Recent Advances in Facial Appearance Capture

Oliver Klehm[1,2] Fabrice Rousselle[1] Marios Papas[1,3] Derek Bradley[1]
Christophe Hery[4] Bernd Bickel[1,5] Wojciech Jarosz[1] Thabo Beeler[1]

[1]Disney Research, Zürich [2]MPI Informatik [3]ETH Zürich [4]Pixar Animation Studios [5]IST Austria

Abstract

Facial appearance capture is now firmly established within academic research and used extensively across various application domains, perhaps most prominently in the entertainment industry through the design of virtual characters in video games and films. While significant progress has occurred over the last two decades, no single survey currently exists that discusses the similarities, differences, and practical considerations of the available appearance capture techniques as applied to human faces. A central difficulty of facial appearance capture is the way light interacts with skin—which has a complex multi-layered structure—and the interactions that occur below the skin surface can, by definition, only be observed indirectly. In this report, we distinguish between two broad strategies for dealing with this complexity. "Image-based methods" try to exhaustively capture the exact face appearance under different lighting and viewing conditions, and then render the face through weighted image combinations. "Parametric methods" instead fit the captured reflectance data to some parametric appearance model used during rendering, allowing for a more lightweight and flexible representation but at the cost of potentially increased rendering complexity or inexact reproduction. The goal of this report is to provide an overview that can guide practitioners and researchers in assessing the tradeoffs between current approaches and identifying directions for future advances in facial appearance capture.
The definitive version is available at diglib.eg.org and onlinelibrary.wiley.com.

1. Introduction

The human face is arguably the most distinctive aspect of human appearance. The face is used to immediately identify a person, infer their emotions, portray their physical condition, and can even communicate their current state of mind. Accurately capturing the facial appearance of a real subject can lead to a wide range of applications, from virtual makeup systems [SRH*11, TOS*03] or computer vision [LCQ*04, SPF*13], to the design of prosthetics or animatronics. One particular successful application is that of virtual characters, which are now ubiquitous in video games and the movie industry [ARL*10, Sey13]. Recent advances in facial appearance capture have enabled the creation of life-like digital renditions of real actors. In all these applications, photo-realism is generally desired, but also elusive, as humans are extremely sensitive to subtle aspects of facial appearance. The way light penetrates below the skin surface, the change of skin color due to fluctuations of blood circulation, and fine scale surface details like wrinkles are all cues that audiences expect from a realistic face.

This state of the art report will survey a variety of techniques from the last two decades that are concerned with achieving photo-realistic facial appearance capture. Ideally, the ultimate goal would be to have the ability to render a virtual actor under arbitrary lighting conditions and from any viewing position, including fine geometric details such as facial hair and peach fuzz, and to dynamically change the skin appearance (for instance, to simulate variations in blood flow). While capturing a complete virtual actor generally implies reconstructing the (macroscopic) geometry as well as the scalp hair, we will not consider these aspects within this report, and instead restrict ourself to capturing the interaction of light with the skin, though we do consider mesoscopic geometry and facial hair as part of facial appearance. While the problem of performance capture is orthogonal to capturing appearance, geometry and corresponding surface normals are often required for appearance capture methods. Here, high-quality geometry is usually employed, which can be reliably reconstructed with state-of-the-art capture methods assuming controlled illumination [ZSCS04, BHPS10, BHB*11, HCTW11].

For an extended view of the field of digital characters, we refer the reader to previous reports. Igarashi et al. [INN07] cover existing models with a focus on accurate bio-physical modeling of the skin, up to the cellular level. Weyrich et al. [WLL*09] consider the principles of appearance acquisition in general, where facial appearance is one of many applications. Jung et al. [JKK*11] consider interactive virtual characters, including behavior control, real-time animation, and real-time rendering. Ward et al. [WBK*07] survey hair modeling, including capture from photographs.

The central component of facial appearance is skin itself—which has a complex multi-layered structure—and more specifically, how light interacts with skin. A portion of light will reflect directly off the oily skin surface, resulting in view-dependent highlights that expose the finer geometric details of skin, such as its pores and wrinkles. Light that does not reflect off the skin surface then travels through the various layers of the skin, each of which has different characteristics, and therefore interacts differently with the light. In addition to the complex cumulative effect of the various skin layers on color, this travel allows light to exit the skin from a different position than it entered, which effectively smooths the reflectance, giving skin its characteristic soft appearance. All interactions that occur below the surface are by definition difficult to measure, and this problem is compounded by the fact that capturing facial appearance has to be done rapidly to be practical. Methods that aim to capture facial appearance must cope with the complexity of light interactions with skin, and two broad categories can be distinguished: *image-based methods* and *parametric methods*. Image-base methods exhaustively capture the exact face appearance under (potentially many) different lighting and viewing conditions, and then solve the rendering problem through weighted image combinations. Parametric methods instead aim at modeling the structure of skin with suitable approximations, which then allows a more lightweight and flexible representation of the skin at the cost of a potentially inexact reproduction.

The structure of this report is build around the broad classification between image-based and parametric methods. We start by presenting general capture methodologies developed over the years, which introduce different illumination and capture configurations that have been used throughout the capture methods (Sec. 2). We then present a variety of image-based (Sec. 3) and parametric (Sec. 4) methods that have been proposed. We focus mainly on how the various methods relate to each other, and therefore keep the exposition at a high level. Lastly, we discuss more practical concerns that arise when capturing or editing facial appearance (Sec. 5) and identify some considerations for future work (Sec. 6).

2. Capture Methodologies

Our goal is to capture the appearance of skin from real-world measurements. As we have mentioned, the skin structure is very complex including an uneven oily surface and several different scattering layers. The overall appearance is defined by the way light from all directions and wavelengths interacts with the skin and surrounding objects, and then bounces into the observer's eye from any particular viewing angle at any particular moment in time. This appearance problem is incredibly complex and it is generally not feasible to capture all possible light paths. For this reason, various different capture methodologies have been developed to sample the illumination, the viewing angles, and the temporal variation, which we will now describe.

2.1. Illumination

We first consider methods to sample the incoming illumination. These approaches fundamentally rely on the linearity of light transport, which states that the light reflected from a surface illuminated simultaneously by two different light sources will equal the sum of the reflectance when illuminating by each light source independently. This implies that if a subject is measured while individually lit from n different directions, the appearance under any linear combination of these n directions can be synthesized by linearly combining the n acquired measurements. One methodology is thus to illuminate the subject from a series of independent directions, which we will refer to as the canonical basis. The canonical basis cannot be used to synthesize light from directions that were not sampled, so the more illumination directions acquired the better. The downside, of course, is the time it takes to perform these individual measurements. This may be overcome by using spherically smooth and continuous illumination bases, such as spherical harmonics, which allow the acquisition of reflectance statistics and require far fewer measurements. The need to sample the incoming illumination directions has led to the design of a series of *light stages* over the past two decades, so-called because the idea is to surround an actor with light. The main practical challenges in the design and construction of these devices are sampling density and acquisition time.

A basic approach is to sample light directions individually by moving a single point light to the desired position in space and recording a measurement for each light position [MWL*99, DHT*00, FBLS05]. Most of these devices have mechanical components with two degrees of freedom and effectively rotate the light to the desired position, allowing to continuously sample both longitude and latitude (see Fig. 1). While this approach has the flexibility to very densely sample incident angles, it comes at the cost of acquisition time, often requiring several minutes for a complete measurement during which time the subject must remain very still. Acquisition time can be reduced by distributing more lights spatially around the actor, since turning a light on and off is considerably faster than mechanically moving it to the desired position. Matusik et al. [MPN*02] add lights along the vertical axis on an arc, which they rotate around the object. This replaces one of the mechanical degrees of

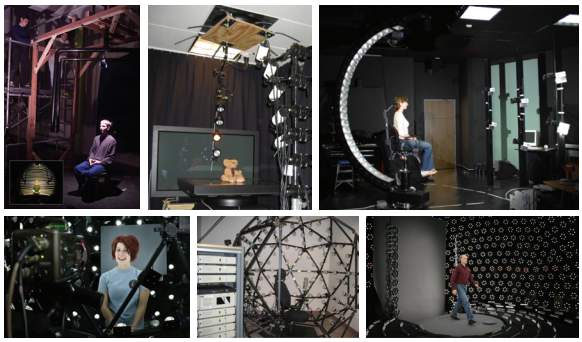


Figure 1: *Upper row: early illumination capture methodologies sampled light directions by rotating lights on mechanical gantries around the subject [DHT*00, MPN*02, HWT*04]. Lower row: Illumination can be sampled faster using a sparse set of fixed lights mounted around the subject [WGT*05, WMP*06, ECJ*06].*

freedom with a set of fixed light positions. Activating these lights synchronized to the rotation of the arc allows to sample the incident illumination continuously in longitude but discretely in latitude. This is sufficient to capture a canonical basis in just a few seconds by strobing the lights [HWT*04]. Tunwattanapong et al. [TFG*13] extend this by increasing the light density on the arc to the point that they can produce a smooth incident illumination condition, allowing capture under the canonical and also the spherical harmonics basis.

To further decrease acquisition time and to reduce complexity and physical limitations introduced by the mechanical components, Debevec et al. [DWT*02] discretely distribute fixed lights all around the actor. These devices have been employed both to capture canonical [DWT*02, WMP*06] and smooth bases [MHP*07, GCP*09, TGD11], as well as other illumination conditions [WGT*05]. Furthermore, without moving parts one could also consider larger capture volumes [ECJ*06]. These approaches, while faster in acquisition than rotating light sources, come at the cost of sparser directional sampling because it becomes impractical to place more than a few hundred lights in a dome (see Fig. 1).

All these devices so far are concerned with generating distant illumination, where the light incident on the surface does not vary spatially. Spatially varying illumination can be achieved by employing projectors. On the one hand, projected stripe patterns are often used to estimate facial geometry [ZSCS04, MHP*07, MJC*08, WLVGP09], but a complete survey is beyond the scope of this paper. On the other hand, light projection can also be used for appearance capture to help estimate how light scatters beneath the skin by sampling a subset of light paths using a variety of binary patterns [NKGR06, TGL*06], and this approach can even be combined with fixed spherical illumination [GHP*08].

Figure 2: *Image-based approaches inherently capture all real world effects, allowing for highly realistic relighting effects [WGT*05].*

2.2. Viewing Direction

We now turn to approaches for sampling the outgoing illumination. Not surprisingly, the most straightforward method is to capture the fixed bundle of rays that lead to a single photographic camera. Naturally, angular sampling can be increased with a multi-camera setup. Additionally, the subject can be placed on a turn table to increase sample density [MPN*02, ECJ*06, TFG*13]. One can also use mirrors to add additional virtual views, by spatially multiplexing them onto the same sensor. This obviously comes at the cost of reduced spatial resolution for the individual views [GHAO10, IRM*12].

2.3. Temporal Domain

Finally, it is also important to consider the temporal domain, as facial appearance changes over time. Temporal variation in position or shape of the skin can affect global illumination due to changes in occlusion and interreflections. Furthermore, physical changes in skin appearance may be caused by time-varying blood flow or perspiration. A basic approach is to sequentially acquire multiple individual measurements spaced out over time, for example as the subject undergoes extreme facial expressions [JSB*10]. Alternatively, appearance information from live performances can be acquired using high-speed video cameras [FHW*11] or restricting the performance to repetitive motions [ECJ*06].

In summary, the capture methodologies described above are used in various approaches for image-based and parametric facial appearance capture, as we describe in the following sections.

3. Image-based Approaches

The key idea of image-based methods is to perform a weighted superposition of captured responses to different illumination patterns. The data is directly used for rendering and not fit to any model, which is the main difference to parametric approaches. Image-based methods are not bound to the limitations of a specific model and, hence, are able to reproduce all real-world effects accurately, if an appropriate

basis of image data is available (Fig. 2). As the data is directly used for rendering, generating images usually does not require extensive computation. On the downside, however, image-based methods are bound to the data that has been captured and are usually very data intensive. Artistic control and editing of the face appearance also poses a formidable challenge.

3.1. Reflectance Field

The interactions of light with an object can be arbitrarily complex, making explicit modeling of these interactions extremely challenging. For image-based techniques however, only the observed result of these interactions is relevant, which can be mathematically encapsulated with the notation of reflectance field. Lets assume the object is enclosed in an imaginary convex manifold $\mathcal{M}$, as shown in Fig. 3. Then the complex light interactions taking place inside this manifold can be considered a black box which is fully described by the boundary conditions: the incoming and outgoing light rays at the manifold surface. This reflectance field has eight degrees of freedom. Four degrees describe the incoming light field: light entering the manifold $\mathcal{M}$ at a point $\mathbf{x}_i$ from a direction $\vec{\omega}_i$. The other four degrees specify the outgoing light field: light exiting the manifold $\mathcal{M}$ at a point $\mathbf{x}_o$ in direction $\vec{\omega}_o$. The concept of a light field was proposed simultaneously in graphics by Levoy et al. [LH96] and Gortler et al. [GGSC96] and extended to a reflectance field in the context of facial appearance capture by Debevec et al. [DHT*00]. In addition to these eight dimensions, one may also consider time as additional ninth dimension.

The above description of the reflectance field uses an entirely manifold centric parametrization. Since the reflectance field is traditionally captured using cameras—each of which captures a 2D slice of the outgoing light field—it can often be more convenient to use a view-centric perspective for the outgoing portion of the reflectance field. In this view-centric parametrization the locations of the cameras and pixels of the captured images define the 4D set of rays of the outgoing light field. We refer to the *viewpoint domain* as the locations of the cameras, and the *image plane domain* as the resolution of each camera in this parametrization.

Sampling the complete 9D reflectance field is still challenging and depending on the use case authors reduce the dimensionality by fixing certain dimensions. The most obvious one is certainly to only consider static scenes, thus fixing the temporal dimension and reducing it to the eight dimensions originally introduced. Another very common reduction to assume distant illumination. In this case light from a given direction $\vec{\omega}_i$ hits the manifold at all points equally, thus removing the spatial variability and reducing the dimensionality by two. Capturing from only a single camera removes the viewpoint domain sampling of the outgoing light field, reducing the dimensions by two as well.

As we will elaborate in the following sections, different authors have focused on sampling different sub-spaces of the reflectance field. We first start with papers that are concerned with static faces only and then look at subsequent work that samples the temporal dimension as well.

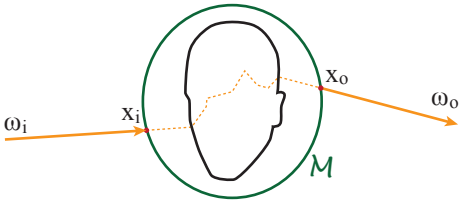


Figure 3: *The 8D reflectance function encapsulates the complex light interactions inside an imaginary manifold $\mathcal{S}$ via the boundary conditions on the surface of the manifold: the incident position $\mathbf{x}_i$ and direction $\vec{\omega}_i$, as well as the outgoing position $\mathbf{x}_o$ and direction $\vec{\omega}_o$.*

3.2. Static Faces

Building on early work in image-based rendering [NSD95], the seminal work of Debevec et al. [DHT*00] and the advent of a series of light stages introduced effective methods for acquiring the reflectance field of a human face. This work sparked a large body of research, and also had high impact on the movie industry. Their main goal was to enable relighting of an actor after capturing, an important step towards the virtual actor. The basic approach is to measure the reflectance field of a face introduced in the previous section. To make the problem tractable, the sampling domain of the reflectance field was reduced to 6D by assuming distant illumination. Furthermore, they captured the outgoing illumination from only two fixed viewpoints. Therefore the viewpoint domain is only very sparsely sampled, leaving only four dimensions. These four dimension are sampled by moving a light to fixed positions around the subject (64×32 light directions covering a full sphere) and acquiring images with the two cameras. According to the capture methodologies introduced in Sec. 2 this sampling pattern is a canonical basis. Due to the linearity of light transport these canonical images can be linearly combined to synthesize any illumination condition that could be physically produced by combinations of the sampled illumination directions.

While densely sampling the incident illumination directions enables high quality relighting, the synthesis is tied to the viewpoint used to acquire the images. To allow synthesis under novel viewpoints, the viewpoint domain needs also to be densely sampled, which amounts to sampling a 6D reflectance field. This requires excessive capturing as well as data storage.

Instead, Debevec et al. [DHT*00] suggest to only sparsely sample different viewing directions and additionally estimate

the geometry and surface normals to guide a view interpolation technique. They estimate the surface normals from the acquired image basis and employ an external system to acquire the geometry. However, even with geometry, the view interpolation is not a trivial task as the skin does not exhibit Lambertian reflection behavior. Therefore the authors split the reflectance into a view-independent (diffuse reflection) and a view-dependent component (specular reflection). Separating these two components will be a recurring theme in subsequent papers and several different methods have been explored as summarized in Sec. 4.1. To synthesize a novel view, the specular component is transformed using the estimated surface normals and merged with the constant diffuse component to form the final image.

To reduce the number of required photographs, Tunwattanapong et al. [TGD11] propose an optimization procedure to determine an optimal illumination basis, taking into account typical surface reflectance functions. The resulting basis is a combination of a smooth illumination basis (spherical harmonics) and a set of distant local lights, representing both low and high frequency components, and requires an order of magnitude less data compared to conventional sampling.

3.3. Dynamic Faces

The goal of the first approach that supports live performances is to ease the composition of an actor into a virtual environment [DWT*02]. For this the illumination within the virtual scene and illumination of the actor needs to be matched. To do so, Debevec et al. [DWT*02] constructed the first light dome with 156 static red/green/blue LEDs to illuminate a subject inside the dome. By sampling the potentially continuous environmental illumination down to the 156 lights, they can physically capture the subject under a discretized approximation of the target illumination. Since the lights are controllable, they can also be used to simulate dynamic illumination, e.g., capturing the face of a subject as if the person is walking through a virtual environment. This approach, however, does not allow for changing the illumination or viewpoint post-capture as in their earlier work [DHT*00].

A first step towards increased post-processing capabilities for an animatable face was proposed by Hawkins et al. [HWT*04] (Fig. 4). Their main assumption is that facial geometry fully defines the appearance of a face, hence, there is a direct mapping of geometry to appearance. The key idea is that a database with associations of geometry to reflectance could be queried to return the appearance for any target geometry. They start by building the database with the reflectance information (using Debevec et al.'s [DHT*00] approach) for various static facial expressions, but also for different head poses as well as mouth and eye positions. They reconstruct the corresponding geometry with the help of facial markers. A typical database contains about 60 expressions, imaged from 6 cameras, each with geometry data and 4D reflectance information based on 480 distant lights. They then estimate

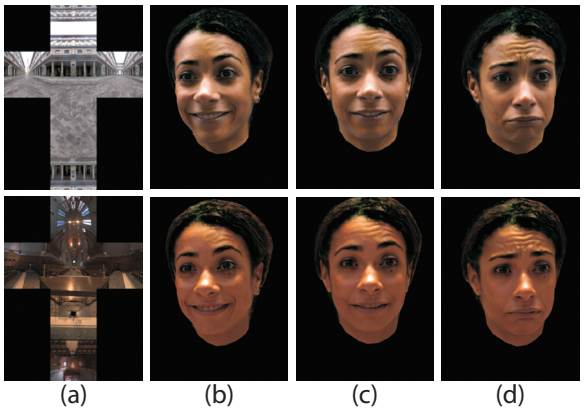


Figure 4: *Animatable facial reflectance fields. Given novel distant illuminations (a), an actor can be relit under different points of view (b,d), and expression (b,c,d) [HWT*04].*

the reflectance for any novel geometry and illumination in a two step approach. First, they relight the captured images as in Debevec et al.'s earlier approach [DHT*00]. Second, they linearly combine the relit images, weighted according to the similarity of the target geometry and the geometry of the database entries. The quality of the synthesized reflectance thus highly depends on how similar the closest match in the database is to the current geometry. This restricts novel performances to only similar expressions as have been stored in the database. Further, geometry estimation and shape retrieval are non-trivial and active research fields of their own.

A more basic approach was specially designed for the movie *The Matrix*. Borshukov et al. [BPL*03, BL03] capture a running footage of a registered soft ambient lit performing face. Reconstructing geometry not just allows them to transform the capture to a geometry-independent texture space, but also serves as a starting point for artistic shape manipulations. They then apply the animated albedo texture onto the virtually deformed face geometry. For relighting they add two secondary layers of illumination from an assumed directional light: a specular and an ad-hoc diffuse component directly convolved in texture space so as to simulate subsurface scattering. Details, potentially changing over time (such as underlying blood flow and pore structure) are effectively preserved up-close through the capture. However, relighting capabilities are limited due to the lack of capturing other lighting bases than the ambient or fitting more material parameters such as parametric methods (Sec. 4).

To avoid geometry completely, Wenger et al. [WGT*05] use a more direct approach to extend reflectance scanning to live performances. They time-multiplex illumination patterns and capture the performance at high framerates with a highspeed camera. Using the lighting apparatus of Debevec et al. [DWT*02], each light source can be switched on and off rapidly. Although they manage to capture all 156 illu-

minations within a twelfth of a second, motion within this time still causes smearing effects after linearly combining the captured images for relighting. The key is to apply motion compensation via an estimate of optical flow [HS81]. Since optical flow does not handle varying illumination well, Wenger et al. interleave additional fully lit *tracking* frames with the different illumination patterns. This allows for accurate motion compensation and for final relighting at a target framerate of 24 or 30fps. The method is limited to relighting within the captured viewpoints, as it samples only five of the nine dimensions of the reflectance field.

Einarsson et al. [ECJ*06] go one step further and additionally sample different viewpoints, effectively sampling seven dimensions. Naïve sampling of these additional two dimensions requires many synchronized cameras capturing the action from different points of view. While this has been done for BRDF scanning [SSWK13], it poses severe practical and financial challenges to do so with highspeed cameras for animated subjects. Einarsson et al. reduce the amount of cameras required substantially by only placing cameras latitudinally. The longitudinal dimension is then sampled temporally by repeating the same performance several times, each time rotated a few degrees. This trick obviously only works for very repeatable motions, such as walk cycles and even this requires strong actor discipline and complex motion compensation.

Spatially-varying Illumination. All the methods presented so far are only considered with unoccluded distant illumination. These methods could for example not synthesize a shadow cast by a virtual object onto the subject. For a purely image-based approach, this requires sampling the full 4D incident illumination. Jones et al. [JGB*06] propose an alternative by leveraging synthesis to simulate 4D incident illumination, although having only captured reflectance caused by distant illumination. While this allows to add virtual shadows, it cannot be used to simulate local light sources. The idea is simple though; turning a spatially-constant capture into a virtual spatially-varying image requires a scaling/removal of the occluded light transport paths. For the example of spatially-varying shadows, it requires to estimate what fraction of light was captured, but should be excluded. To do so Jones et al. estimate the indirect illumination that is caused by the distant illumination in the canonically-lit images. They do this using a global illumination rendering approach, for which they require estimates of the geometry and normals. These are computed from structured light scanning, time-multiplexed with the canonical basis. The key missing component for a simulation is reflectance information, hence, they assume a lambertian reflectance behavior scaled by a recovered albedo map. Having an estimate of indirect illumination allows Jones et al. to remove the indirect illumination from the captured images, modulate the remaining direct illumination in a desired manner, and re-compose the two illumination components back together. While the quality of the result directly depends on the accuracy of the estimated indirect illumination, they argue that a more accurate surface reflectance simulation could be used. This directly hints towards parametric approaches, which try to fit models to allow to completely simulate the light/skin interaction for any illumination.

3.4. Conclusion

Image-based methods are a very powerful in reproducing highly realistic detailed facial features and other subtle effects that cannot be described by existing parameteric appearance models. However, they usually come at the cost of extensive data capture, and require storing and processing large amounts of data. In addition, using this approach for modeling the appearance of virtual face models that go beyond cloning of a real actor or to artistically edit the appearance is highly challenging. In these cases, often a parametric model is the method of choice.

4. Parametric Approaches

In contrast to image-based approaches which use the captured data directly, parametric approaches use the measurement data to determine the parameters of some parametric forward appearance model used during rendering. After fitting parameters from the measurements, the original data can be discarded. This strategy has a number of compelling advantages. Firstly, since the parameters of the forward model are typically lower dimensional than the original measurements, the fitting process can massively reduce the amount of data that needs to be stored. Moreover, appropriately chosen parametric models can inform the capture process, potentially reducing the amount of measurements needed while simultaneously enabling interpolation and even extrapolation beyond sparse observations. Finally, parametric models typically allow for easier editing and re-use of appearance after capture, as long as the employed parametric model is intuitive.

For parametric approaches to work well, however, the employed models need to accurately reproduce all the important visual effects present in the real world. To make the problem tractable, most approaches decompose the full physical appearance into several parametric appearance models, each accounting for a specific component of facial appearance. For instance, the appearance of a face could be represented by a set of parametric models responsible for surface (Sec. 4.2) and subsurface (Sec. 4.3) scattering, surface details due to mesogeometry (Sec. 4.6), and facial hair (Sec. 4.7). Such parametric approaches therefore need to address two key challenges: 1) how to separate the different effects from one another in the measured data; and 2) how to fit the model parameters efficiently and accurately, often from measurements that only indirectly observe the desired effect.

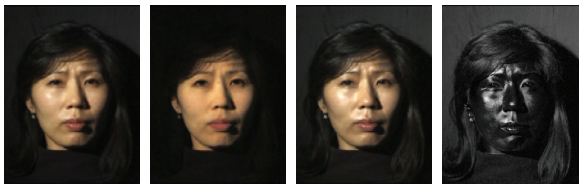


Figure 5: *Diffuse–specular separation using polarization [DHT*00], from left to right: reference image without polarizers; cross-polarization yields diffuse reflectance only, parallel-polarization yields diffuse and specular reflectance, subtracting the cross-polarization image from the parallel-image yields the specular reflectance only.*

4.1. Diffuse–Specular Separation

A shared challenge of most parametric face capture approaches is to separate the diffuse and specular component of reflectance, which model the subsurface scattering in the skin and reflection on the oily layer of skin, respectively. It is possible to perform this separation by purely optical means that rely on light polarization or through computational methods that exploit known characteristics of light transport.

Polarization. Methods that leverage polarization to perform the diffuse–specular separation exploit the fact that specular reflections preserve the polarization state of light, while light that has scattered multiple times rapidly loses polarization. Consequently, a setup with linear polarizers can be used to perform the diffuse–specular separation. When using cross-polarization (where the light and camera have orthogonal polarizers), the specular reflection is eliminated, while half of the unpolarized diffuse reflection survives. Parallel-polarization, on the other hand, preserves the specular reflection as well as half of the unpolarized diffuse reflection. In both cases only half of the diffuse light is observed as half of the randomly oriented (unpolarized) light waves are blocked by the polarizer in front of the camera. The diffuse reflectance can be extracted by multiplying the cross-polarization capture by a factor of two. Conversely, the specular reflectance can be obtained by subtracting the cross-polarization measurement from the parallel-polarization measurement (see Fig. 5). Even though polarization requires two captures (cross- and parallel-polarization) to get diffuse and specular reflection, its simple implementation make it the preferred choice for diffuse–specular separation [MGR00, TGL*06, MHP*07, GHP*08, DWd*08, BBB*10, WGP*10, FHW*11, GFT*11, GTB*13]. It should be noted however that the separation is not perfect. What is separated as specular component with polarization, can also contain other diffuse like transport [NFB93] such as single scattering [GHP*08]. Further practical issues for setups with multiple or moving cameras are discussed in Sec. 5.1.

Computational. Various computational approaches have been proposed to perform the diffuse–specular separation. Debevec et al. [DHT*00] consider the captured photograph to be a linear combination of the diffuse color, $\vec{d}$, the specular color, $\vec{s}$, and some error color, $\vec{e} = \vec{d} \times \vec{s}$, assumed to be orthogonal to the diffuse and specular colors. This defines a simple linear system which can be solved to get the weight of each component (diffuse, specular, and noise), and therefore extract images having only diffuse or specular reflectance. The color of the specular component is assumed to be that of the light, while the chromaticity of the diffuse component is recovered as the median of the red–green and green–blue ratios over the reflectance for pixels in a certain brightness range. The luminance of the diffuse color can be fixed arbitrarily, as this only incurs a scaling factor on the weight of the diffuse component. They also introduce further refinements to the diffuse color estimation to account for an observed desaturation at large viewing and lighting angles.

Weyrich et al. [WMP*06] propose an approach based on the observation that specular reflection vanishes for at least some view/light direction pairs. Using a light stage capture device, they look for the pair that has the minimum intensity (since diffuse reflection is uniform for all viewing angles), while penalizing observations at grazing angles and discarding the k smallest values to improve the robustness of the estimation. Once the diffuse reflectance is characterized, it is removed from the captured data, to isolate specular reflectance.

Tunwattanapong et al. [TFG*13] use spherical harmonic illumination, and characterize the specular reflectance using the response of the third- to fifth-order spherical harmonics. This becomes possible since diffuse reflectance has a low frequency that is nearly completely captured by the first two orders of spherical harmonics [RH01]. Once they model the specular component, they remove it from the response of the first two orders to characterize the diffuse reflectance.

A conceptional similar approach is taken by Lamond et al. [LPGD09], who also project distant light of different frequency patterns. One implementation is a binary square wave pattern with the subject illuminated by half of all directions on average. They assume that the diffuse reflectance is therefore simply scaled by 0.5, while the specular lobe from a single viewing direction averages portions of the environment that are lit, unlit, or a combination of both. They further assume that for one of the multiple rotated environmental illumination patterns, the viewing direction's specular lobe covers only the illuminated portions of the environment. Conversely, there needs to be a rotation without any observable specular reflection. Which rotation is required to observe either of these two cases depends on the surface normal and therefore differs between pixels. A simple per pixel min and max of all measurements allows them to extract only diffuse (scaled by 0.5) as well as a diffuse and specular, similar to polarization separation.

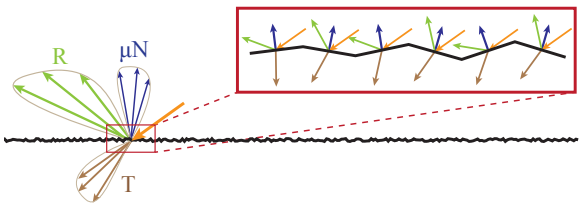


Figure 6: *A diagram showing the various types of interactions between light and a rough dielectric surface. The events depicted in this figure are Surface Reflection (R) and Surface Transmission (T). A zoom-in of the surface which is approximated as a collection of microfacets along with their normals (μN) is highlighted on the top.*

Finally, optical light transport analysis can also be used to perform surface and subsurface reflectance separation [NKGR06, ORK12], using structured light instead of environmental illumination. These methods are, however, not typically used for diffuse–specular separation for faces.

4.2. Surface Reflectance

Since skin is composed out of several layers (Fig. 14) it is commonly decomposed into a top layer with a surface reflectance model and lower layers with a volumetric approximation.

4.2.1. Background BRDF Models

If we assume only surface reflection, the radiance L_o observed at a surface location $\mathbf{x}$ from direction $\vec{\omega}_o$ requires integrating the product of the incident lighting L_i and *bidirectional reflectance distribution function* (BRDF) f over the sphere of directions Ω:

$$L_o(\mathbf{x},\vec{\omega}_o) = \int_\Omega f(\mathbf{x},\vec{\omega}_i,\vec{\omega}_o) L_i(\mathbf{x},\vec{\omega}_i)\, d\vec{\omega}_i^{\perp}, \qquad (1)$$

where $d\vec{\omega}_i^{\perp}$ is the projected solid angle measure. In the general case, a spatially-varying BRDF is a 6 dimensional function (2 for the surface position $\mathbf{x}$ and 2 each for the directions $\vec{\omega}_i$ and $\vec{\omega}_o$).

Parametric approaches assume that this high-dimensional function can be compactly described using some (typically small) set of parameters Θ, which we denote as $f(\mathbf{x},\vec{\omega}_i,\vec{\omega}_o\,|\,\Theta)$. While some work in capturing surface reflectance [FBLS05, GCP*09, TFG*13] has utilized empirical, phenomenological models (such as that of Ward [War92], largely due to its easy fitting), the vast majority of recent techniques have relied on micro-facet models [TS67, CT82, HTSG91, APS00, WMLT07] which have a strong theoretical basis and have been validated against measurements by several researchers [TS67, NDM05, WMLT07].

These models assume that, at a microscopic level, a rough surface is composed of randomly oriented facets—the *micro-facets*—which specularly reflect and refract light (see Fig. 6). The distribution D of micro-facet normal orientations is the primary parameter which characterizes the shape, or "specular roughness," of these parametric BRDF models. This distribution is most often defined statistically by an analytic function such as the Beckmann [Bec87], Blinn [Bli77] or the GGx [WMLT07] distribution, which compactly express the spherical distribution using just one (m isotropic) or two ($m_{u,v}$ anisotropic) numbers. According to this distribution function, a shadowing and masking term is used to account for the fact that micro-facets can be occluded from the incident or outgoing directions. The amount of light reflected and refracted off each micro-facet is dictated by the Fresnel equations which depend on the relative index of refraction (IOR) η of the interface. Additionally, a direction-less albedo term ρ, which scales the amount of scattered light, is sometimes employed to account for imperfections leading to absorption at the micro-facets. In summary, the surface scattering off the top boundary of skin is characterized by the set of parameters $\Theta = \{m_{u,v}, \eta, \rho\}$.

4.2.2. Parameter Estimation

Assuming the specular and diffuse components are separated using one of the techniques described previously in Sec. 4.1, the goal is to capture all the data necessary to reliably fit the parameters of the surface reflectance model while simultaneously trying to minimize capture time by means of different sampling strategies. We summarize the models in Tab. 1.

Ideally, to capture the specular reflectance at every surface point, one would use a gonioreflectometer, but this is not practical for skin, since measurements have to be done rapidly and directly on the subject. One can, however, approximate this setup using a geodesic dome supporting light sources and cameras to sample the BRDF domain (refer to Sec. 2.1). This is the approach taken by Weyrich et al. [WMP*06], who used a dome with 150 LED lights and 16 cameras,

Table 1: *Comparison of surface parameter estimation methods. The acronyms in the sampling strategy column are as follows: G - goniometer, CS - curved surface, GI - gradient illumination, P - polarimetry.*

	Diffuse–specular separation	Roughness	Sampling strategy
[MWL*99]	polarization	object	CS
[Geo03]	-	face	G
[FBLS05]	-	region	CS
[GCP*09]	polarization	pixel	GI
[GCP*10]	polarization	pixel	P
[GTB*13]	polarization	region	CS
[WMP*06]	computational	pixel	G
[GHP*08]	polarization	region	CS

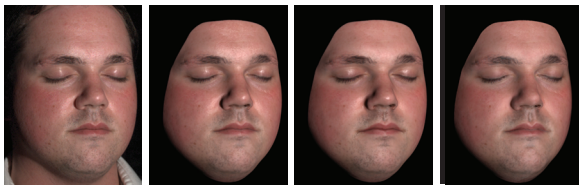


Figure 7: *Data-driven BRDF fitting. From left to right: input photograph, synthetic renderings using a spatially-varying Torrance-Sparrow, spatially-varying Blinn-Phong, and uniform Torrance-Sparrow BRDF model [WMP*06].*

Figure 8: *Segmenting the face into regions of similar reflectance allows the reconstruction of region-specific BRDF parameters accurately and quickly, since a single capture provides many observations. From left to right, the segmentations used by the methods of Weyrich et al. [WMP*06], Ghosh et al. [GHP*08], and Fuchs et al. [FBLS05].*

and sequentially turned on each light while capturing with all cameras. After a computational diffuse–specular separation (Sec. 4.1), they subtract the diffuse component from the observations and fit the specular reflectance parameters using the procedure of Ngan et al. [NDM05]. Weyrich et al. [WMP*06] considered three BRDF models of increasing complexity: Blinn-Phong [Bli77], Torrance-Sparrow [TS67], and Lafortune [LFTG97]. They performed a per-pixel fit and found the Torrance-Sparrow BRDF provided the best results. The Blinn-Phong BRDF tended to over-estimate the specular component, while the Lafortune model—with its higher number of parameters—yielded unstable fits. These observations, in particular the unstability of fitting the Lafortune model, agree with those made previously by Fuchs et al. [FBLS05] in a similar experiment. Interestingly, Weyrich et al. found it better to use a spatially-varying Blinn-Phong fit than a spatially-uniform Torrance-Sparrow fit, which further underlines the importance of a per-pixel parameterization (Fig. 7). The Torrance-Sparrow model has been widely used to model the specular surface reflection of skin [DHT*00, Geo03, FBLS05, DJ06, GHP*08, LPGD09, GTB*13].

While Weyrich et al.'s approach achieves suitable results, it requires extensive sampling and a relatively long capture time, during which the subject has to remain perfectly still. The capture process can be significantly streamlined, however, by leveraging the fact that a single photograph captures the reflectance for many orientations on curved surfaces (as proposed by Marschner et al. [MWL*99]). If we assume uniform surface properties over a given region, it is possible to combine data from all pixels in the same region to obtain denser angular measurement with fewer captures. This idea was leveraged by Ghosh et al. [GHP*08] to estimate the surface reflectance parameters from a single camera position and a single point light source. They used cross-polarization to perform the diffuse-specular separation, and then recovered a per-region surface roughness (forehead, eyelids, nose, cheekbone, lip and lower cheeks). An additional capture with full-on spherical illumination provides an estimate of per-pixel specular albedo. A similar approach to fitting surface reflectance parameters was used by Fuchs et al. [FBLS05], who likewise fitted per-region surface roughness and specular albedo, as well as a per-pixel diffuse albedo. Georghiades [Geo03] also recover surface parameters from a limited set of photographs, fitting a single specular roughness and specular albedo for the whole face. Neither Fuchs et al. nor Georghiades perform an initial diffuse-specular separation, and they instead fit all parameters of the model directly on the captured data.

While the methods of Georghiades, Fuchs et al., and Ghosh et al. reduce the number of required photographs compared to a full dome capture, they come at the price of providing only a per-region (or per-face) estimation of parameters. To get the best of both world—that is, a full per-pixel parameterization using very few photographs—some recent methods have promoted gathering statistical properties of the reflectance (e.g. total energy, mean, and variance) instead of discretely sampling the reflectance function. These methods build on the spherical illumination framework initially proposed by Ma et al. [MHP*07] to recover per-pixel diffuse and specular normals. In particular, Ghosh et al. [GCP*09] use second order gradient illumination to capture the reflectance variance, which they show to be directly proportional to the surface roughness. Using only nine photographs (one to capture the reflectance zeroth moment, used to recover the total energy; three to capture its first moment, used to recover the mean; and five to capture its second moment, used to recover the variance), they can recover a full per-pixel parameterization, including anisotropy.

Ghosh et al. [GCP*10] later proposed an interesting alternative for estimating surface parameters by analyzing the polarization state of light, which can be encoded using Stokes vectors and captured with only four photographs. By leveraging Mueller calculus, describing how a Stokes vector is transformed when light interacts with a surface, they recover a detailed model of per-pixel surface reflectance, including the diffuse albedo, the specular albedo, the specular roughness, and even the index of refraction.

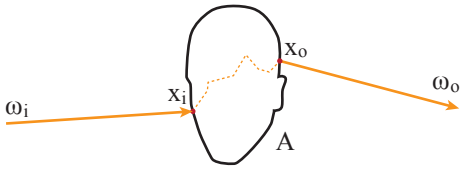


Figure 9: *The 8D BSSRDF function encapsulates the complex light interactions inside of a surface manifold A by relating the incident light at position $\mathbf{x}_i$ and direction $\vec{\omega}_i$ to the outgoing light at position $\mathbf{x}_o$ and direction $\vec{\omega}_o$. This is related to the reflectance field (compare to Fig. 3).*

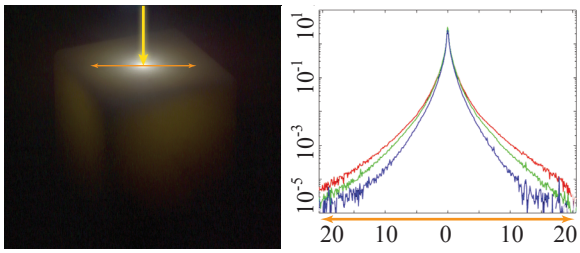


Figure 10: *Diffusion profile of a marble block due to a laser beam at normal incidence [JMLH01]. The x-axis in the diagram denotes distance in millimeters, the y-axis a relative scale of the radiant exitance.*

4.3. Sub-surface Reflectance

We now turn our attention to various approaches that try to capture and parametrically model the sub-surface reflectance of skin. While it is possible to fit parametric BRDF models also to this sub-surface reflectance as in the previous section, since these models assume that light enters and exits at the same surface location, they are not expressive enough to account for the soft, translucent appearance of skin due to sub-surface scattering. BSSRDF models relax this assumption by more accurately modeling the skin surface as a boundary of a volumetric participating medium.

4.3.1. Background BSSRDF Models

The *bidirectional surface scattering reflectance distribution function* (BSSRDF) is the generalization of the BRDF that encapsulates how light incident from direction $\vec{\omega}_i$ at surface boundary location $\mathbf{x}_i$ scatters out at some other location $\mathbf{x}_o$ in direction $\vec{\omega}_o$ (see Fig. 9). To compute the outgoing light, we therefore have to consider the incident radiance at all other locations of the surface, from all directions, which leads to a generalization of Eq. 1:

$$L_o(\mathbf{x}_o, \vec{\omega}_o) = \int_A \int_\Omega S(\mathbf{x}_i, \vec{\omega}_i; \mathbf{x}_o, \vec{\omega}_o) L_i(\mathbf{x}_i, \vec{\omega}_i) \, \mathrm{d}\vec{\omega}_i^{\perp} \mathrm{d}\mathbf{x}_i. \quad (2)$$

In the general case, the BSSRDF is an 8D function (the 6 dimensions from the BRDF, plus an additional 2 for the outgoing position $\mathbf{x}_o$), making it impractical to capture densely using brute-force, image-based techniques. In fact, as can be seen by comparing Fig. 9 and Fig. 3, these can be viewed as different parametrizations of the same function, making dense capturing of the BSSRDF just as difficult as dense capture of the reflectance field. Parametric approaches try to make this more tractable by assuming the high-dimensional BSSRDF arises due to some compact set of parameters Θ, which typically allows collapsing some capture dimensions.

One option is to interpret the tissue beneath the surface as a general (potentially volumetrically-varying) participating medium and obtain the BSSRDF by numerically simulating the light propagation using techniques such as volumetric path tracing [HK93]. While this approach is quite general, the simulation is computationally expensive, and the (potentially volumetrically-varying) parametric description is arguably not much more compact than the desired 8D BSSRDF it is used to describe.

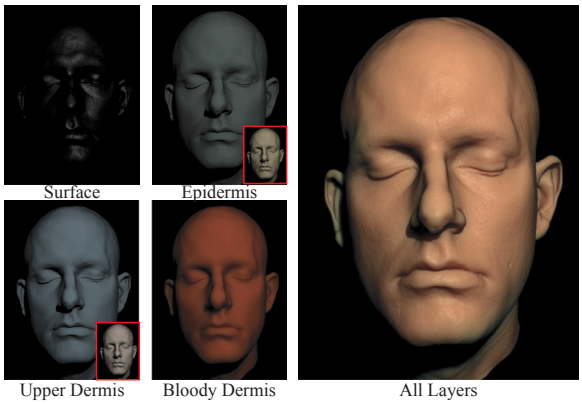


Figure 11: *A breakdown of the multipole diffusion model [DJ05]. Surface reflectance is modeled using the Cook Torrance BRDF. Light that is not reflected off the surface is channeled to the three homogeneous layers of human skin, rendered separately using the parameters obtained from literature [Tuc00]. The final rendering requires a convolution of the contributions of the layers as light travels through the upper layers (epidermis) first before reaching the lower layers (dermis). Transmitted light through epidermis and upper dermis is shown in the insets with a red border. This model forms the basis for current state of the art capture models [GHP*08] as well as more specific biophysical skin models [DWd*08].*

Most practical BSSRDF models therefore make simplifying assumptions about the medium, bounding surface, and/or mathematical form of the resulting BSSRDF to obtain an expressive, yet compact parametric model. A common strategy is to express the multiple scattering in analogy to the so-called "searchlight problem," which considers a focused pencil beam

of light normally incident on the surface of a large (typically infinite), smooth, planar homogeneous medium (see Fig. 10 left). In such a configuration, the light exiting the upper boundary forms a spatial reflectance profile $R(\|\mathbf{x}_o - \mathbf{x}_i\|)$, which is radially symmetric (1D) about the normally-incident beam (see Fig. 10 right). By additionally assuming that the directional and spatial dimensions of the BSSRDF are separable, we obtain a common parametric form:

$$S(\mathbf{x}_i, \vec{\omega}_i; \mathbf{x}_o, \vec{\omega}_o | \Theta) \propto \rho(\vec{\omega}_i \,|\, \Theta) R(\|\mathbf{x}_o - \mathbf{x}_i\| \,|\, \Theta) \rho(\vec{\omega}_o \,|\, \Theta), \quad (3)$$

where $\rho(\vec{\omega})$ is typically collapsed to a 1D angular distribution. This effectively models the behavior of a BSSRDF as a spatial and angular blur, dependent on some parameters Θ.

Jensen et al. [JMLH01] proposed a practical BSSRDF of this form, where the reflectance profile R was expressed using a dipole diffusion approximation, and the angular ρ terms were the Fresnel transmission functions at the incident and outgoing directions. The shape of the profile R is controlled by the relative index of refraction η of the boundary and the medium's absorption σ_a and reduced scattering σ_s' coefficients. Alternatively, the medium parameters can be expressed using the reduced albedo $\alpha' = \sigma_s'/(\sigma_s' + \sigma_a)$ and the diffuse mean free path, or translucency, l_d of the medium.

Notable extensions to this parametric model include the work of Donner and Jensen [DJ05], who introduce the multipole approximation to account for multiple layers of finite thickness—each with its own set of medium parameters (Fig. 11). d'Eon and Irving [dI11] introduced several improvements to the diffusion approximation to better handle the separation of single and multiple scattering and incorporated more accurate boundary conditions. These extensions, as well as many other recent improvements that focus primarily on accelerating the forward rendering problem [JB02, dLE07, d'E12], all follow the general form of Eq. 3, with a few exceptions [DJ07, HCJ13].

4.3.2. Parameter Estimation

When considering subsurface scattering in a homogeneous medium, there are only two parameters to estimate for the diffusion approximation: the absorption coefficient, σ_a, and the reduced scattering coefficient, σ_s'. In their introduction of the dipole method to computer graphics, Jensen et al. [JMLH01] also describe how to fit the parameters from a diffusion profile. The profile can be observed by photographing a material sample illuminated by a normally incident beam of light. Each pixel in the captured images gives a sample point of the profile, resulting in hundreds of observations. Jensen et al. report the fitting to be ill-conditioned and constrain the fit to preserve the total diffuse reflectance. Unfortunately, neither the setup nor the required measurement time of Jensen et al.'s approach are suitable to face measurements.

A more practical approach to the fitting uses the total diffuse reflectance to directly estimate one of the two unknowns. While neither σ_a nor σ_s' can be computed from it, an inversion of the total diffuse reflection allows to compute the reduced albedo α' [JB02], which together with the translucency l_d gives σ_a and σ_s'. The practical advantage is that the measurement of the total diffuse reflectance is very easy, it only requires a capture with spatially-constant irradiance, which is usually achieved by a simple full-on illumination capture. In contrast, this fitting method is potentially less robust compared to Jensen et al.'s constrained profile fitting, as it does not allow for additional constraints. The remaining parameter, translucency, has to be estimated from an observed profile. The following methods differ in the way they observe the profile on the face and how they overcome the difficult fitting procedure, an overview is given in Tab. 2.

Weyrich et al. [WMP*06] follow the idea of a single beam of incident light and built a small sensor head that could be placed safely on the face of a subject. The sensor head contains a linear array of optical fibers, one of which is the light source, while the others are sensors that capture a coarse sampling of the diffusion profile. From fitting σ_a and σ_s' they can derive translucency, which they average over all subjects. The average somewhat counterbalances the ill-conditioned nature of the profile fitting [JMLH01, TGL*06]. Furthermore, it allows them to use an inter-subject translucency, but a subject specific total diffuse reflectance. As total diffuse reflectance is only valid for planar surfaces, they use the average of all pixels of the observed total diffuse reflectance; hence, they obtain homogeneous subsurface scattering parameters. To add back details, they compute a modulation texture, which is simply the ratio of the per-pixel total diffuse reflectance observation and the total diffuse reflectance that one would observe with the per-subject parameters σ_a and σ_s'. Due to the inter-subject shared average translucency, they are able to fit subsurface scattering parameters for a new subject by simply computing total diffuse reflectance, requiring a simple capture with full-on illumination.

Tariq et al. [TGL*06] instead proposed measuring the reflectance profile by projecting stripe patterns on the face. Their approach also allows for spatially-varying parameters, assuming they are locally constant for the parameter fitting. To observe full profiles, the stripe patterns on the face need to be sufficiently spaced for the subsurface reflectance to be

Table 2: *Comparison of subsurface estimation methods. While the direct fitting requires a non-linear optimization, look-up tables (LUT) are queried by a 1D search by minimizing sum-of-squares difference metric.*

	Profile from	Fitting	Granularity
[JMLH01]	beam	optimization	object
[TGL*06]	stripes	LUT	pixel
[WMP*06]	beam	optimization	face
[GHP*08]	black dots	LUT	region
[ZGP*13]	curvature	direct	pixel

Figure 12: *Fitting a subsurface scattering model enables the simulation of the soft appearance of human skin. The right column shows different scattering layers: deep multiple scattering, shallow multiple scattering, single scattering, and specular surface reflectance [GHP*08]. Note that in contrast to Fig. 11, these layers are assumed to be additive.*

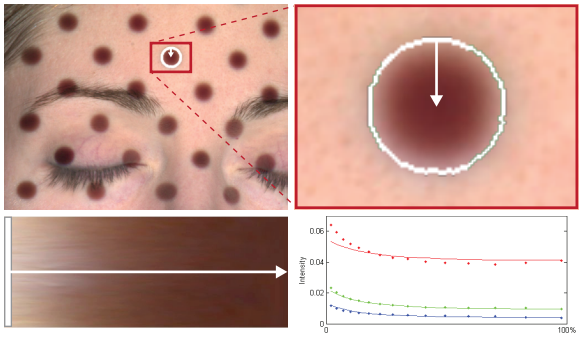


Figure 13: *Ghosh et al. [GHP*08] project black dots to sparsely acquire subsurface scattering profiles to fit the parameters. The lower left image shows the reparametrization, which encodes lines from the dot boundary to the center as rows in the image. Hence, each row represents a scattering profile (lower right).*

near-zero at the center of the unlit regions. As parameters are estimated per pixel, each pixel has a single observation of the profile due to one stripe pattern, although at different distances from the closest directly illuminated stripe. More observations are added by shifting the stripes pattern. To estimate the translucency from the profile, Tariq et al. utilize a 2D look-up table, which is first indexed by the total diffuse reflectance (projected white image), and second by matching the observed profile to precomputed profiles using a sum-of-squares difference metric. They fill the look-up table by rendering profiles assuming a normal incident illumination with stripe patterns.

Tariq et al. and Weyrich et al. estimate parameters for the dipole model, which assumes a semi-infinite homogeneous medium. Inspired by the actual structure of skin, Ghosh et al. [GHP*08] estimate parameters for a multi-layer model (Fig. 12). However, instead of convolving the two layers as dictated by Donner and Jensen's [DJ05] physically-based multi-layer extension of the dipole, the contributions of two layers are simply added. This simplification makes the fitting tractable, but results in a data-driven two-layer model without bio-physical meaning. As with diffuse–specular reflectance, the difficulty lies in separating the shallow and deep scattering contribution to fit the parameters of each layer individually. To get a separate total diffuse reflectance, they adapt the direct/indirect illumination technique of Nayar et al. [NKGR06]. Projecting four phase-shifted high-frequency patterns of 1.2mm stripes gives per-pixel $(\min, \max)$ values. The difference $(\max - \min)$ is taken as the shallow (direct in Nayar et al.) and $2 \times \min$ the deep reflectance (indirect). To estimate translucency, Ghosh et al. propose to capture the diffusion profile by projecting a white pattern with black dots on the subject's face (Fig. 13). The diffusion profile can be directly observed going from the edge of a dot to its center. Separating the shallow and deep scattering profiles follows from an assumption that shallow scattering decays quickly. Hence, they use the inner two thirds of the profile (from the dot center) to estimate the deep scattering, subtract the fitted profile from the observations, and fit the residual as the shallow scattering profile. The actual profile fitting uses a look-up table similar to Tariq et al. [TGL*06]. The dot patterns allow them to measure per-region translucency.

Following a recent trend in illumination bases, Zhu et al. [ZGP*13] derive the diffusion parameters from spherical gradient illumination. To avoid profile sampling and fitting, they require a curved sample and add a curvature parameter to the dipole equations, although this violates the boundary constraints of the dipole. Finally, they estimate the translucency by assuming an observation of the same sample point with and without curvature under different illumination, which gives them a ratio from which they derive translucency.

4.4. Biophysically-based Skin Models

While the previous methods use a combination of generic surface and subsurface models to fit observations, researchers have also proposed models specific to skin. The main advantage of such models is that they are tightly coupled to the actual structure and biological nature of skin. This allows the use of measured data of skin pigments from the biological and dermatological community, similar to the recent trend in rendering to use physically-based materials, which leads to models with an increased predictable behavior that can represent different skin types as well as biological processes such as blushing or tanning.

Before introducing the various skin models, let us first describe the actual structure of skin, illustrated in Fig. 14, as well as its interactions with light. Skin is covered by an oily layer, on which light can reflect directly. Light that does not reflect off the oily layer then travels through the various layers

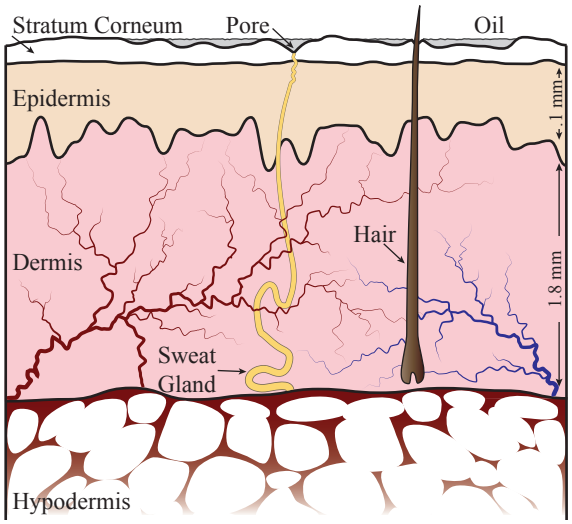


Figure 14: *Diagram showing a decomposition of skin into layers. From top to bottom we see the oil layer mixed with the Stratum Corneum. Below that we have the epidermis layer consisting mainly of melanin pigments responsible for the brownish-yellowish color of skin. Then we have dermis mostly consisting of blood cells with hemoglobin which give the strong red color. Finally there is hypodermis which consists mostly of fat cells that reflect most of the light.*

of skin. The outer layer of skin, the stratum corneum, consists mostly of dead cells and only has a marginal influence on the overall appearance since it is a thin forward-scattering medium with a very high albedo [KB04]. The next layer, the epidermis, consists mostly of melanin that reduces UV radiations, but also affects the visible spectrum, and causes colors from light yellow to dark brown or black [DJ06]. The dermis layer is thicker and consists mostly of hemoglobin, a pigment present in blood cells that binds oxygen. Hemoglobin comes in two different forms, oxygenated and deoxygenated, both having complex spectral scattering properties, that need to be properly modeled and simulated in rendering. Oxygenated hemoglobin has a strong red appearance and a characteristic W-shape around the 550nm wavelength, which is not present for deoxygenated hemoglobin. The difference in the absorption spectra is distinctive that it is used to measure the amount of oxygen in blood. All of the skin models described in the following section employ a spectral description of the scattering coefficients to accurately model the scattering. Additional small-scale cellular structures in the dermis (sweat glands and ducts, hair follicles, muscles, capillary veins, etc.) also influence the appearance, though each one only has a minor individual influence on the overall appearance. Finally, the hypodermis layer consists largely of fat cells, which strongly reflect light [KB04], and it is therefore reasonable to neglect light that penetrates deeper into the skin. Other features are

Table 3: *Comparison of specific skin models.*

	Type	Layers	Parameters	Evaluation
[KB04]	BSDF	5	>10	Monte Carlo
[DJ06]	BSSRDF	2	4	Multipole
[DWd*08]	BSSRDF	2	6	Multipole

visually important, such as the veins that are roughly at the boundary between the dermis and the hypodermis, or freckles and other surface imperfections. Skin thickness, which varies across the face and the body as a whole, also affects appearance. For example, the ears and nose wings are very thin with a low amount of fat cells, which causes a very characteristic translucent look.

For most purposes in graphics, modeling skin and its interaction with light at a molecular or cellular level would be prohibitive. Skin is therefore usually modeled at a macroscopic level, simplifying light interaction to scattering with particles following statistical distributions. In practice, the various skin-specific models make a tradeoff between quality and performance, which is an important practical concern. The complexity of skin, however allows for various choices in abstraction and level of detail, which is one of the main differentiators among the available models (see Tab. 3).

Krishnaswamy and Baranoski [KB04] present the BIOSPEC model that accurately describes the structure of human skin. The model mimics a five layer skin consisting of the stratum corneum, epidermis, dermis (papillary and reticular), and hypodermis (cf. Fig. 14). Due to the multilayer nature and the description of the layers with pigment concentrations, analytic subsurface models such as dipole diffusion cannot be applied for the simulation and Krishnaswamy and Baranoski rely on a brute-force Monte Carlo estimation. To make the evaluation of the model feasible, an interface-to-interface tracing approach is used instead of a physically-based volume rendering—somewhat counteracting the original goal of an accurate model. Hence, scattering is solely computed at the layer interfaces, although scattering in real skin is caused by pigments, structures, as well as a continuous change of index of refraction everywhere within the skin. In the BIOSPEC model, pigments within the layers only absorb light due to Mie and Rayleigh scattering. While the simulated photons have fixed direction along the surface normal, they compute random directions (using a Rayleigh phase function) at the interfaces to adapt the absorption computation to the next interface. Hence, the spatial position is not changed and the overall result is a BSDF (a BRDF defined over the full sphere). Due to the shear number of—and complex interaction between—parameters, it is difficult to deviate from previously measured values to simulate different skin types.

Donner and Jensen [DJ06] were the first to propose a biophysically motivated BSSRDF model, but with the two main

shortcomings of the BIOSPEC model in mind: parameter space and performance. They model the human skin with two-layers and a minimal amount of biophysically-based user parameters that cover a large class of human skin types. More importantly, they chose parameters with an intuitive behavior due to their distinctive effect on appearance: melanin and hemoglobin fraction, a blending weight between eumelanin and pheomelanin, and an oiliness parameter. These parameters allow simulating Caucasian, Asian, as well as African skin types. While in reality more parameters are involved, Donner and Jensen reasoned that the remaining parameters can be easily fixed or derived with only a minor impact on the space of possible skin types. To make the approach computationally efficient, they make use of the multipole diffusion model [DJ05] instead of a Monte Carlo evaluation.

In follow-up work, Donner et al. [DWd*08] increase accuracy by adding more parameters, but also extend the model to handle heterogeneous parameters. Heterogeneous media is difficult to simulate as the dipole and multipole approaches assume a homogenous medium which induces a radially symmetric reflectance profile. To make the diffusion model applicable, Donner et al. assume a "slow variation relative to the mean free path of light". Hence, the final reflectance profile is the convolution of multiple radially symmetric intermediate profiles due to reflectance and transmittance, just as for the multi-layer model [DJ05]. By making the profile dependent on the point of exitance of a layer, and storing intermediate convolved results in texture space, Donner et al.'s [DWd*08] final reflectance profiles can be non-symmetric. To make this computationally feasible, they express radially-symmetric profiles using a Gaussian mixture model [dLE07], which turns the profile convolutions into Gaussian blurs. They start by applying a spatially-varying Gaussian blur in *uv*-texture space to an irradiance map, which simulates one convolution. The overall reflectance is calculated in a multi-pass procedure convolving the results of the previous pass with the spatially-varying parameters of the next layer. While slowly-varying heterogeneous media was already considered in the original dipole approach [JMLH01] (by making the profile size dependent on the light incidence position), Jensen et al.'s resulting profiles were always radially symmetric since they only considered a single semi-infinite medium. Finally, Donner et al. [DWd*08] add a modulation texture between the two layers to support strongly absorbing structures not on the skin surface, such as veins or tattoos. Since these are contained beneath the epidermis, their appearance is soft and difficult to model using textures on the skin surface.

The skin models are able to represent typical skin appearance quite accurately. Thanks to the low parameter models of Donner et al. [DJ06, DWd*08], the models can be practically used for actual modeling of virtual actors. However, these models have two practical shortcomings. Firstly, it is difficult to fit parameters from observations due to the strong non-linearity of the parameters. While Donner et al. [DWd*08] fit parameters for small skin patches and use texture synthesis to model an entire hand, a fitting for a complete face remains open. Furthermore, their fitting excludes the visually important inter-layer absorption texture, which needs to be modeled manually. Additionally, the models are missing distinctive skin features that make the appearance unique. Manual modeling is difficult and tedious, and alleviating this is the main motivation of the capture techniques discussed in this STAR. It remains open if such models could avoid fitting the non-physical albedo map used to represent details.

4.5. Dynamic Appearance

The aforementioned parametric approaches ignore the change of facial appearance over time. This is because, in contrast to image-based methods, effects such as change in viewpoint, occlusion, or interreflection can be fully simulated during rendering by means of global illumination algorithms. An animated mesh with a consistent parametrization between surface points and texture space easily allows propagating material data from one frame to another, guided by the change of the geometry. Hence, a large body of work focuses on performance capture to estimate geometry and establish a consistent texture parametrization. A combination of dynamic geometry and static skin properties allows for convincing results. While image-based, the method of Hawkins et al. [HWT*04] follows this approach and assumes a unique mapping from geometry to appearance. Yet, there are also dynamic skin effects that affect appearance but do not cause a change in geometry such as sweating, cold, blushing, etc. [WMP*06].

Jimenez et al. [JSB*10] focused on the visual effect of blood perfusion due to changes in geometry. Their appearance model is based on a simplified version of Donner et al. [DWd*08]. For dynamic appearance, they only change the hemoglobin concentration, but leave melanin untouched. They observed subjects performing different facial expressions and conclude as follows. Firstly, there is a measurable change in appearance due to a change of blood perfusion. Secondly, there is a minimal delay between change in expression and change in appearance. Finally, they noted that the appearance for different expressions is similar between subjects on a large scale, but has a subject-specific pattern on a small scale. They propose to model expression-dependent spatially-varying concentrations of hemoglobin by storing a base concentration for a neutral expression as well as the per-expression relative change. To encode the relative changes, they first note that a hemoglobin concentration map can be represented by spatially-varying histograms of local neighborhoods. Interestingly, the histograms exhibit a Gaussian shape, and hence, can be represented by their mean and variance. Due to this encoding, the relative change boils down to a per-pixel shift and scale operation. There are two advantages of this representation. The base concentration map can be artistically altered while preserving a valid change to other expressions, i.e., appearance transfer between subjects becomes possible. Furthermore, the shift and scale maps are of

Figure 15: *Dynamic skin appearance can be modeled as a function of facial expression. The model of Jimenez et al. [JSB*10] estimates blood perfusion of facial expressions, defining parameters for subsurface scattering. The effect is mainly perceived as increased redness in the appearance (e.g., see the forehead). Although subtle, this is an important step towards highly believable virtual characters.*

low frequency, allowing for compression. To limit the amount of capture and storage, Jimenez et al. model only six different expressions, i.e., one base map and six shift and scale maps. They follow the discriminative Ekman [EF78] emotions that fit facial expressions: anger, disgust, fear, happiness, sadness, and surprise. To support other intermediate expression, they rely on the concept of blendshapes, which uses a weighted combination of vertex positions to allow for seamless blending between expression. They couple the weights with a linear blending of the shift and scale maps. However, the validity of this blending is not clear. While they note that there is no delay in the blood perfusion between expressions, the transition, which is required to depict other expressions, is not necessarily linear. The results of their method are subtle, but likely important to overcome the uncanny valley of virtual actors (Fig. 15). Besides blood perfusion changes, they also show results for alcohol drinking and exercises. Recently Iglesias-Guitian et al. [IGAJG15] extended parametric models to also take into account the aging of skin, which allows to simulate how the appearance of a person changes over the years. Extensions to sweat, heat, or long term effects such as tanning remain to be explored.

4.6. Surface Details

Geometric details on the skin surface can greatly affect the appearance of the skin, in particular fine scale variations of the surface orientation. At a microscopic level, these variations are not individually perceivable and are thus often quantified statistically and directly integrated into the appearance model (e.g. Torrance-Sparrow). At a macroscopic level, the surface normals can be computed from the reconstructed geometry. On a mesoscopic scale, surface features are visually apparent but too small to be reconstructed as geometry directly. Skin pores and fine scale wrinkles typically fall into this last category. Over the last decades several methods have been proposed that try to estimate the normals directly, and providing a comprehensive overview of this body of work is outside the scope of this report. We therefore focus on work that deals with estimating mesoscopic surface structure in the context of faces. Generally speaking, this work can be categorized into active and passive techniques.

Active Illumination. Active techniques illuminate the face with a well defined set of lighting conditions. Photometric stereo [Woo80] is certainly the most widespread active method for facial mesostructure estimation. While, in theory, three lighting conditions suffice to estimate normals on Lambertian surfaces, in practice, the non-flat, non-convex shape of the human face causes self-shadowing and the reflectance is far from Lambertian. Therefore, Weyrich et al. [WMP*06] make use of a light stage (Sec. 2) and capture 150 different illumination conditions, temporally multiplexed over 25 seconds. This long exposure is problematic, since the person unavoidably moves during the capture process, leading to mis-registrations of the data. Ma et al. [MHP*07] drastically reduce the illumination conditions required by using a different illumination basis. They further show that normals computed from the direct surface reflection only contain higher spatial detail and they thus separate the specular and diffuse reflection. Still, their method requires eight frames which are temporally multiplexed and thus require the actor to remain still during acquisition. Wilson et al. [WGP*10] aim at overcoming this issue by time-multiplexing complementary illumination conditions and use optical flow to align them. Still, optical flow will only work for small motion during acquisition, which entails fast cameras or slow moving actors. To overcome the temporal integration problem completely, authors have started to look into spectrally multiplexing the illumination conditions instead [BHV*11, FYD11, VH12]. These methods introduce a different cost, however, in that that they require white makeup to be applied or they need to consider the spectrally dependent attenuation of the skin.

To summarize, active techniques can generally recover the surface details at extremely high resolution, but usually require time multiplexing to acquire the different lighting conditions, which is problematic if one wants to capture objects in motion.

Passive Methods. In contrast to active methods, passive techniques have the advantage that they only observe the scene. Typically these methods require only a single frame to estimate the structure but yield less accurate results than active methods.

Beeler et al. [BBB*10] introduce the concept of mesoscopic augmentation, where they combine multi-view stereo with shape-from-shading. Their work assumes constant omnidirectional illumination and is thus limited to studio conditions. Wu et al. [WWMT11] extend this concept to arbitrary illumination by also estimating the environment map. One of the core problems of these methods is to distinguish between texture and shading. This separation is ill-posed for a single image and thus Beeler et al. [BBZG12] propose a follow-up approach that uses temporal information to overcome this.

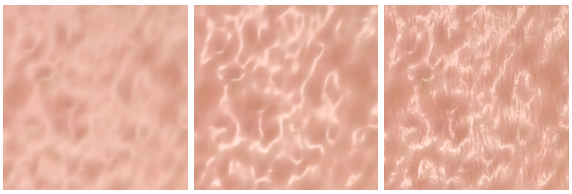
Level of Detail. All above methods, both passive and active, create surface details at a single scale only. The structure of surface detail at various scales, however, has a combined impact on appearance. This coupling must be accounted for both during rendering and during capture.

On the rendering side, displaying high-frequency content at a distance requires filtering to avoid aliasing artifacts or expensive super-sampling. The most common way of filtering is texture mip-mapping, which blurs the input at multiple levels to remove high-frequencies. Unfortunately, such linear filtering of geometric detail in normal or displacement maps induces a strong change in appearance, resulting in overly specular appearance for far away objects. To preserve appearance, variations present in the high resolution normal or displacement mapped geometry must be transferred to the surface roughness as part of the BRDF. Several recent techniques [OB10, DHI*13, BN12] have considered this for the general rendering problem. In the context of faces, von der Pahlen et al. [vdPJD*14] employed a custom mip-map filtering for the normal map, which concurrently filters a gloss map to increase roughness.

The problem occurs in exactly the opposite direction during capture. Due to the averaging occurring within the extent of a single pixel, fitting micro-facet roughness to lower resolution captures results in overly diffuse appearance when the surface is view up close. To combat this problem, Graham et al. [GTB*13] capture microscale detail patches up close using a miniaturization of the system proposed by Ma et al. [MHP*07] and propose fitting the specular roughness from these samples instead of the mesoscopic full facial scan. This has a dramatic effect on the specular appearance of the skin as illustrated by the difference between Fig. 16 (a) and (b). Furthermore, they use the microscale patches to inject high-frequency normal map details into the mesoscopic full face scan using a constrained texture synthesis framework. This results in further resolved details as seen in Fig. 16 (c).

4.7. Facial Hair & Velvety Skin

Large parts of the human skin are covered with hair of different size, shape and color—ranging from the dense and very prominent scalp hair to the sparse and almost invisible vellus hair coating the skin, often referred to as "peach fuzz." In this report we limit ourselves to methods that are concerned with sparse, short hair coverage, where the underlying skin is visible and the hair thus contributes primarily to the skin appearance.



(a) mesoscale scan + mesoscale BRDF (b) mesoscale scan + microscale BRDF (c) microscale scan + miscroscale BRDF

Figure 16: *Graham et al. [GTB*13] capture the full facial mesoscale geometry, which they enhance with microscale details from zoomed-in measurements of small skin patches.*

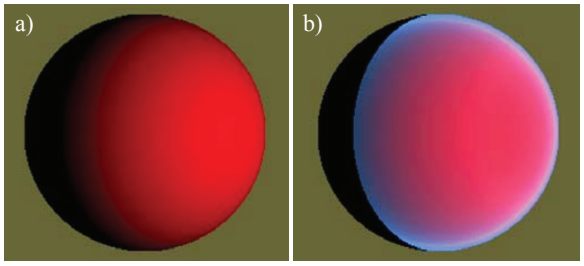


Figure 17: *A sphere rendered with Lambertian shading (a), and the same sphere rendered with additional asperity lobe (b) [KP03]. The typical effects of asperity scattering are clearly visible: the light bleeding at the shadow boundary to the left as well as the bright halo at the occluding edge to the right. These effects are in opposition to those predicted by standard Lambertian shading.*

Koenderink and Pont [KP03] motivate the importance of peach fuzz. They show that vellus hair produces bright halos at occlusion boundaries and light bleeding at shadow boundaries—effects which directly oppose Lambertian shading. These effects are strongest when the light and viewing directions both span a large angle with the surface normal. Koendrik and Pont account for these effects by adding an additional lobe to the surface BRDF, called the asperity lobe. They model the asperity scattering by assuming the surface is covered with an atmosphere of a certain thickness and extinction coefficient in which single scattering occurs. This extended BRDF reproduces the asperity effects directly and homogeneously on the surface as shown in Fig. 17. Similarly, a layer of short non-shadowing curves was found necessary to provide some missing visual elements on the rendered baby for the movie Lemony Snicket [Kur11].

The asperity lobe approach may suffice for distant viewing, where the thickness becomes negligible and the distribution of the scatterers approaches visual homogeneity. In closer

Figure 18: *Vellus hair on a female face and the visual importance on appearance [KP03]. The effect of this "peach fuzz" is most apparent under extreme illumination conditions, such as contre-jour. At this scale, individual vellus hair fibers are clearly visible and modeling their effect with a spatially homogeneous asperity lobe would not suffice.*

depictions however, such as the ones shown in Fig. 18, it is important to capture and represent the spatially varying distribution and shape of the vellus hair.

The work by Beeler et al. [BBN*12] presents a step in this direction as it attempts to capture and reconstruct the individual hair fibers that cover the face. This allows re-rendering the data faithfully, reproducing the spatially varying look of facial hair. This approach, however, requires individual hair fibers to be visible by multiple cameras to conduct multi-view reconstruction, making the method suitable for capturing thicker, more prominent facial hair. Capturing and reconstructing the near-invisible vellus hair is still an open problem.

4.8. Conclusion

Splitting the problem of describing the facial appearance into sub problems, which can be handled by specific models independently, enables a full parametric description. The main benefit is not just the reduced storage, but a complete description of appearance, i.e., a rendering from any viewpoint or under any illumination with virtually any geometry becomes possible. It also allows for artistic control, which we will detail further in Sec. 5.5. However, the observations have to be separated first, such the respective data can be actually fitted to the corresponding model. The fitting procedure itself is difficult, especially for the subsurface scattering component due to the complex structure of skin, which causes the reconstruction to be ill-conditioned. Hence, accurate parameter estimation requires many observations, which prevents current parametric methods from real-time appearance capture.

The discussion so far excluded any details about the actual capture process, which also poses problems.

5. Practical Considerations

So far we have described a number of methods for capturing skin appearance, focusing primarily on the scientific aspects and theoretical capabilities of each method, which at least partially assume an ideal world. In practice, appearance capture has a number of additional constraints that need to be considered. This list is not meant to be exhaustive, but rather hints to some of the practical challenges that need to be considered for face appearance capture.

5.1. Polarization

In the context of face appearance capture, polarization is predominantly used to separate the diffuse and specular components of reflectance, by exploiting the fact that specular reflection preserves the polarization state of light, whereas the diffuse part does not (refer to Sec. 4.1). Practically, however, multiple factors constrain the use of polarization: firstly, the view-dependency of the polarization orientation; secondly, the need for physically changing the filter orientation (of either the camera or light source) to capture parallel- as well as cross-polarization; and lastly, the dependence of the amount of reflected light on the angle of incidence.

The physical basis for the latter is as follows. Light reflection at a smooth surface follows the two Fresnel equations for specular reflections of s- and p-polarized light. The s-polarized light is orthogonal, while p-polarized light is parallel with respect to a reference plane, defined by the two vectors, light incident direction and the viewing direction. As the reflection is perfectly specular, the surface normal needs to be the half vector of both vectors. Importantly, the s and p components are scaled depending on the angle of the incident light (as well as the indices of refraction; see Fig. 19 left). This needs to be considered to scale the observed specular reflection accordingly. Further, the p component goes to zero at the Brewster angle, hence, even without a filter in front of the camera, no specular reflection is observed if the incident light is fully p polarized, which depends on the orientation of the polarizer in front of the light source.

Linear Polarization. Ma et al. [MHP*07] use a single-camera setup, and fix the camera viewing direction and its polarizer orientation and optimize the orientation of the polarizers of the light sources. This causes perfect cross polarization and can be turned into perfect parallel polarization by rotating the polarizer of the camera by 90 degrees. The incident angle dependence of the reflection of s- and p-polarized light leads to holes in the captured data, as shown in Fig. 19, right.

In a multi-camera setup, it is not possible to achieve perfect cross- or parallel-polarization, since the various cameras

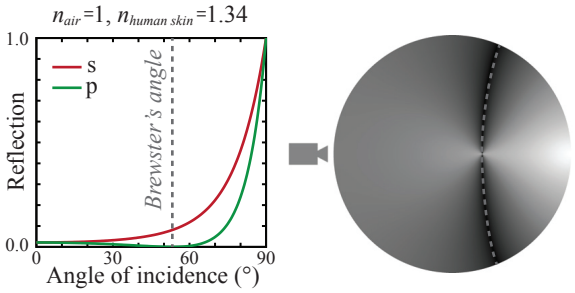


Figure 19: *Linear polarization and Brewster's angle. Left: The dependency of the fraction of reflected light on the angle of incidence of the light. Right: Fraction of expected observed reflected light for parallel polarization with an optimal setup for a single viewing direction [MHP*07]. At the Brewster angle, no reflection can be observed.*

have different viewing directions. To address this limitation, Ghosh et al. [GFT*11] propose a latitude and longitude polarization pattern. The introduced error is low and due to the symmetry in the patterns, it allows for placing cameras around the equator. More interestingly, this setup addresses the issue of physically rotating polarizers. To get parallel- and cross-polarization Ghosh et al. switch between the latitude and longitude pattern in two subsequent captures, i.e., they virtually change the orientation of polarizers of the lights by using a different set of physical light sources. This approach, however, only works for smooth illumination bases such as the spherical gradient illumination to avoid differences in the illumination for the parallel- and cross-polarization captures. Finally, the latitude and longitude patterns allow for a horizontal as well as vertical polarization filter in front of the camera. However, due to the incident angle dependence, the vertical orientation is superior [GFT*11].

Circular Polarization. Circular polarization (light with equal amount of *s* and *p*-polarization, but both phase-shifted) is an alternative to the commonly used linear polarization [MHP*07, GCP*10, GFT*11]. The main advantage is that it does not share the view-dependency of linear polarization. However, incident circularly-polarized light becomes skewed after reflection (as the *s* and *p*-components are scaled differently at a reflections; see Fig. 19 left), resulting in elliptical polarization with unequal *s* and *p* components. Elliptically-polarized light is then only partially filtered by a circular-polarizer. The effect increases with the steepness of the incident light, largely preventing any cancellation for angles larger than 70 degrees [MHP*07]. While this makes it impractical for diffuse–specular separation, Ghost et al. [GCP*10] still capture circular polarization as part of their estimation of the Stokes vector that describes the polarization state of light. They further show how the degree of circular polarization can be correlated to the roughness of the surface (refer to Sec. 4.2).

5.2. Cameras

While a pinhole camera model allows for elegant mathematical modelling, physical realities oftentimes pose severe limitations to the theory. Some of these limitations, such as lens distortion and vignetting, can be alleviated through careful calibration. Some, however, cannot be avoided. For example, cameras have a physical extent and block light, which can pose problems for algorithms that assume omnidirectional illumination, especially if multiple cameras are employed.

The Sharpness/SNR Dilemma. The main physical limitation of cameras is the tradeoff between sharpness and the amount of incident light, which directly impacts the signal-to-noise-ratio (SNR) of the camera. Lenses integrate light angularly by redirecting and focusing rays from different directions onto the sensor. Unfortunately, this limits the depth range that may be imaged in focus, the so called depth-of-field. The camera aperture is used to control the angular domain that is integrated, and reducing it will increase the depth-of-field, but at the same time also reduce the amount of incident light. Furthermore, lenses are typically built for a specific aperture range, and reducing it beyond this range might cause an overall loss of sharpness due to diffraction effects. The second option is to integrate light temporally through increased exposure times, which will cause motion blur for moving objects. A third option is to increase light spatially by employing larger pixel areas, which limits the spatial frequencies that can be resolved and thus reduces sharpness. Finally, integrating light spectrally leads to an averaging of the color features which can be considered a loss of sharpness as well. In the end, a tradeoff has to be made between depth-of-field, SNR, and motion-blur and the optimal tradeoff depends on the use-case. Additionally, the use of polarization filters (as described above) reduces the amount of incident light substantially ($\sim 50\%$).

Dynamic Range. Another point to consider is dynamic range: the ratio between the maximum and minimum measurable light intensity. Even modern cameras have limited dynamic range and may not suffice to capture all elements in a scene at a suitable exposure. Beeler et al. [BBN*12], for example, report that they were not able of capturing dark facial hair and fair skin in a single exposure. High dynamic range imaging (HDR) could alleviate this problem, but most existing HDR cameras require temporal multiplexing [RBS03], which again requires the subject to remain very still. Therefore, most facial appearance capture approaches do not make use of HDR. An exception is the work of Tunwattanapong et al. [TFG*13], which thus requires an acquisition time of approximately 10 minutes. While suitable for inanimate objects, this is not practical when acquiring human faces.

5.3. Illumination

Just as for cameras, physical realities also impose limitations on the illumination hardware, some of which may be compen-

sated through careful calibration and others that unfortunately cannot be avoided. For one, distant lights are often not really distant, and as a consequence incident rays are not parallel. Also, point lights are not really points but area light sources. Furthermore, light has a quadratic spatial falloff and typically also an angular one. These effects may be overcome with careful calibration [WMP*05, WMP*06] or can be safely ignored when the relative distances of the light to acquired surface area is sufficiently large.

The sharpness/SNR dilemma discussed previously could in theory be solved with stronger illumination. Current illuminators such as LEDs can provide high illumination density (the amount of light emitted per area) but typically still have to employ lenses to focus the light, and as such suffer from analogous problems as cameras. Apart from angular amplification through lenses, another option is to add more lights spatially and turn them on at the same time. The downside here is that the angular resolution of the incident illumination is reduced since multiple light sources act effectively as a point sampled area light source. Wenger et al. [WGT*05] investigate different spatial arrangements in their light stage. They show how activating three neighboring lights increase the SNR of the captured data over the canonical single light case, and report only a small decrease of quality due to the skin BRDF. They also investigate the use of a different distribution pattern, the Hadamard basis, which randomly activates half of the lights. In theory, this basis allows for the same spatial resolution of the environment light as the canonical case, but with drastically higher incident light per measurement. In practice, however, the required inversion performs poorly in low light areas such as shadows, and Wenger et al. conclude that it is inferior compared to the canonical case.

Perhaps the best illumination solution would be to increase the illumination density, and future technology is likely to do so. However, as lights reach ever higher illumination densities, the ultimate upper limit on the amount of illumination will be imposed by the human factor, as described next.

5.4. The Human Factor

One aspect that is oftentimes ignored in theory is the fact that the intent is to capture a human being. This puts into place additional limitations, for example the amount of illumination that can be applied before the subject starts squinting or has to close his eyes altogether. Furthermore, flashing lights may not only cause discomfort but could potentially be hazardous [WGT*05]. Also, even if the subject is not performing, but trying to hold an expression, he will not be able to remain perfectly still during longer stretches of time. This poses limitations to methods that require temporal multiplexing or integration, which are even more severe when acquiring dynamic effects such as blood flow. Outside of the research environment, the fact that the person whose facial appearance you want to capture may be a highly-paid, famous actor puts even more serious constraints on capture time and safety concerns.

An interesting track for mitigating human constraints is to reuse, or transfer, as much data as possible, in order to minimize the incremental data that actually has to be captured. Peers et al. [PTMD07] propose a method based on quotient images [RRS99] that capture the ratio of the reflectances due to a desired illumination and constant spherical illumination. Assuming large scale geometry can be warped between two subjects A and B, this allows transferring the reflectance of A to B as well. To do so, the reflectance of subject A needs to be captured using any of the methods discussed in Sec. 3 to obtain the quotient image Q. Capturing the performance of subject B only under uniform illumination and multiplying by a warped version of Q gives a relit B. Tunwattanapong et al. [TGD11] show how multiple quotient images can be combined artistically.

5.5. Artistic Control

In a production environment, it is common to have artists manually alter the appearance of virtual characters, usually by editing texture maps that encode the various attributes of the rendering model. In practice however, parameters often have complex, unintuitive behaviors that are difficult for artists to master. A simple example is drawing or editing albedo maps, since artists are used to incorporating shading cues in their art, which would result in an invalid albedo. While this issue can be addressed through practice and education, some other cases are more problematic. In the context of subsurface scattering for instance, the effect of some parameters can be very difficult to predict, even for more technical users with the proper mathematical background. The absorption and scattering coefficients of BSSRDFs models interact in a non-linear fashion that is not only proving problematic when fitting data (see Sec. 4.3), but also making them highly abstract for artists. Even the biophyiscally-based model of Donner et al. [DJ06], with its parameterization specialized to human skin (inputs covering such quantities as hemoglobin and melanin contents), does not lend itself to, for instance, easily editing in a mole on the subject. An alternative, initially explored by Jensen and Buhler [JB02] and Hery [Her03, Her12], instead obtains the BSSRDF parameters using an albedo inversion, giving much more meaningful and tunable knobs to the artists, namely the translucency and a regular diffuse albedo map.

6. Conclusion

The state of the art in facial appearance capture has evolved drastically over the last two decades. Today, image-based methods can achieve stunning photo-realistic results, though they do so at the cost of an extensive capture process (see Tab. 4). However, practical applications typically need a more flexible framework, where data can be captured rapidly,

Table 4: *Comparison of image-based methods that support relighting and are discussed in Sec. 3. SV denotes spatially-varying. (*) This method does not sample SV illumination but can synthesize it.*

	Relighting	Dynamic geometry & appearance	Novel viewpoints	SV illumination	Sampling dimensionality	Illumination basis
[DHT*00]	✔	✘	✔	✘	6D	canonical
[DWT*02]	✘	✔	✘	✘	3D	canonical
[HWT*04]	✔	✔	✔	✘	7D	canonical
[WGT*05]	✔	✔	✘	✘	5D	canonical & hadamard
[ECJ*06]	✔	✔	✔	✘	7D	canonical
[JGB*06]	✔	✔	✘	✔*	5D	canonical & structured
[TGD11]	✔	✘	✘	✘	4D	canonical & spherical

edited, and extrapolated. This has motivated the development of parametric methods that can offer more of this flexibility. This is illustrated by the recent trend in research, with a clear focus on parametric methods (see Tab. 5). While this has resulted in vast improvements—to the point where parametric methods can now achieve near photo-realism—there are still significant challenges ahead that should be addressed in future work.

Accuracy of Parametric Models. Current parametric models are still not able to accurately represent the full and often subtle complexities of facial appearance. For instance, detailed vellus hair is not considered by existing methods. In general, most existing models rely on assumptions that are actually invalid in practice. In the case of subsurface scattering, many methods rely on the dipole (or multipole) diffusion model and use its BRDF approximation under the assumption of uniform illumination. This has clear shortcomings. These models are only valid for homogeneous media, but are often used for the estimation of spatially-varying parameters. Furthermore, they assume a flat surface, which is not true especially in the regions of the nose and ears. While these limitations do not necessarily preclude reaching acceptable results, there is still clearly room for improvement.

Model/Capture Discrepancy. The inaccuracy of existing parametric models is compounded by the fact that discrepancies in the capture setup often introduces further approximations. For example, when fitting surface parameters, interreflections on the face (such as light reflecting off to nose towards the cheek) is neglected. Another example is the optical separation of diffuse and specular components using cross-polarization. This does not strictly filter out specular reflection but also filters light that scattered a few times and did not lose polarization yet. This additional component is included in the fitting of surface reflectance, which is not explicitly accounted for by the surface reflectance model. Also, when fitting the dipole parameters, it is common to assume the skin surface to be a smooth dielectric, even though it is known to be rough (and usually modeled as such when fitting the specular reflectance). Future work could consider improving upon the state of the art by either relaxing the assumptions of the parametric model to better match reality, or by ensuring that these assumptions are better satisfied during capture. One example of the latter strategy was proposed by Donner et al. [DWd*08], who applied a thin film of ultrasound gel to skin, effectively smoothing out the rough surface. This approach, may however, not be entirely practical for full facial appearance capture.

Editing and Control. Parameters of existing models are currently not easily editable by artists [SPN*14]. Ideally, models should be efficient to evaluate, their parameters should have a physical meaning, should be easy to fit, intuitive to edit, and the required capture should be practical and cheap. In practice, however, many of these goals are conflicting with each other. In particular, bridging the gap between parameters that are intuitive to edit while having a physical meaning is an interesting area for future work.

Human Perception. A wide range of research has aimed at photorealism, but it is unclear which effects are essential for bridging the uncanny valley of human facial appearance. Deeper insights into human perception of facial appearance could help to derive new methods that offer better trade-offs between accuracy and efficiency. Unfortunately, this is a complex problem where the visual impact and perception of one feature is difficult or impossible to isolate from others.

Beyond Humans. Movie productions occasionally feature alien creatures, or other creations, for which facial appearance must be created. Interestingly, humans tend to map their every day observations to these creations and have implicit expectations of what alien organic skin should look like. It would be interesting to develop a framework where current facial appearance data could be extrapolated to produce fantasy creatures that still appear plausible to human observers.

Live Action. Since the estimation of parameters or the estimation of the reflectance function requires the observation of many measurements, it is in general difficult to capture them on a per frame basis. Motion compensation is strictly necessary for spatially high-frequency changing parameters such as albedos or normals. Clearly, this becomes easier with

Table 5: *Comparison of parametric approaches the estimate model parameters as described in Sec. 4. In this table we make the explicit distinction between local point illumination basis that are not assumed to be at infinity and distant canonical illumination basis, which denotes point light sources at infinity as used by the image-based methods. We use the following abbreviations for the surface models: LUT – lookup table, TS – Torrance-Sparrow model [TS67], Ward – Ward model [War92], KSK – Kelemen model [KSK01]*

	Dynamic appearance	Diffuse–specular separation	Surface model	Translucency	Subsurface model	Illumination basis
Surface appearance						
[MWL*99]	✘	polarization	LUT	✘	✘	local point
[MGR00]	✘	polarization	LUT	✘	✘	local point
[Geo03]	✘	✘	TS	✘	✘	local point
[FBLS05]	✘	✘	TS, Ward	✘	✘	local point
[GCP*09]	✘	polarization	TS, Ward	✘	✘	spherical gradient
[GCP*10]	✘	polarization	TS	✘	✘	spherical gradient
[GTB*13]	✘	polarization	TS	✘	✘	spherical gradient
Subsurface appearance						
[WMP*06]	✘	computational	TS	✔	dipole	distant canonical
[TGL*06]	✘	polarization	✘	✔	dipole	structured
[GHP*08]	✘	polarization	TS	✔	di-/multipole	spherical gradient & structured
[ZGP*13]	✘	polarization	✘	✔	dipole	spherical gradient
[JSB*10]	✔	✘	KSK	✔	screen-space diffusion	-

high-speed cameras as inter-frame motion is reduced. Going even one step further, it would be interesting to estimate the parameters in real-time, similar to recent advances in performance capture [BWP13, LYYB13, CHZ14].

Capture in the Wild. Most setups operate in very controlled environments, with a confined working volume which restricts the motion of the actor. Relaxing these constraints already works for performance capture [KSS11], however, for appearance capture it still poses interesting research challenges and opens up new application areas.

Fabrication. Current methods have achieved impressive results for acquiring real-world data and replicating it in a virtual environment. It would be interesting to investigate to what extent these methods could be employed for the computational design of physical replicas, such as prosthetics or animatronics, effectively closing the loop to the real world. The more stringent requirements of physical appearance replication will likely put increased pressure on the approximations currently deemed acceptable for virtual reproduction.

Scalability. From a practical point of view, creating a photo-realistic virtual actor currently requires a big budget, extensive capture hardware, and the talent of experienced artists. Investigating capture approaches and models that are more scalable would be extremely important to transfer what is currently possible for feature films to other application domains such as virtual reality, telepresence, games, or post-processing of personal video footage.

While addressing these problems is important for a faithful capture and reproduction of facial appearance, it is only one aspect of a complete representation of the human character. There are plenty of remaining challenges that are actively researched, including capturing eyes [BBN*14], teeth, hair [PCK*08], the tongue, the rest of the body, and clothing, which would have to all be integrated in a comprehensive model.

Contributor Biographies

Oliver Klehm, *Max Planck Institute for Informatics*
Oliver Klehm is a Ph.D. student of the computer graphics group at the Max Planck Institute (MPI) for Informatics and interned at Disney Research, Zürich. He received a B.Sc. in 2008 and a M.Sc. in 2011, both from the Hasso-Plattner-Institut. Oliver has published in the areas of interactive rendering as well as computational photography. He has developed methods for efficient forward rendering, but also investigated how these methods can be used for stylistic editing purposes—as its core, the fitting of parameters. He has also been involved in the construction of a recent generic camera add-on that allows for single-shot polarization, multispectral, high dynamic range, and light field capture.

Fabrice Rousselle, *Disney Research, Zürich*
Fabrice Rousselle is a Post-Doctoral Associate at Disney Research Zurich. He obtained a B.Ing. in computer engineering from the Ecole Polytechnique de Montréal in 2000, a M.S. in computer science from the University of Montreal in 2007, and a PhD in computer science from the University of Bern in 2014. Fabrice also worked as a software developer on video editing software at Matrox Video Group from 2000 to 2005, and as a research assistant in the colorimetry lab at the EPFL, where he worked on spectral reflection models from 2007 to 2009. Fabrice has been working on adaptive

Monte Carlo rendering for the last four years, leveraging image-based techniques to improve rendering times.

Marios Papas, *ETH Zürich / Disney Research, Zürich*
Marios Papas is a Ph.D. student at the Computer Graphics Laboratory of ETH Zürich. He received a B.Sc. in 2007 from the University of Cyprus and a M.Sc. in 2009 from UC San Diego, both in Computer Science. Marios has published in the areas of Appearance Capture and Fabrication. During his M.Sc. he developed and validated a physically based analytic model for modeling the appearance of rough, highly scattering, thin paper-like materials. During his Ph.D. he developed methods and apparatus for capturing and fabricating the appearance of translucent materials such as milk, wax and soap.

Derek Bradley, *Disney Research, Zürich*
Derek Bradley is a Research Scientist at Disney Research Zurich. He obtained a BCS from Carleton University in 2003, an MCS from Carleton University in 2005, and a PhD from the University of British Columbia in 2010. Derek has been studying photorealistic virtual humans for 5 years, and has published several papers on capturing facial geometry and performance animation of faces with high-level surface details. He has also been involved in the construction of numerous multi-camera multi-illumination capture setups for facial measurements.

Christophe Hery, *Pixar Animation Studios*
Christophe Hery joined Pixar in June 2010, where he holds the position of Senior Scientist. After writing new lighting models and rendering methods for Monsters University and The Blue Umbrella, Christophe is currently working on Finding Dory. An alumnus of Industrial Light & Magic, Christophe previously served as a research and development lead, supporting the facility's shaders and providing rendering guidance. He was first hired by ILM in 1993 as a senior technical director. In 2003, Christophe was the recipient of a Technical Achievement Award for the development of practical methods for rendering skin and other translucent materials using subsurface scattering techniques. In 2010, along with Per Christensen, he received a Scientific and Engineering Award for the development of point-based rendering for indirect illumination and ambient occlusion.

Bernd Bickel, *IST Austria / Disney Research, Zürich*
Bernd Bickel is an Assistant Professor at the Institute of Science and Technology Austria (IST Austria), where he leads the Computer Graphics and Digital Fabrication group. Prior to this appointment he was a Research Scientist at Disney Research Zurich and a Visiting Professor at the Technical University of Berlin. His research involves acquiring, simulating, modeling, and physically realizing the visual appearance and deformation behavior of objects, with applications in animation, biomechanics, material science, and computational design for digital fabrication. He has published several papers on capturing facial appearance, facial geometry, and facial animation. Bernd received a PhD in computer science from ETH Zurich.

Wojciech Jarosz, *Disney Research, Zürich*
Wojciech Jarosz is a Senior Research Scientist in charge of the rendering group at Disney Research Zürich, and an adjunct lecturer at ETH Zürich. He obtained his Ph.D. and M.S. in computer graphics from UC San Diego, and his B.S. in computer science from the University of Illinois, Urbana-Champaign. Wojciech's main research interests lie broadly in the areas of visual appearance and realistic image synthesis, and he has published several papers related to capturing human faces and simulating the complex interaction of light with scattering media like human skin. His work in these areas has been incorporated into production rendering systems, been used in the making of feature films, and earned him the Eurographics Young Researcher Award.

Thabo Beeler, *Disney Research, Zürich*
Thabo Beeler is a Research Scientist at Disney Research Zurich, where he is heading the Capture and Effects group. Prior to that he obtained his PhD from ETH Zurich, for which he was awarded with the Eurographics PhD award. Over the past 5 years, Thabo has been working on digital humans and reflectance acquisition, focusing on high-quality facial geometry reconstruction and dense performance capture. He has published several papers to these topics and has contributed to different feature films that involve digital humans.

Acknowledgments

We thank Steve Marschner for providing the image in Fig. 10.

References

[APS00] ASHIKMIN M., PREMOŽE S., SHIRLEY P.: A microfacet-based BRDF generator. In *Proc. SIGGRAPH* (2000), pp. 65–74. 8

[ARL*10] ALEXANDER O., ROGERS M., LAMBETH W., CHIANG J.-Y., MA W.-C., WANG C.-C., DEBEVEC P.: The digital emily project: Achieving a photorealistic digital actor. *IEEE Computer Graphics and Applications 30*, 4 (July 2010), 20–31. 1

[BBB*10] BEELER T., BICKEL B., BEARDSLEY P., SUMNER B., GROSS M.: High-quality single-shot capture of facial geometry. *ACM Trans. Graph. (Proc. SIGGRAPH) 29*, 4 (2010), 40:1–40:9. 7, 16

[BBN*12] BEELER T., BICKEL B., NORIS G., BEARDSLEY P., MARSCHNER S., SUMNER R. W., GROSS M.: Coupled 3D reconstruction of sparse facial hair and skin. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (2012), 117:1–117:10. 17, 18

[BBN*14] BÉRARD P., BRADLEY D., NITTI M., BEELER T., GROSS M.: High-quality capture of eyes. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 33*, 6 (2014), 223:1–223:12. 21

[BBZG12] BEELER T., BRADLEY D., ZIMMER H., GROSS M.: Improved reconstruction of deforming surfaces by cancelling ambient occlusion. In *European Conference on Computer Vision (ECCV)*. 2012, pp. 30–43. 16

[Bec87] BECKMANN P.: *The scattering of electromagnetic waves from rough surfaces*. 1987. 8

[BHB*11] BEELER T., HAHN F., BRADLEY D., BICKEL B., BEARDSLEY P., GOTSMAN C., SUMNER R. W., GROSS M.: High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (2011), 75:1–75:10. 1

[BHPS10] BRADLEY D., HEIDRICH W., POPA T., SHEFFER A.: High resolution passive facial performance capture. *ACM Trans. Graph. (Proc. SIGGRAPH) 29*, 4 (2010), 41:1–41:10. 1

[BHV*11] BROSTOW G. J., HERNÁNDEZ C., VOGIATZIS G., STENGER B., CIPOLLA R.: Video normals from colored lights. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 33*, 10 (2011), 2104–2114. 15

[BL03] BORSHUKOV G., LEWIS J. P.: Realistic human face rendering for "The Matrix Reloaded". In *ACM SIGGPAPH Sketches* (2003), pp. 1–1. 5

[Bli77] BLINN J. F.: Models of light reflection for computer synthesized pictures. *Computer Graphics (Proc. SIGGRAPH) 11*, 2 (1977), 192–198. 8, 9

[BN12] BRUNETON E., NEYRET F.: A survey of non-linear prefiltering methods for efficient and accurate surface shading. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 18*, 2 (2012), 242–260. 16

[BPL*03] BORSHUKOV G., PIPONI D., LARSEN O., LEWIS J. P., TEMPELAAR-LIETZ C.: Universal capture: Image-based facial animation for "The Matrix Reloaded". In *ACM SIGGPAPH Sketches* (2003), pp. 1–1. 5

[BWP13] BOUAZIZ S., WANG Y., PAULY M.: Online modeling for realtime facial animation. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4 (2013), 40:1–40:10. 21

[CHZ14] CAO C., HOU Q., ZHOU K.: Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Trans. Graph. (Proc. SIGGRAPH) 33*, 4 (2014), 43:1–43:10. 21

[CT82] COOK R. L., TORRANCE K. E.: A reflectance model for computer graphics. *ACM Trans. Graph. 1*, 1 (1982), 7–24. 8

[d'E12] D'EON E.: *A better dipole*. Tech. rep., 2012. 11

[DHI*13] DUPUY J., HEITZ E., IEHL J.-C., POULIN P., NEYRET F., OSTROMOUKHOV V.: Linear efficient antialiased displacement and reflectance mapping. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 32*, 6 (2013), 211:1–211:11. 16

[DHT*00] DEBEVEC P., HAWKINS T., TCHOU C., DUIKER H.-P., SAROKIN W., SAGAR M.: Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH* (2000), pp. 145–156. 2, 3, 4, 5, 7, 9, 20

[dI11] D'EON E., IRVING G.: A quantized-diffusion model for rendering translucent materials. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (2011), 56:1–56:14. 11

[DJ05] DONNER C., JENSEN H. W.: Light diffusion in multilayered translucent materials. *ACM Trans. Graph. (Proc. SIGGRAPH) 24*, 3 (2005), 1032–1039. 10, 11, 12, 14

[DJ06] DONNER C., JENSEN H. W.: A spectral BSSRDF for shading human skin. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2006), pp. 409–417. 9, 13, 14, 19

[DJ07] DONNER C., JENSEN H. W.: Rendering translucent materials using photon diffusion. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2007), pp. 243–251. 11

[dLE07] D'EON E., LUEBKE D., ENDERTON E.: Efficient rendering of human skin. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2007), pp. 147–157. 11, 14

[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. Graph. (Proc. SIGGRAPH) 27*, 5 (2008), 140:1–140:12. 7, 10, 13, 14, 20

[DWT*02] DEBEVEC P., WENGER A., TCHOU C., GARDNER A., WAESE J., HAWKINS T.: A lighting reproduction approach to live-action compositing. *ACM Trans. Graph. (Proc. SIGGRAPH) 21*, 3 (2002), 547–556. 3, 5, 20

[ECJ*06] EINARSSON P., CHABERT C.-F., JONES A., MA W.-C., LAMOND B., HAWKINS T., BOLAS M., SYLWAN S., DEBEVEC P.: Relighting human locomotion with flowed reflectance fields. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2006), pp. 183–194. 3, 6, 20

[EF78] EKMAN P., FRIESEN W.: The facial action coding system: A technique for the measurement of facial movement. In *Consulting Psychologists* (1978). 15

[FBLS05] FUCHS M., BLANZ V., LENSCH H., SEIDEL H.-P.: Reflectance from images: A model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 11*, 3 (2005), 296–305. 2, 8, 9, 21

[FHW*11] FYFFE G., HAWKINS T., WATTS C., MA W.-C., DEBEVEC P.: Comprehensive facial performance capture. *Comp. Graph. Forum (Proc. Eurographics) 30*, 2 (2011), 425–434. 3, 7

[FYD11] FYFFE G., YU X., DEBEVEC P.: Single-shot photometric stereo by spectral multiplexing. In *IEEE Computational Photography (ICCP)* (2011), IEEE, pp. 1–6. 15

[GCP*09] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P.: Estimating specular roughness and anisotropy from second order spherical gradient illumination. *Comp. Graph. Forum (Proc. EG Symposium on Rendering)* (2009), 1161–1170. 3, 8, 9, 21

[GCP*10] GHOSH A., CHEN T., PEERS P., WILSON C. A., DEBEVEC P.: Circularly polarized spherical illumination reflectometry. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 29*, 6 (2010), 162:1–162:12. 8, 9, 18, 21

[Geo03] GEORGHIADES A. S.: Recovering 3-D shape and reflectance from a small number of photographs. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2003), pp. 230–240. 8, 9, 21

[GFT*11] GHOSH A., FYFFE G., TUNWATTANAPONG B., BUSCH J., YU X., DEBEVEC P.: Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 30*, 6 (2011), 129:1–129:10. 7, 18

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. In *Proc. SIGGRAPH* (1996), pp. 43–54. 4

[GHAO10] GHOSH A., HEIDRICH W., ACHUTHA S., O'TOOLE M.: A basis illumination approach to BRDF measurement. *International Journal of Computer Vision (IJCV) 90*, 2 (2010), 183–197. 3

[GHP*08] GHOSH A., HAWKINS T., PEERS P., FREDERIKSEN S., DEBEVEC P.: Practical modeling and acquisition of layered facial reflectance. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 27*, 5 (2008), 139:1–139:10. 3, 7, 8, 9, 10, 11, 12, 21

[GTB*13] GRAHAM P., TUNWATTANAPONG B., BUSCH J., YU X., JONES A., DEBEVEC P., GHOSH A.: Measurement-based synthesis of facial microgeometry. *Comp. Graph. Forum (Proc. Eurographics) 32*, 2 (2013), 335–344. 7, 8, 9, 16, 21

[HCJ13] HABEL R., CHRISTENSEN P. H., JAROSZ W.: Photon beam diffusion: A hybrid Monte Carlo method for subsurface scattering. *Comp. Graph. Forum (Proc. EG Symposium on Rendering) 32*, 4 (2013), 27–37. 11

[HCTW11] HUANG H., CHAI J., TONG X., WU H.-T.: Leveraging motion capture and 3d scanning for high-fidelity facial performance acquisition. *ACM Trans. Graph. (Proc. SIGGRAPH) 30*, 4 (2011), 74:1–74:10. 1

[Her03] HERY C.: Implementing a skin BSSRDF (or several). In *RenderMan, Theory and Practice, ACM SIGGRAPH Courses* (2003). 19

[Her12] HERY C.: *Texture mapping for the Better Dipole model*. Tech. rep., 2012. 19

[HK93] HANRAHAN P., KRUEGER W.: Reflection from layered surfaces due to subsurface scattering. In *Proc. SIGGRAPH* (1993), pp. 165–174. 10

[HS81] HORN B. K., SCHUNCK B. G.: Determining optical flow. *Artificial Intelligence 17* (1981), 185–203. 6

[HTSG91] HE X. D., TORRANCE K. E., SILLION F. X., GREENBERG D. P.: A comprehensive physical model for light reflection. In *Computer Graphics (Proc. SIGGRAPH)* (1991), pp. 175–186. 8

[HWT*04] HAWKINS T., WENGER A., TCHOU C., GARDNER A., GÖRANSSON F., DEBEVEC P.: Animatable facial reflectance fields. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2004), pp. 309–319. 3, 5, 14, 20

[IGAJG15] IGLESIAS-GUITIAN J. A., ALIAGA C., JARABO A., GUTIERREZ D.: A biophysically-based model of the optical properties of skin aging. *Comp. Graph. Forum (Proc. Eurographics)* (2015). 15

[INN07] IGARASHI T., NISHINO K., NAYAR S. K.: The appearance of human skin: A survey. *Foundations and Trends in Computer Graphics and Vision 3*, 1 (2007), 1–95. 2

[IRM*12] IHRKE I., RESHETOUSKI I., MANAKOV A., TEVS A., WAND M., SEIDEL H.-P.: A kaleidoscopic approach to surround geometry and reflectance acquisition. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (2012), pp. 29–36. 3

[JB02] JENSEN H. W., BUHLER J.: A rapid hierarchical rendering technique for translucent materials. *ACM Trans. Graph. (Proc. SIGGRAPH) 21*, 3 (2002), 576–581. 11, 19

[JGB*06] JONES A., GARDNER A., BOLAS M., MCDOWALL I., DEBEVEC P.: Simulating spatially varying lighting on a live performance. In *Conference on Visual Media Production (CVMP)* (2006). 6, 20

[JKK*11] JUNG Y., KUIJPER A., KIPP M., MIKSATKO J., GRATCH J., THALMANN D.: Believable virtual characters in human-computer dialogs. In *Eurographics State of the Art Reports* (2011). 2

[JMLH01] JENSEN H. W., MARSCHNER S. R., LEVOY M., HANRAHAN P.: A practical model for subsurface light transport. In *Proc. SIGGRAPH* (2001), pp. 511–518. 10, 11, 14

[JSB*10] JIMENEZ J., SCULLY T., BARBOSA N., DONNER C., ALVAREZ X., VIEIRA T., MATTS P., ORVALHO V., GUTIERREZ D., WEYRICH T.: A practical appearance model for dynamic facial color. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 29*, 6 (2010), 141:1–141:10. 3, 14, 15, 21

[KB04] KRISHNASWAMY A., BARANOSKI G. V.: A biophysically-based spectral model of light interaction with human skin. *Comp. Graph. Forum (Proc. Eurographics) 23*, 3 (2004), 331–340. 13

[KP03] KOENDERINK J., PONT S.: The secret of velvety skin. *Mach. Vision Appl. 14*, 4 (2003), 260–268. 16, 17

[KSK01] KELEMEN C., SZIRMAY-KALOS L.: A microfacet based coupled specular-matte brdf model with importance sampling. In *Eurographics Short Presentations* (2001), pp. 1–11. 21

[KSS11] KEMELMACHER-SHLIZERMAN I., SEITZ S. M.: Face reconstruction in the wild. In *International Conference on Computer Vision (ICCV)* (2011), ICCV '11, pp. 1746–1753. 21

[Kur11] KURACHI N.: *BSSRDF Models in Motion Picture Production*, 1 ed. 2011, ch. 4.3.6, The Magic of Computer Graphics, pp. 116–119. 16

[LCQ*04] LI Y.-M., CHEN J., QING L.-Y., YIN B.-C., GAO W.: Face detection under variable lighting based on resample by face relighting. In *Conference on Machine Learning and Cybernetics* (Aug 2004), vol. 6, pp. 3775–3780. 1

[LFTG97] LAFORTUNE E. P. F., FOO S.-C., TORRANCE K. E., GREENBERG D. P.: Non-linear approximation of reflectance functions. In *Proc. SIGGRAPH* (1997), pp. 117–126. 9

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. In *Proc. SIGGRAPH* (1996), pp. 31–42. 4

[LPGD09] LAMOND B., PEERS P., GHOSH A., DEBEVEC P.: Image-based separation of diffuse and specular reflections using environmental structured illumination. In *IEEE Computational Photography (ICCP)* (2009). 7, 9

[LYYB13] LI H., YU J., YE Y., BREGLER C.: Realtime facial animation with on-the-fly correctives. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4 (2013), 42:1–42:10. 21

[MGR00] MARSCHNER S. R., GUENTER B. K., RAGHUPATHY S.: Modeling and rendering for realistic facial animation. In *Rendering Techniques (Proc. EG Workshop on Rendering)* (2000), pp. 231–242. 7, 21

[MHP*07] MA W.-C., HAWKINS T., PEERS P., CHABERT C.-F., WEISS M., DEBEVEC P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2007), pp. 183–194. 3, 7, 9, 15, 16, 17, 18

[MJC*08] MA W.-C., JONES A., CHIANG J.-Y., HAWKINS T., FREDERIKSEN S., PEERS P., VUKOVIC M., OUHYOUNG M., DEBEVEC P.: Facial performance synthesis using deformation-driven polynomial displacement maps. *ACM Trans. Graph. (Proc. SIGGRAPH Asia) 27*, 5 (2008), 121:1–121:10. 3

[MPN*02] MATUSIK W., PFISTER H., NGAN A., BEARDSLEY P., ZIEGLER R., MCMILLAN L.: Image-based 3D photography using opacity hulls. *ACM Trans. Graph. (Proc. SIGGRAPH) 21*, 3 (2002), 427–437. 2, 3

[MWL*99] MARSCHNER S. R., WESTIN S. H., LAFORTUNE E. P. F., TORRANCE K. E., GREENBERG D. P.: Image-based BRDF measurement including human skin. In *Rendering Techniques (Proc. EG Workshop on Rendering)* (1999), pp. 131–144. 2, 8, 9, 21

[NDM05] NGAN A., DURAND F., MATUSIK W.: Experimental analysis of BRDF models. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2005), pp. 117–126. 8, 9

[NFB93] NAYAR S., FANG X., BOULT T.: Removal of Specularities using Color and Polarization. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (Jun 1993), pp. 583–590. 7

[NKGR06] NAYAR S. K., KRISHNAN G., GROSSBERG M. D., RASKAR R.: Fast separation of direct and global components of a scene using high frequency illumination. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3 (2006), 935–944. 3, 8, 12

[NSD95] NIMEROFF J. S., SIMONCELLI E., DORSEY J.: Efficient re-rendering of naturally illuminated environments. In *Photorealistic Rendering Techniques*. 1995, pp. 373–388. 4

[OB10] OLANO M., BAKER D.: Lean mapping. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (i3D)* (2010), pp. 181–188. 16

[ORK12] O'TOOLE M., RASKAR R., KUTULAKOS K. N.: Primal-dual coding to probe light transport. *ACM Trans. Graph. (Proc. SIGGRAPH) 31*, 4 (2012), 39:1–39:11. 8

[PCK*08] PARIS S., CHANG W., KOZHUSHNYAN O. I., JAROSZ W., MATUSIK W., ZWICKER M., DURAND F.: Hair photobooth: Geometric and photometric acquisition of real hairstyles. *ACM Trans. Graph. (Proc. SIGGRAPH) 27*, 3 (Aug. 2008), 30:1–30:9. 21

[PTMD07] PEERS P., TAMURA N., MATUSIK W., DEBEVEC P.: Post-production facial performance relighting using reflectance transfer. *ACM Trans. Graph. (Proc. SIGGRAPH) 26*, 3 (2007). 19

[RBS03] ROBERTSON M. A., BORMAN S., STEVENSON R. L.: Estimation-theoretic approach to dynamic range enhancement using multiple exposures. *Journal of Electronic Imaging 12*, 2 (2003), 219–228. 18

[RH01] RAMAMOORTHI R., HANRAHAN P.: An efficient representation for irradiance environment maps. In *Proc. SIGGRAPH* (2001), pp. 497–500. 7

[RRS99] RIKLIN-RAVIV T., SHASHUA A.: The quotient image: Class based recognition and synthesis under varying illumination conditions. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (1999), vol. 2, pp. 566–571. 19

[Sey13] SEYMOUR M.: The art of digital faces at ICT - digital emily to digital ira, Nov. 2013. [Online; posted 25-November-2013: www.fxguide.com/featured/the-art-of-digital-faces-at-ict-from-digital-emily-to-digital-ira/]. 1

[SPF*13] SCHERBAUM K., PETTERSON J., FERIS R. S., BLANZ V., SEIDEL H.-P.: Fast face detector training using tailored views. In *International Conference on Computer Vision (ICCV)* (2013), pp. 2848–2855. 1

[SPN*14] SCHMIDT T.-W., PELLACINI F., NOWROUZEZAHRAI D., JAROSZ W., DACHSBACHER C.: State of the art in artistic editing of appearance, lighting, and material. In *Eurographics State of the Art Reports* (2014). 20

[SRH*11] SCHERBAUM K., RITSCHEL T., HULLIN M., THORMÄHLEN T., BLANZ V., SEIDEL H.-P.: Computer-suggested facial makeup. *Comp. Graph. Forum (Proc. Eurographics) 30*, 2 (2011), 485–492. 1

[SSWK13] SCHWARTZ C., SARLETTE R., WEINMANN M., KLEIN R.: Dome ii: A parallelized btf acquisition system. In *Eurographics Workshop on Material Appearance Modeling: Issues and Acquisition* (2013), pp. 25–31. 6

[TFG*13] TUNWATTANAPONG B., FYFFE G., GRAHAM P., BUSCH J., YU X., GHOSH A., DEBEVEC P.: Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph. (Proc. SIGGRAPH) 32*, 4 (2013), 109:1–109:12. 3, 7, 8, 18

[TGD11] TUNWATTANAPONG B., GHOSH A., DEBEVEC P.: Practical image-based relighting and editing with spherical-harmonics and local lights. In *Conference on Visual Media Production (CVMP)* (2011), pp. 138–147. 3, 5, 19, 20

[TGL*06] TARIQ S., GARDNER A., LLAMAS I., JONES A., DEBEVEC P., TURK G.: Efficient estimation of spatially varying subsurface scattering parameters. In *11th Workshop on Vision, Modeling and Visualization (VMV)* (2006). 3, 7, 11, 12, 21

[TOS*03] TSUMURA N., OJIMA N., SATO K., SHIRAISHI M., SHIMIZU H., NABESHIMA H., AKAZAKI S., HORI K., MIYAKE Y.: Image-based skin color and texture analysis/synthesis by extracting hemoglobin and melanin information in the skin. *ACM Trans. Graph. (Proc. SIGGRAPH) 22*, 3 (2003), 770–779. 1

[TS67] TORRANCE K. E., SPARROW E. M.: Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America 57*, 9 (1967), 1105–1112. 8, 9, 21

[Tuc00] TUCHIN V.: *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*. Spie Press Series. 2000. 10

[vdPJD*14] VON DER PAHLEN J., JIMENEZ J., DANVOYE E., DEBEVEC P., FYFFE G., ALEXANDER O.: Digital Ira and beyond: Creating real-time photoreal digital actors. In *ACM SIGGRAPH Courses* (2014), pp. 1:1–1:384. 16

[VH12] VOGIATZIS G., HERNÁNDEZ C.: Self-calibrated, multi-spectral photometric stereo for 3D face capture. *International Journal of Computer Vision (ICJV) 97*, 1 (2012), 91–103. 15

[War92] WARD G. J.: Measuring and modeling anisotropic reflection. *Computer Graphics (Proc. SIGGRAPH) 26*, 2 (1992), 265–272. 8, 21

[WBK*07] WARD K., BERTAILS F., KIM T.-Y., MARSCHNER S., CANI M.-P., LIN M.: A survey on hair modeling: Styling, simulation, and rendering. *IEEE Transactions on Visualization and Computer Graphics (TVCG) 13*, 2 (2007), 213–234. 2

[WGP*10] WILSON C. A., GHOSH A., PEERS P., CHIANG J.-Y., BUSCH J., DEBEVEC P.: Temporal upsampling of performance geometry using photometric alignment. *ACM Trans. Graph. 29*, 2 (2010), 17:1–17:11. 7, 15

[WGT*05] WENGER A., GARDNER A., TCHOU C., UNGER J., HAWKINS T., DEBEVEC P.: Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph. (Proc. SIGGRAPH) 24*, 3 (2005), 756–764. 3, 5, 19, 20

[WLL*09] WEYRICH T., LAWRENCE J., LENSCH H. P. A., RUSINKIEWICZ S., ZICKLER T.: Principles of appearance acquisition and representation. *Foundations and Trends in Computer Graphics and Vision 4*, 2 (2009), 75–191. 2

[WLVGP09] WEISE T., LI H., VAN GOOL L., PAULY M.: Face/off: Live facial puppetry. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2009), pp. 7–16. 3

[WMLT07] WALTER B., MARSCHNER S. R., LI H., TORRANCE K. E.: Microfacet models for refraction through rough surfaces. In *Rendering Techniques (Proc. EG Symposium on Rendering)* (2007), pp. 195–206. 8

[WMP*05] WEYRICH T., MATUSIK W., PFISTER H., NGAN A., GROSS M.: *Measuring skin reflectance and subsurface scattering*. Tech. rep., Mitsubishi Electric Research Laboratories (MERL), 2005. 19

[WMP*06] WEYRICH T., MATUSIK W., PFISTER H., BICKEL B., DONNER C., TU C., MCANDLESS J., LEE J., NGAN A., JENSEN H. W., GROSS M.: Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph. (Proc. SIGGRAPH) 25*, 3 (2006), 1013–1024. 3, 7, 8, 9, 11, 14, 15, 19, 21

[Woo80] WOODHAM R. J.: Photometric method for determining surface orientation from multiple images. *Optical engineering 19*, 1 (1980), 139–144. 15

[WWMT11] WU C., WILBURN B., MATSUSHITA Y., THEOBALT C.: High-quality shape from multi-view stereo and shading under general illumination. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (2011), IEEE, pp. 969–976. 16

[ZGP*13] ZHU Y., GARIGIPATI P., PEERS P., DEBEVEC P., GHOSH A.: Estimating diffusion parameters from polarized spherical-gradient illumination. *IEEE Computer Graphics and Applications 33*, 3 (2013), 34–43. 11, 12, 21

[ZSCS04] ZHANG L., SNAVELY N., CURLESS B., SEITZ S. M.: Spacetime faces: High resolution capture for modeling and animation. *ACM Trans. Graph. (Proc. SIGGRAPH) 23*, 3 (2004), 548–558. 1, 3