large predicted differences of category Δ_3, the agreement is 93.6% (σ_s=5%) and the null hypothesis can be rejected: $\chi^2(1, n = 1410) = 1074, p << 10^{-5}$.

Our experiment thus shows that human observers agree with our feature-based distance for judgements of similarity between photos and rendered images.

6. Results

We test the appearance estimation on a set of 64 photographs. Figure 6 shows a subset of the results and the associated model parameters are given in Table 1. The new lighting conditions shown in the rightmost column are chosen manually to best match the corresponding photographs. The complete set is given in the additional material and the accompanying video. Over the full set of photos, the best-match EMD distance spans the range $1.35 < \delta < 6.13$, with a mean $\mu = 2.92$, and a standard deviation $\sigma = 1.01$.

Animation: We can easily animate our estimated hair appearance, which remains a challenge for other methods such as [P*08]. Animations of two hair appearances are shown in the accompanying video and Figure 5.

Figure 5: *Animating hair using estimated hair appearance parameters, with the input photo shown on the left.*

Application scenario: Our method is suited to end user-driven content creation such as for the design of avatars for games, in contrast to the methods used in high-end visual effects for film where expert artists are available to help set

Figure 6: *Seven parameter estimation results, showing, from left to right : input photo, nearest-neighbor match, photo of new lighting condition, rendered new lighting condition.*

$\hat{m}_1$	$\hat{m}_2$	α_R	α_{TT}	α_{TRT}	N_g	δ
0.001	0.123	11.9	18.3	9.9	0.927	2.75
0.355	0.572	11.4	20.0	23.8	0.788	3.15
0.052	0.267	14.0	5.3	16.4	0.561	1.63
0.081	0.955	16.0	12.7	16.0	0.969	1.99
0.362	0.916	17.3	19.0	23.5	0.544	3.81
0.910	0.086	11.3	18.9	14.9	0.760	3.70
0.033	0.436	8.1	7.3	12.9	0.400	3.42

Table 1: *Estimated model parameters for Figure 6, as well as their distances, δ, to the photo.*

parameters. We have used a subset of our input photographs to create a prototype interface for character design, as shown in Figure 7 and the accompanying video. In this interface, the user can create a library of hairstyles for game characters by taking flash photographs and picking a template geometry. These hairstyles are then available in the game as shown in our prototype (see accompanying video).

Figure 7: *The use of hair appearance in a prototype character customisation interface for a game.*

7. Discussion

We first discuss the various estimation methods we tested before adopting the approach presented here, then discuss a number of robustness tests we performed and conclude with a discussion of limitations.

Other estimation methods: We tried several machine learning approaches based on features inspired by domain-specific knowledge on hair appearance, but none of them worked as well as the approach that we finally propose. We tested regression methods that seek a function f such that the parameters $\mathcal{P}$ can be expressed as $\mathcal{P} = f(\mathcal{F})$, where $\mathcal{F}$ is a set of image features. The critical part is the choice of $\mathcal{F}$. Low-dimensional feature vectors require only a small training set to cover the feature space. But such feature vectors are exceedingly difficult to design since they are almost of the same size as the parameters, that is, finding good feature vectors is almost equivalent to the initial problem. Larger feature vectors do not suffer from this problem but require a larger training set to obtain f, which quickly becomes the limiting factor. We tested feature vectors based on means and medians of the color components, as well as responses to banks of oriented spatial filters. We implemented Gaussian process or Nadaraya-Watson kernel regression to predict the hair absorption parameters, but none produced reliable estimates, as tested using cross-validation with rendered images. We also tested several segmentation methods to correlate highlights with corresponding lobe parameters; none

reliably identified the desired highlights, because the R and TRT regions overlap too significantly.

Our final use of a color histogram solves this dilemma by using a high-dimensional feature, the weighted color distribution, a perceptually-validated metric on this feature vector, and the use of a sampling and a nearest-neighbor scheme to resolve the final parameter estimates. A somewhat suprising aspect of the solution is that it can be effective without relying on spatial features. The advantage of this lack of spatial sensitivity makes our approach robust to spatial variations.

Dataset sampling: We compute a simple statistic to confirm our intuition that the photo and rendered-image manifolds are widely spaced compared to the sample spacing used by the dataset (see Figure 4). Specifically, we compute the ratio $r = \delta_{12}/\delta_{1P}$ over the full set of photos, where δ_{12} is the distance between the first and second nearest neighbors, and δ_{1P} is distance between the photo and the nearest neighbor. The resulting small values of r, $0.22 < r < 1.08$, $\mu = 0.62, \sigma = 0.17$, support the described intuition regarding this geometry. The use of only 5000 samples to cover a seven-dimensional parameter space is partly enabled by the dimensionality reduction to 6 dimensions and perceptually-uniform sampling of the hair color space. We also note that for particular regions of the parameter space, the rendered image is locally invariant to some parameter values, which allows us to use a smaller number of samples. For example, the appearance of black hair is largely invariant to TT or TRT lobe widths.

Match Sensitivity: We examine how the distance to the photo changes as a function of the model parameters in the region surrounding the nearest neighbor. Figure 8 shows normalized plots of how the distance metric changes as individual parameters are varied around their final estimated value while the remainder are held fixed. The absorption parameters and the R-lobe parametere have well-defined local minima. The TT and TRT lobes should not take on values smaller than their nominal value, but have a minimal effect on the overall appearance for larger values. The noise parameter exhibits a shallow local minimum for this example, although in general it can have a stronger variation – the two images shown in Figure 3, which vary only in their noise parameter, have a relatively large distance of $\delta = 3.2$.

Robustness with respect to lighting variation: To test for the effect of changing the lighting, we compare the result of using flash-lighting placed 15 degrees above the lens with the result of flash lighting placed 15 degrees below the lens. For the case of a person with black hair, the original and modified-lighting parameter estimates are $\mathcal{P}_0 = \{0.0518, 0.1133, 13°, 3.5°, 12°, 0.82\}$ and $\mathcal{P}_1 = \{0.0073, 0.6543, 11°, 15°, 17°, 0.82\}$, respectively. While the estimated aborptions are different in the melanin space, they are very similar when seen in the *rgb* space. α_R also remains very similar. The TT and TRT lobe widths receive different estimates, although this is not unexpected
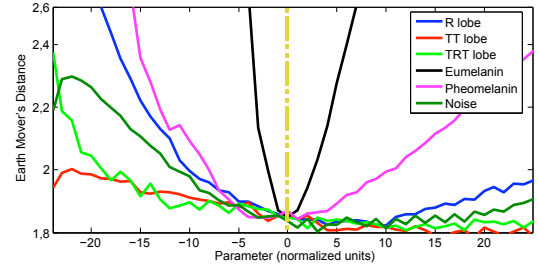
Figure 8: *Example variation of the distance as a function the underlying model parameters around the nearest neighbor. The x-axis spans the following spreads for each model parameter:* $\alpha_R : 28°, \alpha_{TT} : 21°, \alpha_{TRT} : 28°, \hat{m}_1 : 0.65, \hat{m}_2 : 1.9, N_g : 1.92.$

given the negligible role of transmissive scattering in black hair. The distance to the best match changes only marginally ($\delta_0 = 1.93$, $\delta_1 = 2.02$). We also apply the same test for a person with lighter-colored hair, and in this case the same nearest-neighbor is returned ($\delta = 2.36$), thus yielding an unchanged parameter estimate.

Impact of hairstyle geometry: A fundamental question to ask is the extent to which the choice of hairstyle geometry affects the parameter estimation. A first way to measure this is to match a photo using different hairstyles and to observe the resulting parameters and their images. Figure 9 shows an input photo and the nearest neighbors for three different hairstyles. The estimated parameters for the manually chosen target hairstyle are: $\mathcal{P}_0 = \{0.0845, 0.9232, 16°, 13°, 16°, 0.97\}$. With the shown alternate hairstyles, this changes to: $\mathcal{P}_1 = \{0.0137, 0.8786, 21°, 3.9°, 20°, 0.76\}$ and $\mathcal{P}_2 = \{0.0495, 0.9093, 23°, 14°, 23°, 0.68\}$. The differences in estimated parameter values can be attributed in part to the choice of hairstyle and in part to the fact that parameters such as α_{TT} have little effect for dark hair and so they may not be estimated in a consistent fashion.

The impact of hairstyle geometry can also be measured by using each of the seven different hairstyles to do parameter estimation and then using the resulting parameter estimates to render images using the user-selected best-matching hairstyle. We can then compute their respective similarities to the input photo using our metric. The results of this computation show that parameter estimates that come from the user-selected hairstyle are always among the best results, and that largely different hairstyles produce inferior results. This confirms the importance of using similar hairstyles for matching and rendering.

Our library is currently based on seven hairstyle models because these approximately span the range of hairstyles observed in our test set of 64 input photographs, and because of the difficulty of obtaining or creating additional hairstyles.

Figure 9: *The effect of hairstyle choice on parameter estimation: input photo, manually chosen target hairstyle, alternate hairstyle 1, alternate hairstyle 2.*

Figure 10: *The photos and nearest-neighbors for our worst two matches, as measured by the metric. Left pair: $\delta = 6.13$. Right pair: $\delta = 5.08$.*

Enlarging our library of hairstyles may help improve the final rendered likeness to the input photos, and possibly result in small improvements to the parameter estimation.

Limitations: Hair may be dyed, have sun-bleached or dyed highlights, or have a partial distribution of gray hairs, violating our assumption of constant hair aborption values. Visible scalp will also affect the estimations. Our synthetic appearance model does not model wet or greasy hair. Figure 10 illustrates our two worst results as judged by the metric. The gap between the rendered images and the photos for our dataset of tens of heads highlights some of the remaining challenges in hair modeling/capture and rendering. Our appearance estimation technique inherits the limitations of current rendering techniques, but also stands to directly benefit from future advances in hair rendering. In particular, our rendering implementation was unable to obtain realistic images for front lit blond hair, possibly due to the disciplined hair approximation [ZYWK08]. As such we do not present results yet for very ("scandinavian") blond hair.

8. Conclusions

We have presented a novel method for estimating hair appearance parameters from a single flash-lit photo. We develop a hair absorption model based on melanin-based pigmentation, and introduce geometry noise as an appearance parameter. A suitable image feature and distance are defined for measuring hair appearance similarity, and we conduct a perceptual evaluation of this metric, which gives a strong indication of the validity of our choices. The technique has been used to estimate hair appearance parameters for 64 photographs, for which we provide side-by-side comparisons of the input photos and renderings. To our knowledge, this is a significantly larger set of comparative results than those presented to date in prior art on hair modeling and rendering.

We analyze the robustness and sensitivity of the appearance estimates in several ways.

Hair appearance can be captured with a wide range of techniques. Our proposed approach lies at one end of this spectrum, requiring a single flash photograph as input, and producing an estimate of seven appearance parameters in minutes. The technique of Paris et al. [P*08] lies at the other end of the spectrum, requiring a light-stage setup consisting of 16 cameras, 3 DLP projectors, 150 programmable LED lights, 40 000 images, 20 minutes of capture time, hours of compute time, and producing a detailed model of geometry and reflectance that requires 4.3Gb of data during rendering.

There are a number of exciting avenues for further exploration. With the help of lighting and white-point estimation techniques, it may well be possible to do parameter estimation from one or more photos taken in unstructured lighting conditions. We wish to add further geometric expressivity to the model by estimating meso-level geometry such as clumps and wisps of hair that may be identifiable from the photo. Recent progress has been made on this [WYZG09]. The distance we use could be used to define a hair appearance manifold from a large collection of input photos, independent of hair rendering techniques. We believe that the general style of parameter estimation approach may be applicable to other types of phenomena in graphics.

References

[AV07] ARTHUR D., VASSILVITSKII S.: *k*-means++: The advantages of careful seeding. In *ACM-SIAM Symp. on Discrete algorithms* (2007), pp. 1027–1035.

[Ber06] BERTAILS F.: *Simulation de Chevelures Virtuelles*. PhD thesis, Institut National Polytechnique de Grenoble, 2006.

[DWD*08] DONNER C., WEYRICH T., D'EON E., RAMAMOORTHI R., RUSINKIEWICZ S.: A layered, heterogeneous reflectance model for acquiring and rendering human skin. *ACM Trans. on Graphics (Proc SIGGRAPH ASIA)* (2008).

[DWYW09] DUCHENEAUT N., WEN M.-H., YEE N., WADLEY G.: Body and mind: a study of avatar personalization in three virtual worlds. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems* (New York, NY, USA, 2009), ACM, pp. 1151–1160.

[GSML02] GRABLI S., SILLION F., MARSCHNER S. R.,

LENGYEL J. E.: Image-based hair capture by inverse lighting. In *Proc. Graphics Interface* (2002).

[GTHD03] GARDNER A., TCHOU C., HAWKINS T., DEBEVEC P.: Linear light source reflectometry. *ACM Trans. on Graphics 22*, 3 (2003), 749–758.

[HS01] HALAL J., SCHOON D. D.: *Hair Structure and Chemistry Simplified*. Cengage Learning, 2001.

[KK89] KAJIYA J. T., KAY T. L.: Rendering fur with three dimensional textures. In *Proc. SIGGRAPH '89* (1989), pp. 271–280.

[L'O08] L'ORÉAL: http://www.hair-science.com, hair science, accessed Nov, 2008.

[M*97] MARKS J., ET AL.: Design galleries: A general approach to setting parameters for computer graphics and animation. In *Proc. SIGGRAPH* (1997), pp. 389–400.

[MJC*03] MARSCHNER S. R., JENSEN H. W., CAMMARANO M., WORLEY S., HANRAHAN P.: Light scattering from human hair fibers. *ACM Trans. Graph. 22*, 3 (2003), 780–791.

[MM06] MOON J., MARSCHNER S.: Simulating multiple scattering in hair using a photon mapping approach. *ACM Trans. on Graphics 25*, 3 (2006), 1067–1074.

[MTLB03] MIHASHI T., TEMPELAAR-LIETZ C., BORSHUKOV G.: Generating realistic human hair for "the matrix reloaded". In *ACM SIGGRAPH 2003 Sketches and Applications Program* (2003).

[MWM08] MOON J. T., WALTER B., MARSCHNER S. R.: Efficient multiple scattering in hair using spherical harmonics. *ACM Trans. on Graphics 27*, 3 (2008).

[ND06] NGAN A., DURAND F.: Statistical acquisition of texture appearance. In *EG Symp. on Rendering* (2006), pp. 31–40.

[P*08] PARIS S., ET AL.: Hair photobooth: Geometric and photometric acquisition of real hairstyles. *ACM Trans. on Graphics 27*, 3 (2008).

[PBS04] PARIS S., BRICEÑO H., SILLION F.: Capture of hair geometry from multiple images. *ACM Trans. on Graphics 23*, 3 (2004), 712–719.

[RTG00] RUBNER Y., TOMASI C., GUIBAS L.: The Earth Mover's Distance as a Metric for Image Retrieval. *Intl J. of Computer Vision 40*, 2 (2000), 99–121.

[SS88] SARNA T., SWARTZ H. M.: The physical properties of melanins. In *The Pigmentary System* (1988), Oxford University Press.

[Tob08] TOBIN D. J.: Human hair pigmentation: biological aspects. *Intl J. of Cosmetic Science 30*, 4 (2008).

[WBK*07] WARD K., BERTAILS F., KIM T., MARSCHNER S., CANI M., LIN M.: A Survey on Hair Modeling: Styling, Simulation, and Rendering. *IEEE Trans. Vis. and Comp. Graphics* (2007), 213–234.

[WOQS05] WEI Y., OFEK E., QUAN L., SHUM H.: Modeling hair from multiple views. *ACM Trans. on Graphics 24*, 3 (2005), 816–820.

[WYZG09] WANG L., YU Y., ZHOU K., GUO B.: Example-based hair geometry synthesis. *ACM Trans. Graph.* (2009).

[Yu01] YU Y.: Modeling realistic virtual hairstyles. In *Proc. of Pacific Graphics* (2001), pp. 295–304.

[Zin07] ZINKE A.: Light scattering from filaments. *IEEE Trans. Vis. and Comp. Graphics 13*, 2 (2007), 342–356.

[ZYWK08] ZINKE A., YUKSEL C., WEBER A., KEYSER J.: Dual scattering approximation for fast multiple scattering in hair. *ACM Trans. on Graphics 27*, 3 (2008).