Practical Dynamic Facial Appearance Modeling and Acquisition

PAULO GOTARDO, Disney Research
JÉRÉMY RIVIERE, Disney Research
DEREK BRADLEY, Disney Research
ABHIJEET GHOSH, Imperial College London
THABO BEELER, Disney Research

Fig. 1. We present an inverse rendering approach to capture dynamic appearance properties of human skin, including per-frame albedo, high-resolution normals and specular intensity, at high fidelity from a purely passive multi-camera setup.

We present a method to acquire dynamic properties of facial skin appearance, including dynamic diffuse albedo encoding blood flow, dynamic specular intensity, and per-frame high resolution normal maps for a facial performance sequence. The method reconstructs these maps from a purely passive multi-camera setup, without the need for polarization or requiring temporally multiplexed illumination. Hence, it is very well suited for integration with existing passive systems for facial performance capture. To solve this seemingly underconstrained problem, we demonstrate that albedo dynamics during a facial performance can be modeled as a combination of: (1) a static, high-resolution base albedo map, modeling full skin pigmentation; and (2) a dynamic, one-dimensional component in the CIE $L^*a^*b^*$ color space, which explains changes in hemoglobin concentration due to blood flow. We leverage this albedo subspace and additional constraints on appearance and surface geometry to also estimate specular reflection parameters and resolve high-resolution normal maps with unprecedented detail in a passive capture system. These constraints are built into an inverse rendering framework that minimizes the difference of the rendered face to the captured images, incorporating constraints from multiple views for every texel on the face. The presented method is the first system capable of capturing high-quality dynamic appearance maps at full resolution and video framerates, providing a major step forward in the area of facial appearance acquisition.

CCS Concepts: • **Computing methodologies → Reflectance modeling**; **3D imaging**; *Appearance and texture representations*;

Additional Key Words and Phrases: Inverse Rendering, Passive Capture, Dynamic Appearance, Albedo Subspace, Surface Detail

ACM Reference Format:
Paulo Gotardo, Jérémy Riviere, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2018. Practical Dynamic Facial Appearance Modeling and Acquisition. *ACM Trans. Graph.* 37, 6, Article 232 (November 2018), 13 pages. https://doi.org/10.1145/3272127.3275073

Authors' addresses: Paulo Gotardo, Disney Research, paulo.gotardo@disneyresearch.com; Jérémy Riviere, Disney Research, jeremy.riviere@disneyresearch.com; Derek Bradley, Disney Research, derek.bradley@disneyresearch.com; Abhijeet Ghosh, Imperial College London, ghosh@imperial.ac.uk; Thabo Beeler, Disney Research, thabo.beeler@disneyresearch.com.

0730-0301/2018/11-ART232 $15.00
https://doi.org/10.1145/3272127.3275073

1 INTRODUCTION

Creating and rendering realistic humans is becoming ever more important in computer graphics, with applications ranging from visual effects for entertainment, to educational and training scenarios, and even medical use cases. Digital humans pose a formidable research challenge since their virtual appearance is comprised of many different components including shape, motion and material properties. In order to create a compelling and believable overall virtual character each of these components must be modeled realistically.

One of the most important challenges is to faithfully reproduce the way light interacts with skin, which we here refer to as appearance modeling. In the past years the field has made substantial progress in skin rendering, yielding impressive results even in real-time [d'Eon et al. 2007; Jimenez et al. 2009; von der Pahlen et al. 2014]. These methods rely on different parameters to simulate skin appearance, including surface geometry and reflectance properties such as albedo and specular roughness.

Early work has modeled skin appearance using static parameters for diffuse and specular reflectance, as described by a bidirectional reflectance distribution function (BRDF) [Marschner et al. 1999]. General BRDF acquisition from human subjects is extremely challenging since the space is high-dimensional and hence would require dense sampling of incoming and outgoing light rays at every point on the surface, even without considering changes in surface and physiological state. To avoid such a daunting task, considerable

effort has been dedicated to the investigation of skin BRDF models [Ghosh et al. 2008; Weyrich et al. 2006]. However, robustly fitting the parameters of these models to real skin observations still requires a large sample set of incoming/outgoing rays. Such sampling is usually achieved by temporally multiplexing different illumination conditions [Debevec et al. 2000; Wenger et al. 2005; Weyrich et al. 2006], which limits the approach when capturing non-static subjects and leads to complex acquisition setups.

Static appearance capture falls short for creating convincing animations because, as skin shape changes over time, appearance does not remain static. Appearance also changes dynamically as a result of various factors, including changes in blood flow and in skin microstructure. To mitigate this limitation, multiple albedo maps can be employed and blended during the animation [Alexander et al. 2010]. This is a convenient representation as it naturally integrates with traditional blendshape animation, where the surface is also created as a linear combination of a set of base shapes. While this method has enjoyed much popularity, it is becoming more and more obvious that linear blendshapes are not sufficient to faithfully represent the non-linear behaviour of real skin deformation [Ma et al. 2008]. The same holds also for appearance, which is subject to even more non-linearities as blood flow, for example, exhibits hysteresis and is also influenced by physiological effects caused by heat or excitement [Jimenez et al. 2010]. Consequently, just as researchers explore alternatives to linear blendshape models for representing the full complexity of dynamic skin deformation [Beeler et al. 2011; Bradley et al. 2010; Wu et al. 2016], better models for dynamic skin appearance are also needed.

We present a first comprehensive model for dynamic skin appearance, which couples dynamic reflectance parameters for skin (albedo and specular reflectance) with dynamic geometry. It provides a compact time-varying model of appearance and surface detail that can be estimated from multiview image streams without requiring time-multiplexed illumination. However, given our passive acquisition setup, we enforce a few constraints on the appearance estimation: we model the time-varying diffuse reflectance purely as a change in albedo and do not estimate any subsurface scattering parameters (e.g., translucency). We also do not explicitly model any anisotropic skin reflectance changes caused by deformation at the mesoscopic level, but instead model anisotropy in our per-frame high-resolution normal and specular intensity maps to achieve a temporally and spatially varying isotropic BRDF.

Key to our dynamic albedo model is the observation that time-varying skin albedo samples lie near a compact, one-dimensional linear subspace of the standard CIE $L^*a^*b^*$ 1976 color space (denoted Lab, for simplicity). We demonstrate this relation empirically by measuring time-varying albedo of different skin tones undergoing changes in facial expression and blood flow. As a result, our 1-D albedo subspace is restricted to explain changes in appearance that are predominately due to varying skin pigmentation (*e.g.*, concentration of hemoglobin), allowing to separate shading changes from albedo variation. This fact not only removes ambiguities in albedo estimation but also provides sufficient constraints to recover dynamic surface geometry (normal field) and specular reflectance without the need for time-multiplexed illumination.

In addition to introducing this dynamic appearance model, we present a method to fit our model to performance capture data of real subjects. Most notably, our approach requires only a multi-view video camera setup with static illumination. No temporal multiplexing of lighting patterns are required, making our method highly suitable for integration with traditional facial performance capture setups. The skin reflectance maps presented in this paper were computed from just four color cameras. This advantage alleviates the current need for separate scanning sessions that are required to recover both facial motion and appearance. Furthermore, since we go beyond static appearance capture, the dynamic performances contain unprecedented per-frame skin reflectance parameters modeling effects such as hysteresis in blood flow.

We demonstrate our dynamic appearance model and fitting strategy by reconstructing a number of performances given by several different subjects with varying skin tones. The recovered time-varying geometry and appearance maps are directly suitable for relighting in applications such as visual effects, VR/AR simulations, or telepresence.

2 RELATED WORK

While there is significant literature on general reflectance modeling, we will restrict the discussion here specifcally to facial reflectance capture and modeling. We refer the interested reader to two recent surveys on the topic [Klehm et al. 2015; Weyrich et al. 2009]. In the following, we discuss facial reflectance capture and modeling in computer graphics in the context of both static and dynamic facial appearance, using both active and passive capture setups.

Active illumination, static facial appearance: the work by Debevec et al. [2000] first proposed employing a specialized light stage setup to acquire a dense reflectance field of a human face for photo-realistic image-based relighting applications. They also employed the acquired data to estimate a few view-dependent reflectance maps that could be interpolated for viewpoint animation. Fuchs et al. [2005] employed a smaller number of photographs and lighting directions, at the cost of sacrificing continuously-varying specular reflectance. Weyrich et al. [2006] employed an LED sphere and 16 cameras to densely record facial reflectance and computed view-independent estimates of facial reflectance from the acquired data including per-pixel diffuse and specular albedos, and per-region specular roughness parameters. They also employed a specialized skin contact probe to estimate parameters of subsurface scattering based on dipole diffusion [Jensen et al. 2001]. Subsequently, Ma et al. [2007] introduced polarized spherical gradient illumination (using an LED sphere) for efficient acquisition of the separated diffuse and specular albedos and photometric normals of a face using just eight photographs, and demonstrated high quality facial geometry including skin mesostructure as well as realistic rendering with the acquired data. Ghosh et al. [2008] further extended the acquisition method to acquire layered facial reflectance using a combination of polarization and structured lighting. Similar to Weyrich et al., they estimated a per-region specular BRDF, but further include single scattering and a data-driven two-layered subsurface scattering (modeled with multipole diffusion [Donner and Jensen 2005]) in their reflectance model. Later, Ghosh et al. [2011] extended the

view-dependent solution of Ma et al. [2007] for multi-view facial acquisition with polarized spherical gradient illumination. Graham et al. [2013] have proposed augmenting meso-scale facial geometry with micro-geometry of facial skin patches acquired using a combination of macro-photography and polarized spherical gradient illumination. They employ constrained texture synthesis to add microscale details to underlying skin meso-structure and also fit micro-scale skin BRDF for increased realism of skin rendering. More recently, Fyffe et al. [2016] have proposed an alternate solution for static facial capture that employs commodity hardware. Their setup achieves near-instant capture of facial geometry and reflectance using a combination of multiple cameras and multiple flashes that are triggered in sequence within a few milliseconds. However, their approach does not extend to dynamic facial appearance capture.

Active illumination, dynamic facial appearance: work by Hawkins et al. [2004] extended the approach of Debevec et al. [2000] to acquire dynamic facial reflectance fields of a set of key poses. They then interpolated between the reflectance fields of these key poses at run-time for synthesizing relightable facial animations. Wenger et al. [2005] employed an LED sphere and high speed photography to aquire the response to a dense set of illumination conditions in order to relight each frame of a target facial performance. They also proposed employing the data to estimate photometric surface normals and diffuse and specular albedos for a reflectance-model based relighting of the facial performance. Ma et al. [2008] instead employed spherical gradient illumination in conjunction with high speed acquisition to capture short sequences of facial performances (from neutral to various expressions). They then employed the acquired facial displacement maps (in conjunction with marker-based correspondences) to fit polynomial functions over the space of facial expressions as a way of encoding changes in facial mesostructure during a performance. Fyffe et al. [2011] instead applied the complementary spherical gradient illumination based alignment of Wilson et al. [2010] in conjunction with high speed photography to acquire longer facial performance sequences. They further applied a heuristics based diffuse-specular separation on the acquired data to obtain albedo and normal maps for high quality rendering of the acquired facial performance. More recently, Fyffe & Debevec [2015] have proposed employing spectral multiplexing with polarized spherical gradient illumination (using an RGB LED sphere) for facial performance capture at regular video rates. This however requires a complicated setup with multiple cameras per acquisition viewpoint. Gotardo et al. [2015] propose a simpler binocular setup with spectral and temporal multiplexing of nine light sources to compute dynamic albedo and normal maps, but only diffuse reflectance is modeled. Finally, Nagano et al. [2015] have acquired microgeometry of various skin patches under stretch and compression (using polarized spherical gradient illumination) and employed the acquired data for building an efficient real-time rendering technique for dynamic facial microgeometry using texture space filtering of the neutral displacement map. Our work is related to this but we estimate skin surface geometry changes due to stretching and compression at the scale of mesostructure. Furthermore, compared to above related works, we rely purely on passive acquisition for such analysis.

Passive acquisition: In order to overcome the requirements of specialized acquisition setups, researchers have also investigated approaches for passive facial acquisition. Such acquisition is particularly well suited for facial performance capture since active approaches usually require time-multiplexed illumination, imposing requirements of high frame rate acquisition and synchronization. A popular approach has been to employ uniform constant illumination for multi-view facial capture [Beeler et al. 2010; Bradley et al. 2010]. Such an approach enables estimation of an albedo texture under flat lit illumination for rendering purposes besides facial geometry reconstruction based on multi-view stereo. Beeler et al. [2010] further proposed augmenting the reconstructed facial geometry with mesostructure detail extracted from the albedo texture using a high-pass filter ("dark is deep" assumption). The approach was later extended for reconstructing facial performances with drift-free tracking over long sequences using anchor frames [Beeler et al. 2011]. While producing very good qualitative results for facial geometry and a uniformly lit texture for rendering, the estimated albedo is not completely diffuse and contains a small amount of specular reflectance baked into the texture. Furthermore, the approach has thus far not enabled estimation of detailed specular reflectance parameters over the facial surface which we target in this work. Researchers have also extended passive facial geometry and performance capture to simple binocular [Valgaerts et al. 2012] and even monocular setups [Cao et al. 2015; Garrido et al. 2013; Ichim et al. 2015; Shi et al. 2014] under uniform, uncontrolled illumination settings including indoor and outdoor environments. These methods often assume that skin reflectance is Lambertian and constant over time, with lighting estimation limited to low-frequency spherical harmonics. In addition, they strongly rely on facial geometry priors (e.g. blendshape models) and, although shading-based geometry refinement reveals facial wrinkles at larger scales, they cannot resolve the same level of fine detail achieved with our novel approach. Also related to our work is the appearance model of Jimenez et al. [2010], which is used to drive a spectral skin BSSRDF [Donner et al. 2008] for rendering faces with time-varying skin color. They employ passive acquisition to estimate hemoglobin concentration maps for various facial expressions; these maps provide an N-dimensional linear model of hemoglobin variation (due to blood flow) over the face caused by expressions as well as physiological or emotional changes. The exact value for N is not given and can be large, making it difficult to capture the model. In contrast, we model albedo dynamics using a single, one-dimensional basis in Lab space, which is captured with the simple protocol outlined in section 3.2. The compactness of our model is key in making it possible to resolve per-frame albedo, specular intensity, and high-detail normal maps without requiring multiplexed illumination. Furthermore, we do not restrict blood flow to be piecewise linear over time and instead model its full dynamics (including hysteresis effects) over each frame of facial performance. Fyffe et al. [2014] have employed a database of acquired high quality facial scans (using the method of [Ghosh et al. 2011]) to augment a monocular video sequence of a facial performance acquired under passive illumination with high resolution facial geometry and reflectance maps for realistic rendering. The approach achieves impressive qualitative results but requires the existence of a dense set of facial scans with reflectance information of the target subject.

Saito et al. [2017] have proposed a deep learning approach for data-driven inference of high resolution facial texture map of an entire face for realistic rendering from an input of a single low resolution face image with partial facial coverage. This has been recently extended to inference of facial mesostructure given a diffuse albedo texture [Huynh et al. 2018] , and even complete facial reflectance and displacement maps besides albedo texture given partial facial image as input [Yamaguchi et al. 2018]. These approaches focus on easily creating a believable digital avatar rather than accurate reconstruction of facial appearance and rely on a facial database acquired using polarized spherical gradients for training. In this work, we aim to estimate detailed facial appearance information including time varying changes in diffuse albedo and changes in specular reflectance and mesostructure due to skin deformation using a typical passive facial capture setup, without requiring to borrow any information from a database. Unlike previous work, we also target truly dynamic appearance modeling at the temporal resolution of every acquired frame of a facial performance.

3 DYNAMIC APPEARANCE MODEL

Skin appearance does not remain constant over time, but changes at several time-scales. In this section, we explain how we model the time-varying effects of skin appearance such that it can be estimated from our captured data. We start by reviewing the skin reflectance model and subsequently introduce our dynamic appearance model.

3.1 Skin Reflectance Model

In this work, we model skin as a two-layer material composed of a rough dielectric layer, the stratum corneum, which accounts for reflection at the surface of the skin, and a diffuse layer that accounts for body reflection. Following previous work [Weyrich et al. 2006], we model the stratum corneum with the microfacet BRDF model [Cook and Torrance 1981]

$$f_s(\omega_o, \omega_i) = \varrho \frac{D(\omega_o, \omega_i, \mathbf{n}, \alpha)\, G(\omega_o, \omega_i)\, F(\eta, \mathbf{n}, \omega_i)}{4\,|\,\langle \mathbf{n}, \omega_i \rangle\, \langle \mathbf{n}, \omega_o \rangle\,|}, \qquad (1)$$

where D is the distribution term, which we model using a Blinn-Phong lobe with exponent α, G is the standard geometric masking/shadowing term, and F is the Fresnel term, which we model using Schlick's approximation [Schlick 1994]. The specular intensity ϱ controls how strongly the incoming light is reflected at this location, and is influenced by properties such as oiliness or specular ambient occlusion. To make dynamic capture well-constrained, we assume a known index of refraction η for skin and specular lobe α as measured in [Weyrich et al. 2006].

We model the body reflection as a simple diffuse Lambertian lobe

$$f_d(\omega_o, \omega_i) = \psi \frac{\boldsymbol{\rho}}{\pi}, \qquad (2)$$

where $\boldsymbol{\rho}$ is the RGB albedo color. An additional scalar parameter ψ is introduced to capture residual diffuse ambient occlusion in locations where the initial base mesh does not capture fine geometric detail, for example in wrinkle folds (data capture is described in Sec. 5). We employ this simple model for the body reflection instead of a more sophisticated subsurface scattering model (e.g., [Donner and Jensen 2006]) for ease of model-fitting from the acquired data.

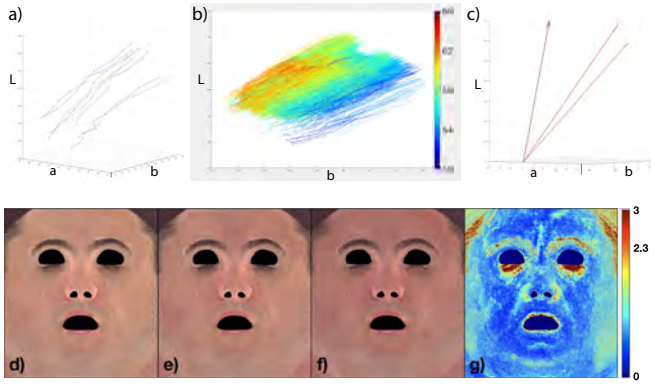

Fig. 2. **Albedo Subspace** – (a) We found that time-varying albedo lies near a straight line in Lab space for the amount of blood flow typically observed during performances. (b) When looking at the lines for different locations over a person's face, we see how they follow the same general direction with only a slight variation in angular slope (line color). (c) Directions vary considerably more when looking across different skin types (line colors). (d)-(f) Sequence of real albedo maps showing face reddening due to blood flow. (g) Quality of our albedo subspace approximation using a standard perceptual metric, CIE ΔE_{2000}, averaged over 10 frames within sequence (d)-(f); $\Delta E_{2000} \leq 3.0$ corresponds to mostly imperceptible differences.

Following the dichromatic reflection model [Shafer 1985], our full appearance model is expressed as the sum of Eq. 1 and Eq. 2,

$$f_r(\omega_o, \omega_i) = f_d(\omega_o, \omega_i) + f_s(\omega_o, \omega_i)\,. \qquad (3)$$

3.2 Dynamic Albedo

Skin albedo is mainly the result of underlying concentrations of melanin and hemoglobin [Anderson and Parrish 1981]. In this work, we assume that albedo changes are only caused by varying hemoglobin concentration due to blood flow, which is a reasonable assumption at the time-scales we are concerned with. When modeling longer time-scales, one might also have to take into account changes in melanin concentrations, for example, due to tanning. The blood concentration in skin may change either due to physiological effects, such as blushing, or physical effects such as muscular activity that actively presses blood out of one part of the skin and into another [Jimenez et al. 2010]. We model this variation in albedo due to blood flow using a compact subspace which we analyze in the following.

Albedo Subspace. Chardon et al. [1991] show that skin albedo with a given melanin concentration projects onto a single line in the Lb plane of the Lab color space. They quantify this line by its angle with the b axis, called the typology angle of skin. Inspired by this work, we analyzed the time-varying component of skin albedo and found that it resides near a line $\boldsymbol{v}$ in Lab space (Fig. 2) as blood flow is observed during facial performance. The albedo values for this subspace analysis were obtained in a separate capture process using cross-polarization to isolate the pure diffuse reflectance.

Thus, for a given skin patch (texel), our subspace models the albedo $\boldsymbol{\rho}_f$ at any point in time (frame) f as a combination of a

base albedo $\boldsymbol{\rho}_0$ in Lab space plus a scalar h_f describing blood-flow-induced change in hemoglobin concentration,

$$\boldsymbol{\rho}_f = \mathcal{T}_{\mathsf{Lab}} \left(\boldsymbol{\rho}_0 + h_f \boldsymbol{\upsilon} \right) , \tag{4}$$

where $\mathcal{T}_{\mathsf{Lab}}$ denotes the transformation from Lab to RGB space. In fact, we define our albedo subspace as a line segment centered at the base albedo, since we expect to observe only a limited amount of blood flow during performance capture. This constraint is enforced during model fitting by penalizing the magnitude of h_f (deviation from the base albedo). In addition, we further constrain the change in hemoglobin concentration h_f to be spatially smooth, while allowing the base albedo to model the full skin pigmentation and spatial detail.

The albedo line direction $\boldsymbol{\upsilon}$ varies considerably among people as a function of their skin typology (Fig. 2 (c)). However, we found that variation over a person's face was limited to ±6 degrees (Fig. 2 (b)). This finding further constrains our model and facilitates capturing $\boldsymbol{\upsilon}$: once its effect is observed on a small face area, its estimate can be applied over the whole face (Sec. 5.5). To validate this claim, we evaluate our blood flow subspace in approximating a 10-frame sequence of albedo maps showing pronounced face reddening, Fig. 2(d)-(f), corresponding to the data in Fig. 2(b). A common line direction was used for all texels. We measure appromixation error using a standard, perceptually-motivated color difference metric, CIE ΔE_{2000}, between original and approximated albedos; $\Delta E_{2000} = 2.3$ corresponds to a Just Noticeable Difference (JND) [Sharma and Bala 2002]. Figure 2(g) shows that the errors of our model range from imperceptible to minimally perceptible.

A key result of our albedo subspace model is that base albedo $\boldsymbol{\rho}_0$ and its hemoglobin direction $\boldsymbol{\upsilon}$ can be pre-acquired (and fixed) using a simple protocol. Then, dynamic albedo capture only requires the estimation of a single degree of freedom h_f per texel and per frame. By constraining the dynamic albedo in this way, our model makes it tractable to estimate dynamic, non-Lambertian BRDF parameters and resolve high-resolution per-frame surface normal without requiring active, polarized illumination as demonstrated next.

4 DYNAMIC APPEARANCE ESTIMATION

This section describes how we solve for the per-frame parameter vector $\boldsymbol{\Theta}_f = [\boldsymbol{\rho}_0, \boldsymbol{\upsilon}, h_f, \psi_f, \varrho_f, \mathbf{n}_f]$ in our dynamic appearance model introduced in Section 3. Here we assume that camera and lighting calibration, and 3D face mesh tracking have been performed *a priori*, as detailed in Section 5. We also assume that the hemoglobin direction $\boldsymbol{\upsilon}$ has been captured from a small face area, using a separate capture protocol detailed in Section 5.

4.1 Inverse Rendering

At the core, our inverse rendering pipeline estimates optimal parameters by minimizing the residual between a synthesized pixel and its captured color $\mathbf{c}_{f\omega_o}$ in the camera views $\omega_o \in \mathcal{V}_f$ where it is visible. We model incident illumination as a set of directional light rays ω_i that are uniformly sampled over the incident sphere (Ω) and present constant illumination color $\mathbf{c}_{\omega_i}$ and uniform solid angle $\Delta\omega = \frac{4\pi}{|\Omega|}$. For each texel, we denote the set of unoccluded lights at that texel location as $\mathcal{L}_f$. Using Eqs. 1–4, our rendering loss is formulated for each frame and texel as

$$E_f(\boldsymbol{\Theta}_f) = \sum_{\omega_o \in \mathcal{V}_f} w_{f\omega_o} \left\| \mathbf{c}_{f\omega_o} - \sum_{\omega_i \in \mathcal{L}_f} f_r(\omega_o, \omega_i, \boldsymbol{\Theta}_f)(\mathbf{n}_f^T \omega_i) \mathbf{c}_{\omega_i} \Delta\omega \right\|_2^2 . \tag{5}$$

Here, $w_{f\omega_o}$ is a precomputed per-camera weight that encodes how reliable the observation $\mathbf{c}_{f\omega_o}$ is, based on factors such as focus, motion blur and view foreshortening.

For efficiency purposes, we operate entirely in the texture space of the tracked 3D face mesh, which facilitates pooling data across views and, when necessary, also across time. All input data is converted into texture domain and visibility information is precomputed and stored in the input texture maps (Figs. 3 and 4). For each frame, we also precompute self-shadowing maps given the light rays and 3D face geometry. The final output of the method is a per-frame, multichannel parameter map with per-texel vectors $\boldsymbol{\Theta}_f$ (Fig. 6).

To estimate this parameter map, we implemented our model in Eq. 5 as an auto-differentiable renderer using Ceres Solver [Agarwal et al. 2016]. To navigate around local minima and improve robustness, we optimize using block coordinate descent and compute the solution in three main steps. In each step we optimize a different subset of the parameters $\boldsymbol{\Theta}_f$, with different constraints, as detailed next. Albedo is first computed in RGB space, given the other fixed parameters, then projected onto its precomputed Lab subspace. By working in Lab space, this projection step minimizes a perceptually more meaningful error metric.

4.2 Tangent Space Normal Paramaterization

We represent each normal $\mathbf{n}_f = \mathbf{R}_f \mathbf{n}_t$ in terms of its corresponding tangent space normal $\mathbf{n}_t$. The tangent space of the texel is given by the (known) 3D rotation $\mathbf{R}_f = [\ \mathbf{t}_f\ \mathbf{b}_f\ (\mathbf{t}_f \times \mathbf{b}_f)\ ]$, where $\mathbf{t}_f$ and $\mathbf{b}_f$ are the unit tangent and bitangent directions precomputed from the tracked 3D face mesh at frame f and texel (u, v). Considering all texels, this tangent space normal field is parameterized using a height surface map $z(u, v)$, which presents integrability as a hard constraint and only a single degree of freedom per texel (instead of 2), making normal estimation better constrained [Onn and Bruckstein 1990]. A tangent space normal $\mathbf{n}_t$ is encoded by the partial derivatives (forward differences) z_u and z_v of z,

$$\mathbf{n}_t(u, v) = \begin{bmatrix} -z_u(u, v) \\ -z_v(u, v) \\ 1 \end{bmatrix} (1 + z_u(u, v)^2 + z_v(u, v)^2)^{-\frac{1}{2}} . \tag{6}$$

Initializing $z(u, v) = 0, \forall u, v$, corresponds to initializing all $\mathbf{n}_f$ to the normals of the base mesh at frame f. Note that $\mathbf{n}_f$ does not depend on the absolute values in z, only on its derivatives. We therefore constrain z to remain near 0 by penalizing its magnitude squared. This parameterization based on derivatives of z couples the solutions of all texels; however, these solutions are easily parallelized via an iterative, alternated optimization strategy on a Red-Black texel grid.

4.3 Step 0 – Base Albedo ρ_0 and Specular Intensity ϱ_0

This first optimization step can be considered a calibration step and is required only once per actor. Given the pre-acquired hemoglobin direction $\boldsymbol{\upsilon}$ (Section 5.5), our main goal now is to capture the origin of our albedo subspace for every texel. The base $\boldsymbol{\rho}_0$ captures the

full skin pigmentation and its spatial detail. To achieve this goal, we require the actor to hold a neutral expression while also slowly rotating their head up-down, left-right, to form a cross pattern. This is a multi-frame step that simulates temporal multiplexing of illumination by moving the face instead of lights, thus varying the relative direction of illumination incident on the face. Note that light directions ω_i and colors $\mathbf{c}_{\omega_i}$ remain constant, but lighting visibility $\mathcal{L}_f$ and the integration hemisphere do vary with the changing tangent space of each surface patch, as obtained from the tracked 3D face mesh. This simple protocol leads to a well-constrained photometric stereo problem without requiring active illumination: it provides $F \approx 30$ frames at different illumination conditions (up to $4F \approx 120$ image samples $\mathbf{c}_{f\omega_o}$ per texel, depending on camera visibility $\mathcal{V}_f$), to which we fit 5 parameters in our model, as described next.

For these neutral frames, we fix base hemoglobin concentration $h_f = 0$ and $\psi_f = 1, \forall f$. We solve for temporally constant scalar $\varrho_f = \varrho_0$ and $\boldsymbol{\rho}_f = \boldsymbol{\rho}_0$ in RGB space, before converting albedo to Lab. Given the rigid face motion, we also compute a new per-texel tangent space normal $\mathbf{n}_t = \mathbf{n}_0$ (represented by a single height surface z_0) that is also constant over these neutral frames. We thus solve

$$\min_{\boldsymbol{\rho}_0, \varrho_0, z_0} \sum_f E_f(\boldsymbol{\Theta}_0) + \lambda_1 \|z_0\|_F^2 + \lambda_2 \|\varrho_0 - \bar{\varrho}\|_F^2 \,, \tag{7}$$

where $\lambda_1 = 1^{-5}$ and $\lambda_2 = 0.05$ are small regularization weights; $\bar{\varrho} = 1$ is used to regularize towards Fresnel reflection for skin, and $\|\cdot\|_F^2$ denotes the Frobenius norm.

In summary, this calibration step estimates 5 degrees of freedom per texel ($\boldsymbol{\rho}_0, \varrho_0, z$) via a multi-frame fit to nearly $4F \approx 120$ image samples acquired under varying illumination due to relative motion between head and light rig.

4.4 Step 1 – Per-Frame Normals $\mathbf{n}_f$

Once the calibration stage above is done, the only remaining degree of freedom in our albedo subspace is h_f. We now turn to a more difficult problem in which we independently process new frames with arbitrary facial expressions. Given a single frame f (up to 4 RGB samples per texel from 4 views), Step 1 estimates 3 degrees of freedom per texel, $\{h_f, \varrho_f, z_f\}$, as to minimize the rendering loss in Eq. (5). In this stage, our main goal is to estimate a high-detail normal field, parameterized by height surface z_f as above. To avoid ambiguities in representing shading in the input face images, we initially maintain $\psi_f = 1$ fixed; h_f and ϱ_f are allowed to vary but both are constrained to be spatially smooth. The intended effect is to push as much geometric detail as possible into the normal map represented by z_f, which is responsible for explaining most of the observed high-frequency shading. We solve

$$\begin{aligned}\min_{h_f, \varrho_f, z_f} \; & E_f(\boldsymbol{\Theta}_f) + \lambda_1 \|z_f\|_F^2 + \lambda_2 \|\varrho_f - \varrho_0\|_F^2 + \lambda_3 \|h_f\|_F^2 \\ & + \lambda_4 \|\nabla^2 z_f\|_F^2 + \lambda_5 \Big(\|\nabla \varrho_f\|_F^2 + \|\nabla h_f\|_F^2 \Big) \,.\end{aligned} \tag{8}$$

where ∇ denotes the gradient (forward differences) and ∇^2 is the Laplacian operator on a 3×3 neighborhood in texture space. We also found it beneficial to weakly constrain z_f to be smooth in small regions with ambiguous normal ($\lambda_4 = 0.005, \lambda_5 = 1.0$).

Our albedo subspace in Eq. 4 actually defines a sector along a line (*i.e.*, observable concentrations of hemoglobin), with origin at the base albedo. We thus regularize the estimates h_f to remain near 0 ($\lambda_3 = 0.002$). Another weak regularizer acts on ϱ_f to bias it towards the neutral ϱ_0 when data evidence is weak (λ_1, λ_2 as in Step 0).

To further improve detail resolution in z_f, we apply different weights per color channel ($w_R = 0.1, w_G = 0.3, w_B = 1.0$) to the loss in $E_f(\boldsymbol{\Theta}_f)$ to account for wavelength-dependent blurring due to subsurface scattering.

4.5 Step 2 – Per-Frame Albedo h_f, Specular Intensity ϱ_f, and Diffuse Ambient Occlusion ψ_f

In this step, we fix the normals estimated above and focus on recovering the other BRDF parameters (3 degrees of freedom per texel). To estimate optimal appearance parameters, we now weigh color channels uniformly ($w_R = w_G = w_B = 1$). In addition, we now also fit ψ_f and remove the spatial smoothness constraint from ϱ_f. The intended effect is to allow both to explain any residual shading (ambient occlusion on both diffuse and specular layers) not captured by the high-detail normals and base 3D face mesh. We solve

$$\min_{h_f, \varrho_f, \psi_f} E_f(\boldsymbol{\Theta}_f) + \lambda_2 \|\varrho_f - \varrho_0\|_F^2 + \lambda_3 \|h_f\|_F^2 + \lambda_5 \|\nabla h_f\|_F^2 \,. \tag{9}$$

Note that we still require that hemoglobin concentration values h_f be spatially smooth and not too far from the base albedo. Also, we maintain the regularizer on specular intensity, ϱ_f, biasing it towards the better constrained base ϱ_0 estimated in Step 0.

4.6 Sensitivity to Regularization Weights

Normal optimization requires a very small number of passes (approximately 5) on the Red-Black texel grid. Thus, the final displacement values $z(u, v)$ remain near 0, reducing sensitivity to weight λ_1; if λ_1 is too high, less surface detail is recovered. Weight λ_4 is also very small and only needs to be strong enough to smooth small regions with temporally noisy normals. Weights λ_2 and λ_3 regulate variability in appearance relative to that of the neutral face; if set too high, the dynamic model behaves more like a static one. Finally, if λ_5 is too weak, high-frequency shading is more prominently encoded as specular ambient occlusion, leading to diminished recovery of surface detail in z (Fig. 12).

5 DATA ACQUISITION AND PREPROCESSING

In this section we describe how we acquired the input data for the presented method, as well as data preprocessing steps to compute derived data using prior art algorithms.

5.1 Hardware Setup

Our capture setup (shown in Fig. 3) consists of a multi-view stereo-rig composed of eight 12MP Ximea CB120MG monochrome cameras arranged in four stereo-pairs in order to cover the entire face of our actor, which are used to reconstruct our base 3D model. We interleave four additional color cameras (20MP Ximea CB200CG), one between each stereo-pair, to record RGB color data for facial appearance estimation. Both geometry and appearance data are acquired at 30 frames per second. During performance capture, we illuminate our actors with constant white illumination provided by

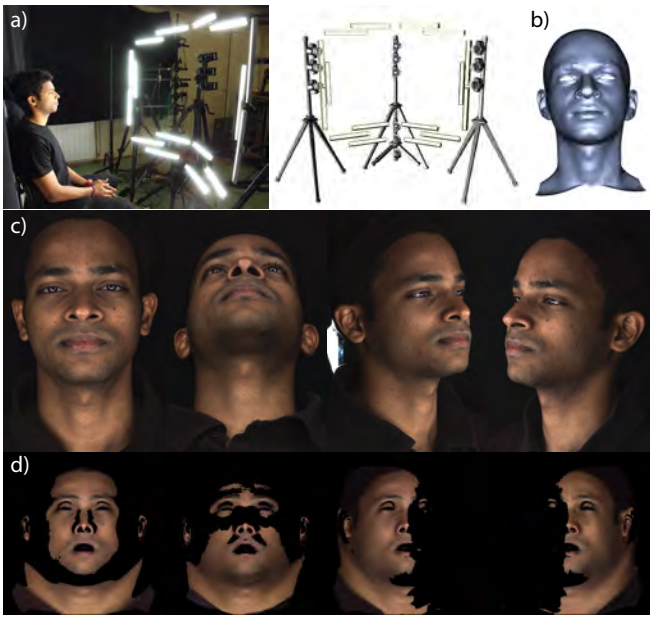

Fig. 3. **Capture setup and pre-processing pipeline** – Our capture setup (a) consists of 12 cameras: 8 monochrome for Multi-View Stereo geometry reconstruction (b), and 4 colour cameras (c) for appearance capture. From these, we compute per-camera textures (d) which allow us to efficiently formulate our inverse rendering problem in texture-space.

16 LED strips placed in front of the actor. The strips were clustered to produce two horizontal and two vertical linear light sources, where the horizontal ones illuminate the face slightly from below and above and the vertical ones from each half-profile.

5.2 Calibration

We require both geometrically and photometrically calibrated cameras. After each acquisition session, we capture a planar calibration target with fiducial markers [Garrido-Jurado et al. 2014] for geometric calibration, plus an X-Rite ColorChecker® chart for photometric calibration of the acquired footage with respect to a linear sRGB color space.

5.3 Environment Map

We also need to accurately model the incident illumination for inverse rendering. For this purpose, we acquire an HDR light probe of the surrounding environment by capturing a mirror sphere at several exposures using the frontal color camera. From our calibrated cameras, we estimate the position of the mirror sphere in the scene and compute a latitude-longitude environment map (1024×512). We compress this environment map to 900 uniformly distributed light directions by integrating for each light direction the radiance within the corresponding Voronoi area in the environment map. For human skin, reducing to a few hundred light directions is reasonable and yields a lighting resolution comparable to the typical one of Light Stages [Fyffe et al. 2014; Ma et al. 2007; Weyrich et al. 2006].

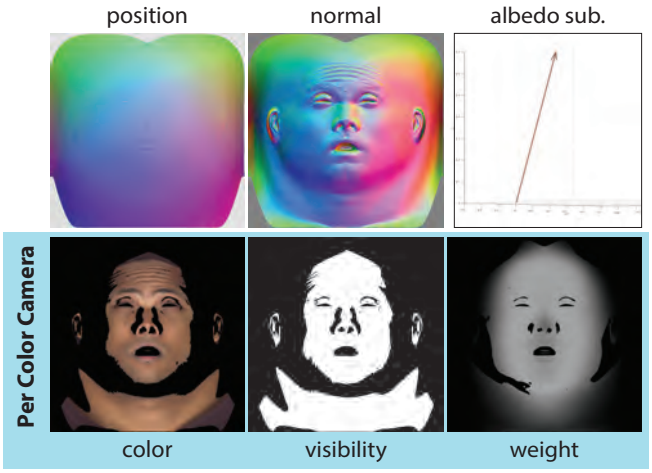

Fig. 4. **Input Data** – We prepare the input data for the inverse renderer in texture domain, computing per frame position and normal maps. We also pre-compute the dynamic albedo blood flow subspace as a line in Lab. Furthermore, we generate for every color camera a color texture and visibility maps, as well as a weight map that indicates sharpness and reliability.

5.4 Base Geometry Reconstruction

For the presented dynamic appearance capture we require a base mesh, fully tracked over time. We apply a state-of-the-art passive multi-view performance capture system to reconstruct geometry using the eight monochrome cameras [Beeler et al. 2010] and track a consistent topology to all frames [Beeler et al. 2011]. The resulting shapes are stabilized with respect to the neutral face [Beeler and Bradley 2014]. From the four color cameras we compute high-resolution texture maps. Since our inverse rendering framework will operate in texture space, we encode the mesh vertex positions and base normals as texture maps for every frame. We further compute for each color camera per-frame visibility textures as well as weight textures (Fig. 4). These per-texel weights measure how sharp the texel is, integrating information from camera focus and motion blur.

5.5 Albedo Blood Flow Subspace

The dynamic albedo will be described by varying blood flow over time. As detailed in Section 3.2, this blood flow is parameterized by an albedo subspace, characterized by a single line in Lab color space. Since the slope of the line is person-specific and depends on skin type, we propose a simple method to pre-compute the line for the given capture subject. Using a digital SLR camera with a mounted ring flash, we photograph a small patch of skin in burst mode, immediately after the actor presses firmly on the skin with their fingers. This sequence of photos provides a time-varying measure of hemoglobin concentrations, to which we fit a line in Lab space. We use linear cross-polarization on the flash and camera lens to filter out specular highlights, and we align the images using optical flow [Brox et al. 2004] to account for small motion. The images are color calibrated using an X-Rite ColorChecker, and we place white markers in the scene to compute and account for any variability in the ring flash from photo to photo; Fig. 5 (1^{st} row)

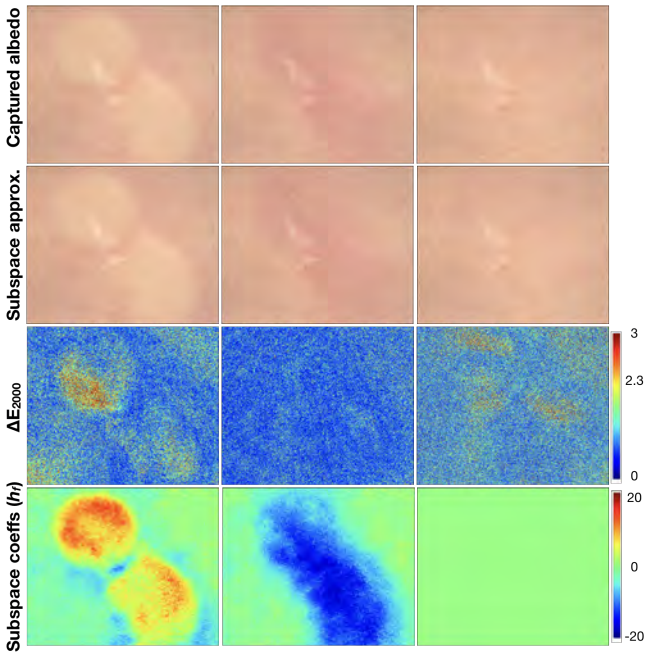

Fig. 5. **Blood Flow Subspace** – We precompute the color subspace for dynamic albedo by photographing the subject in burst mode after pressing firmly on a forehead skin patch (1^{st} row). These images are aligned and calibrated photometrically, before the person-specific albedo line is computed in Lab space. The resulting albedo subspace is validated both qualitatively (2^{nd} row) and quantitatively (3^{rd} row) using a standard perceptual metric; $\Delta E_{2000} \leq 3.0$ corresponds to mostly imperceptible differences. Row 4 shows the distribution of coefficients along the albedo subspace right after pressure from the fingers has been released (left), after blood flows back in the affected region (middle) and after blood flow has settled (right).

shows a subset of captured albedos for one actor. The corresponding approximated albedos, using the calculated subspace for the same person, are also shown in Fig. 5 (2^{nd} row) and closely match the captured data. We further evaluate our blood flow subspace in Fig. 5 (3^{rd} row) by computing the standard, perceptually-motivated color difference metric CIE ΔE_{2000} between ground-truth and approximated albedos. Finally, we show in Fig. 5 (4^{th} row) how blood flow affects the distribution of coefficients (h_f) along the albedo subspace: positive values correspond to blanching while negative values show reddening of the skin patch.

6 RESULTS AND EVALUATION

We now assess the individual appearance maps computed by our method and show the outcome of the complete pipeline. Our experiments were run on a 12-core Mac Pro desktop computer with average runtimes of 250 min for Step 0 (processing 30 frames on average), 15 min/frame for Step 1, and 2 min/frame for Step 2.

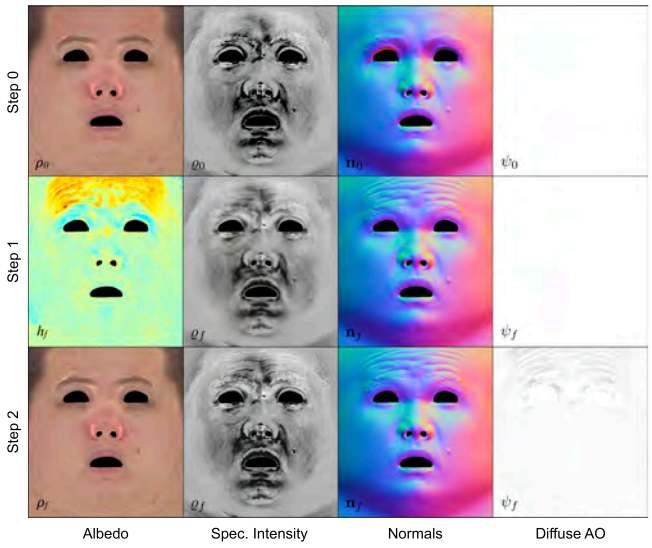

Fig. 6. **Recovered Maps** – Our method recovers per-frame appearance maps, including albedo, specular intensity, high-detail normals, and diffuse ambient occlusion (AO). We show the resulting maps from Step 2 for a single frame (bottom row). Also shown are the intermediary results for Step 0 (top row), where neutral maps are computed only once per actor, and Step 1 (middle row), where smooth specular intensity ϱ_f pushes detail into the normal map $\boldsymbol{n}_f$ and single-channel hemoglobin map h_f shows blood flow (in red). The final albedo $\boldsymbol{\rho}_f$ is defined by h_f and $\boldsymbol{\rho}_0$.

6.1 Dynamic Appearance Maps

The output of the proposed system is a set of four parameter maps per frame, namely albedo, diffuse ambient occlusion, specular intensity, and high-resolution normals. This output is illustrated in Fig. 6, which also shows the neutral maps computed in Step 0 and intermediary maps obtained in Step 1. Except for the results in Step 0 (computed once per actor), these maps are time-varying and can be used with existing rendering packages to render a face under different illumination as shown in the next section.

Albedo Map. The albedo map contains the shading-free color of the face. When acquiring albedo in film production, an actor's face is typically lit as uniformly as possible and captured using cross-polarization. While the cross-polarized filters can succeed at removing direct specular reflection from the skin, diffuse shading will remain baked into the resulting map, Fig. 7 (a). Applying the proposed inverse rendering pipeline on cross-polarized data allows to remove this shading and produces a shading-free albedo, Fig. 7 (b). Finally, the presented method succeeds at extracting a very similar albedo from regular un-polarized data, showing that it can effectively separate diffuse and specular reflection computationally at a similar quality as physical polarization.

Albedo changes over time due to blood flow (Fig. 8), either caused by physiological effects such as exercise (a) or due to physical pressure exerted onto the skin when activating facial muscles (b). Blood flow is not instantaneous, which causes hysteresis effects over time. This effect is shown in Fig. 8 (b), where it takes several frames

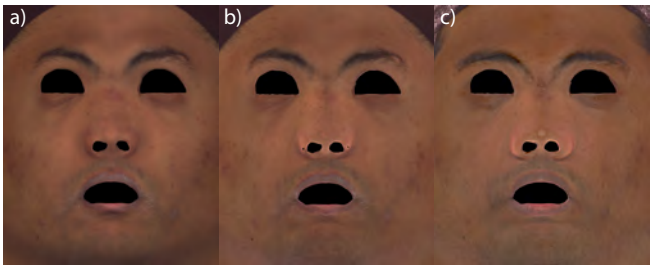

Fig. 7. **Albedo Validation** – (a) Using cross-polarization, an albedo texture can be extracted directly from the cameras without specular shading. However, diffuse shading remains baked in. (b) From the same data, our inverse rendering pipeline can provide an excellent albedo. (c) Even from regular unpolarized footage (captured at a different time from our setup in Fig. 3), our pipeline yields an albedo of similar quality with negligible shading.

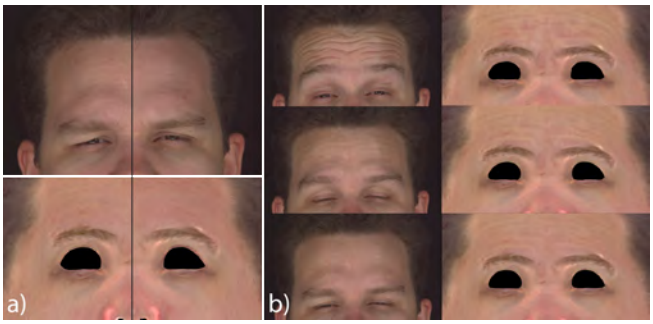

Fig. 8. **Dynamic Albedo Map** – We compare the input images to our dynamic albedo maps. (a) Physiological effects such as exercise or overheating can alter blood flow which we see here by splitting two different frames left-right, particularly in the forehead. (b) Facial expressions also alter blood flow (shown as forehead crop over time). Blood flow can be apparent for several frames after the expression returns to neutral due to hysteresis over time. Our method recovers both of these effects in the captured performance.

until blood has fully returned after releasing an expression. By constraining albedo to change along a one-dimensional line, which we precompute per actor as described in Section 5.5, the proposed method recovers high-quality per-frame albedo maps.

Diffuse Ambient Occlusion Map. We introduce this map in order to capture residual diffuse shading that stems from the base mesh not faithfully capturing the geometry everywhere, in particular in wrinkle folds (e.g. refer to Fig. 6).

Specular Intensity Map. This map modulates the light reflected off the skin surface. The amount of light reflected depends on a variety of factors, such as oiliness or wetness of the skin (Fig. 9 (left)) or changes in skin microstructure due to stretch (Fig. 9 (right)) and tissue scarring (Fig. 10). This map also accounts for specular ambient occlusion caused by mesoscopic skin detail, such as pores, when not completely explained by the normal map. Some of these properties change over time and motivate per-frame specular intensity maps.

Dynamic Normal Map. Skin surface is not flat but covered by mesoscopic detail that is too finescale to be picked up by the coarse

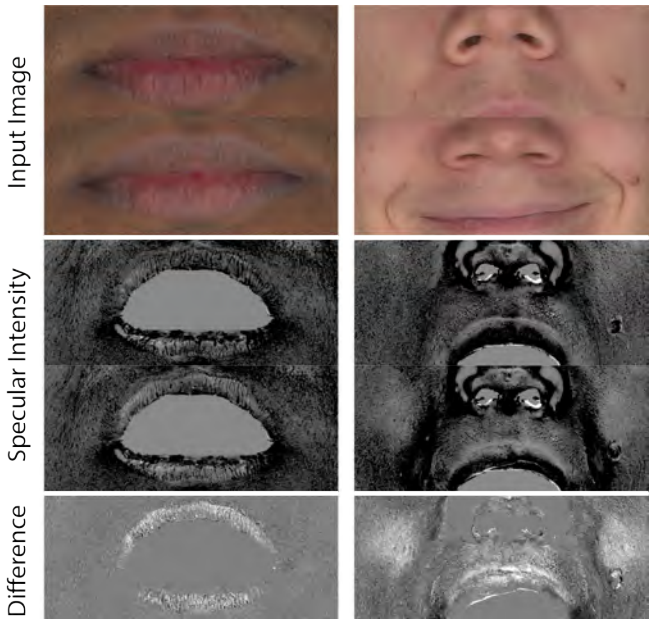

Fig. 9. **Dynamic Specular Intensity Map** – The amount of light reflected off the skin changes over time. The left column shows an example where the actor wet his lips between two takes, which increases specular reflectance. And in the right column the expression of the actor causes skin to stretch as the cheeks bulge, which in amounts in an increase in specular reflection.

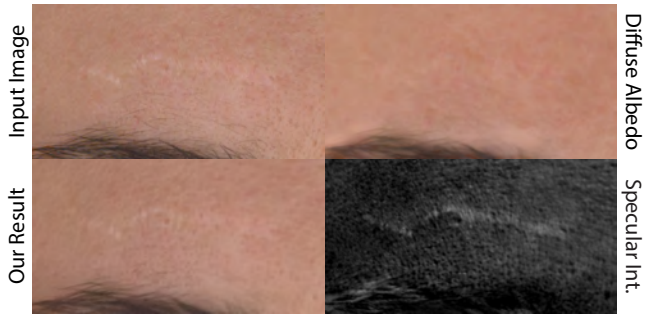

Fig. 10. **Scar** – A small scar is shown on the subject's forehead. As the scar is not deep, it influences predominately surface reflectance, namely specular intensity, since it is smoother than the surrounding skin tissue.

base mesh, such as pores and fine wrinkles. As skin stretches or compresses, these details change dramatically and strongly influence the appearance of the face. The proposed method can recover high-quality per-frame normal maps that encode this dynamic geometric detail (Fig. 11). The effect of the different constraints used in resolving fine geometric detail in Step 1 is illustrated in Fig. 12. We also compare the level of detail obtained by our method to that of the mesoscopic augmentation approach by Beeler et al. [2010]; Fig. 13 shows improved skin detail in our results computed from a cheek patch at both 2K and 4K texture resolutions.

Fig. 11. **Skin Detail** – Closeups of various areas on the face show the level of detail the method can recover, ranging from pores to finescale wrinkles. The last row shows two examples of dynamically changing detail: (left) a patch on the forehead exhibits strong anisotropic wrinkling when the eyebrows are raised; and (right) a patch around the chin shows deformation caused by a muscle pulling skin tissue towards the upper left side of the image, causing pores and wrinkles to stretch in an elliptical pattern.

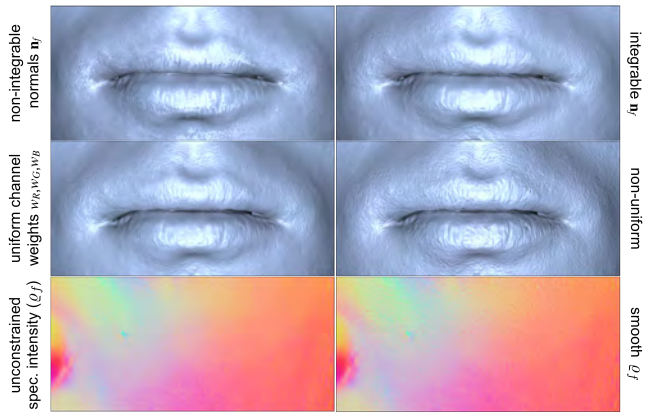

Fig. 12. **Ablation study** – The left and right images in each row illustrate the effect of the different constraints used in Step 1 to resolve fine geometric detail in the per-frame normal map (lips and left cheek).

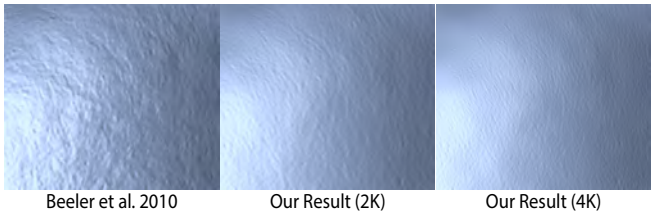

Fig. 13. **Mesoscopic Comparison** – We compare recovered skin detail against mesoscopic augmentation ("dark is deep") by Beeler et al. [2010]. Finest scale detail is obtained by our method at 4K texture resolution.

6.2 Dynamic Appearance Relighting

The recovered maps can be used to create renders of the face with a high degree of realism as demonstrated in Fig. 16 (col. 2). From the per-frame normal maps (col. 8), diffuse surface shading (col. 5) is computed using a Lambertian BRDF modulated by diffuse ambient occlusion, and specular surface shading (col. 6) is computed using the Cook-Torrance model with a Blinn-Phong distribution. The final rendering is computed by multiplying the diffuse shading with the dynamic albedo (col. 4) and adding the specular shading modulated by specular intensity (col. 7). The maps have been recovered by minimizing the difference of the render to the input image (col. 1). An evaluation of the dynamic appearance is provided as an error per texel, computed as the absolute pixel re-render error averaged over all channels and views, illustrated as a percentage (col. 3).

For validation, we use the maps recovered for a neutral face to re-render the face under a different illumination condition (Fig. 14), where we configured the lights to illuminate the face only from the left side. The rendering can reproduce the surface appearance of the face under this novel condition very well, as compared to a reference photo under the same conditions. Note that our result is lacking a more elaborate sub-surface contribution, as the scope of this work is somewhat focused on surface reflection, Fig. 10. This is akin to existing skin appearance acquisition used for visual effects. Nevertheless, our method can readily make use of simple approximate techniques for sub-subsurface scattering in texture space [d'Eon et al. 2007].

Finally, we demonstrate the ability to relight the captured faces under various illumination conditions by re-rendering in a commercial renderer (Autodesk Maya®). Although our method does not readily capture subsurface scattering parameters, these additional parameters can be manually set during relighting as to achieve higher realism, as in Fig. 1 (right) and Fig. 15. In our relighting results, we used the same subsurface scattering parameters for all actors (weight 1.0, subsurface color set to albedo texture, RGB scattering radii [0.324, 0.148, 0.064] and scale 0.25). For more results, we refer the reader to the accompanying video.

7 DISCUSSION

In this work, we have presented a practical approach for measurement-based modeling of dynamic facial appearance. Unlike previous works that have modeled appearance dynamics as a linear blend between a few acquired key poses, we present a method that achieves truly dynamic appearance capture at video framerates of acquisition, and under standard uniform illumination setups that are commonly employed for facial performance capture. We believe our approach takes a big step forward in bridging the gap in rendering fidelity for dynamic facial appearance acquired with passive acquisition compared to that achieved using specialized active illumination setups such as Light Stages. Given limited measurements from passive acquisition in few viewpoints, robust fitting of the variability in diffuse albedo during a facial performance is made possible with our novel albedo subspace and a comprehensive set of constraints on appearance and geometry parameters.

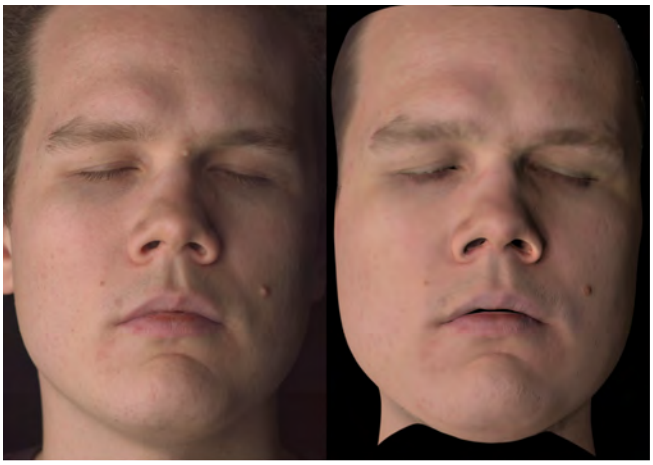

Fig. 14. **Appearance Validation** – The face is rendered under a novel lighting configuration with half the lights turned off. Our result closely matches a reference image recorded under the same conditions.

Fig. 15. **Relighting** – We demonstrate the ability to re-render the captured faces with our appearance parameters under novel environment lighting.

However, given limited input, we do make a few simplifications to the overall dynamic facial appearance model. We currently model the body (subsurface) reflection purely with a Lambertian BRDF and only model albedo change during skin dynamics. Modeling changes in additional parameters of a more sophisticated subsurface scattering model might be required for increased realism for some applications – for instance, modeling any change in spatially varying skin translucency, or explicit modeling of changes in melanin *vs.* hemoglobin concentrations. Our proposed albedo subspace is based on the assumption of blood flow being the dominant factor for changes in albedo which is true for typical facial performances. However, our dynamic albedo model does not consider the effects of any change in melanin concentration or changes due to application of any cosmetics on skin. Our formulation for skin dynamics, while effective in anisotropically updating the surface normal, currently enforces the specular lobe (roughness) to remain isotropic. A more accurate modeling of skin appearance under deformation will additionally require anisotropic modeling of the specular BRDF under stretch and compression. This remains an important challenge for future work, as capturing the shape of specular lobe can be an ill-posed problem even in the static scenario with active illumination [Ghosh et al. 2008]. Despite these current limitations, we demonstrate high fidelity results with dynamic appearance changes for several subjects with different skin types which we believe highlight the unprecedented capabilities of the proposed approach.

ACKNOWLEDGMENTS

We wish to thank our 3D artist Maurizio Nitti for his help with face rendering for the relighting experiments. We also thank Virginia Ramp, Anurag Vempati and Tejaswi Digumarti for posing as capture subjects.

REFERENCES

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

Oleg Alexander, Mike Rogers, William Lambeth, Jen-Yuan Chiang, Wan-Chun Ma, Chuan-Chang Wang, and Paul Debevec. 2010. The digital emily project: Achieving a photorealistic digital actor. *Computer Graphics and Applications, IEEE* 30, 4 (2010), 20–31.

R Rox Anderson and John A Parrish. 1981. The optics of human skin. *Journal of investigative dermatology* 77, 1 (1981), 13–19.

Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-Quality Single-Shot Capture of Facial Geometry. *ACM Transactions on Graphics (TOG)* 29, 3 (2010), 40:1–40:9.

Thabo Beeler and Derek Bradley. 2014. Rigid stabilization of facial expressions. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–9.

Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics (ACM)* 30, Article 75 (August 2011), 10 pages. Issue 4.

Derek Bradley, Wolfgang Heidrich, Tiberiu Popa, and Alla Sheffer. 2010. High resolution passive facial performance capture. *ACM Transactions on Graphics (TOG)* 29, 4 (2010), 41.

Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert. 2004. High accuracy optical flow estimation based on a theory for warping. In *ECCV*. 25–36.

Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time High-fidelity Facial Performance Capture. *ACM Transactions on Graphics (TOG)* 34, 4, Article 46 (July 2015), 9 pages.

A Chardon, I Cretois, and C Hourseau. 1991. Skin colour typology and suntanning pathways. *International journal of cosmetic science* 13, 4 (1991), 191–208.

Robert Cook and Kenneth E. Torrance. 1981. A reflectance model for computer graphics. *Computer Graphics (SIGGRAPH '81 Proceedings)* 15, 3 (1981), 301–316.

Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. 2000. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 145–156.

Eugene d'Eon, David Luebke, and Eric Enderton. 2007. Efficient Rendering of Human Skin. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques (EGSR'07)*. 147–157.

Craig Donner and Henrik Wann Jensen. 2005. Light Diffusion in Multi-Layered Translucent Materials. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 1032–1039.

Craig Donner and Henrik Wann Jensen. 2006. A Spectral BSSRDF for Shading Human Skin. In *Proceedings of the 17th Eurographics Conference on Rendering Techniques (EGSR '06)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 409–417. https://doi.org/10.2312/EGWR/EGSR06/409-417

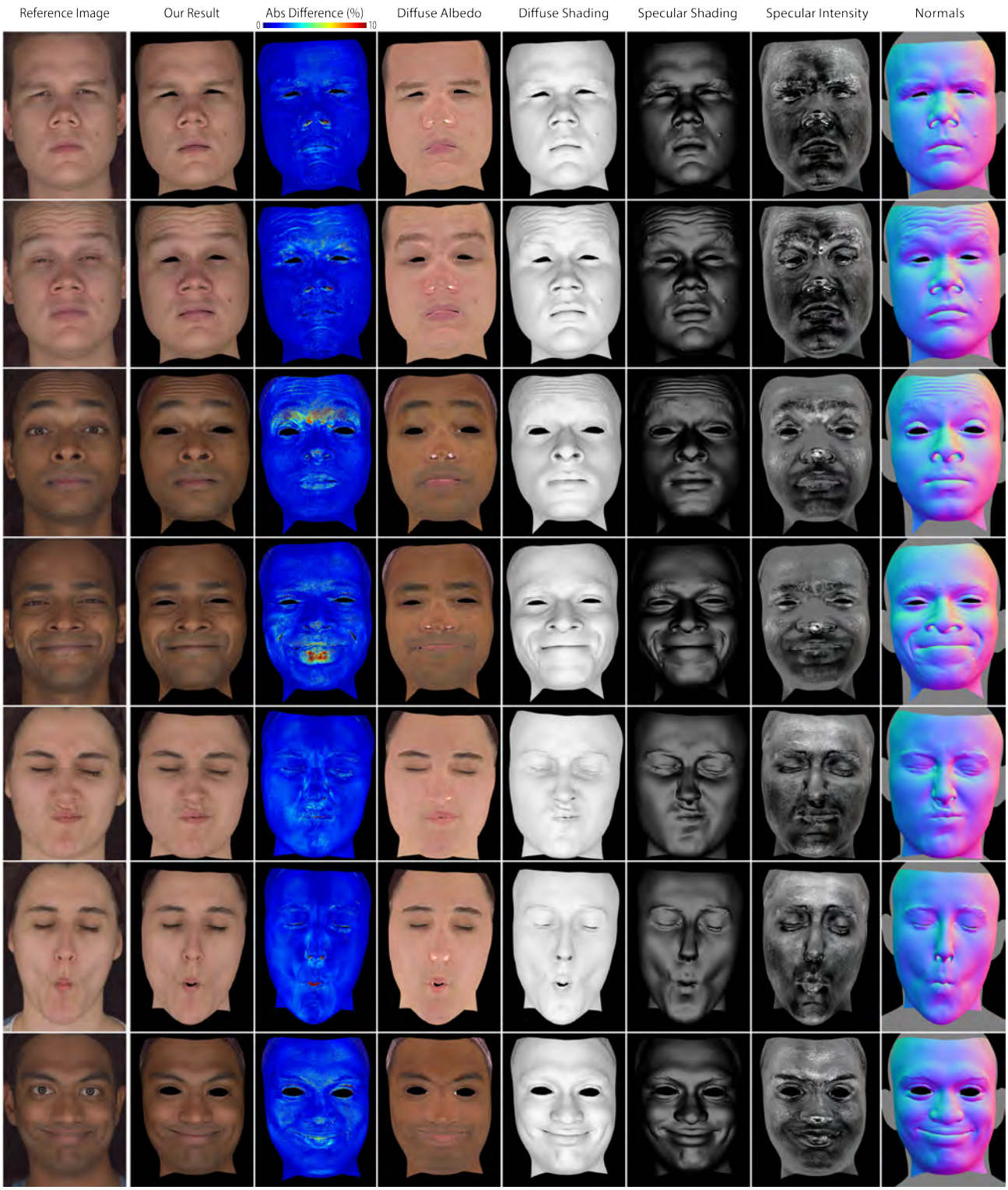

Fig. 16. **Dynamic Appearance** – The proposed system recovers dynamic albedo, diffuse ambient occlusion, dynamic specular intensity and per-frame normals by rendering images from these maps and minimizing the difference to the captured reference images. Per-texel error values are averaged over all views and color channels, relative to the observed range of image values.

Craig Donner, Tim Weyrich, Eugene d'Eon, Ravi Ramamoorthi, and Szymon Rusinkiewicz. 2008. A Layered, Heterogeneous Reflectance Model for Acquiring and Rendering Human Skin. *ACM Transactions on Graphics (TOG)* 27, 5 (Dec. 2008), 140:1–140:12.

Martin Fuchs, Volker Blanz, Hendrik Lensch, and Hans-Peter Seidel. 2005. Reflectance from Images: A Model-Based Approach for Human Faces. *IEEE Transactions on Visualization and Computer Graphics* 11, 3 (May 2005), 296–305. https://doi.org/10.1109/TVCG.2005.47

Graham Fyffe and Paul Debevec. 2015. Single-Shot Reflectance Measurement from Polarized Color Gradient Illumination. In *International Conference on Computational Photography (ICCP)*.

Graham Fyffe, Paull Graham, Borom Tunwattanapong, Abhijeet Ghosh, and Paul Debevec. 2016. Near-Instant Capture of High-Resolution Facial Geometry and Reflectance. *Computer Graphics Forum (CGF)* 35, 2 (2016), 353–363. https://doi.org/10.1111/cgf.12837

Graham Fyffe, Tim Hawkins, Chris Watts, Wan-Chun Ma, and Paul Debevec. 2011. Comprehensive Facial Performance Capture. *Computer Graphics Forum (CGF)* 30, 2 (2011).

Graham Fyffe, Andrew Jones, Oleg Alexander, Ryosuke Ichikari, and Paul Debevec. 2014. Driving High-Resolution Facial Scans with Video Performance Capture. *ACM Transactions on Graphics (TOG)* 34, 1, Article 8 (Dec. 2014), 14 pages. https://doi.org/10.1145/2638549

Pablo Garrido, Levi Valgaerts, Chenglei Wu, and Christian Theobalt. 2013. Reconstructing Detailed Dynamic Face Geometry from Monocular Video. *ACM Transactions on Graphics (TOG)* 32, 6 (November 2013), 158:1–158:10. https://doi.org/10.1145/2508363.2508380

Sergio Garrido-Jurado, Rafael Mu noz Salinas, F.J. Madrid-Cuevas, and Manuel Jesus Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280 – 2292. https://doi.org/10.1016/j.patcog.2014.01.005

Abhijeet Ghosh, Graham Fyffe, Borom Tunwattanapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics (TOG)* 30, 6 (2011), 129.

Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. 2008. Practical Modeling and Acquisition of Layered Facial Reflectance. *ACM Trans. Graph.* 27, 5 (Dec. 2008), 139:1–139:10.

Paulo Gotardo, Tomas Simon, Yaser Sheikh, and Iain Matthews. 2015. Photogeometric Scene Flow for High-Detail Dynamic 3D Reconstruction. In *IEEE International Journal of Computer Vision*. 846–854.

Paul Graham, Borom Tunwattanapong, Jay Busch, Xueming Yu, Andrew Jones, Paul Debevec, and Abhijeet Ghosh. 2013. Measurement-Based Synthesis of Facial Microgeometry. *Computer Graphics Forum (CGF)* 32, 2 (2013), 335–344.

Tim Hawkins, Andreas Wenger, Chris Tchou, Andrew Gardner, Fredrik Göransson, and Paul Debevec. 2004. Animatable Facial Reflectance Fields. In *Proceedings of the Fifteenth Eurographics Conference on Rendering Techniques (EGSR'04)*. Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 309–319. https://doi.org/10.2312/EGWR/EGSR04/309-319

Loc Huynh, Weikai Chen, Shunsuke Saito, Jun Xing, Koki Nagano, Andrew Jones, Paul Debevec, and Hao Li. 2018. Mesoscopic Facial Geometry Inference Using Deep Neural Networks. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Alexandru Eugen Ichim, Sofien Bouaziz, and Mark Pauly. 2015. Dynamic 3D Avatar Creation from Hand-held Video Input. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 45:1–45:14.

Henrik Wann Jensen, Steve Marschner, Marc Levoy, and Pat Hanrahan. 2001. A practical model for subsurface light transport. In *In Proceedings of ACM SIGGRAPH*. 511–518.

Jorge Jimenez, Timothy Scully, Nuno Barbosa, Craig Donner, Xenxo Alvarez, Teresa Vieira, Paul Matts, Verónica Orvalho, Diego Gutierrez, and Tim Weyrich. 2010. A Practical Appearance Model for Dynamic Facial Color. *ACM Transactions on Graphics (TOG)* 29, 6, Article 141 (Dec. 2010), 10 pages. https://doi.org/10.1145/1882261.1866167

Jorge Jimenez, Veronica Sundstedt, and Diego Gutierrez. 2009. Screen-space perceptual rendering of human skin. *ACM Transactions on Applied Perception* 6, 4 (2009), 23:1–23:15.

Oliver Klehm, Fabrice Rousselle, Marios Papas, Derek Bradley, Christophe Hery, Bernd Bickel, Wojciech Jarosz, and Thabo Beeler. 2015. Recent Advances in Facial Appearance Capture. *Computer Graphics Forum (CGF)* 34, 2 (May 2015), 709–733.

Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. 2007. Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques (EGSR'07)*. Eurographics Association, 183–194.

Wan-Chun Ma, Andrew Jones, Jen-Yuan Chiang, Tim Hawkins, Sune Frederiksen, Pieter Peers, Marko Vukovic, Ming Ouhyoung, and Paul Debevec. 2008. Facial Performance Synthesis Using Deformation-driven Polynomial Displacement Maps. *ACM Transactions on Graphics (TOG)* 27, 5, Article 121 (Dec. 2008), 10 pages. https://doi.org/10.1145/1409060.1409074

Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. 1999. Image-based BRDF Measurement Including Human Skin. In *Proceedings of the 10th Eurographics Conference on Rendering (EGWR'99)*. 131–144.

Koki Nagano, Graham Fyffe, Oleg Alexander, Jernej Barbic, Hao Li, Abhijeet Ghosh, and Paul E Debevec. 2015. Skin microstructure deformation with displacement map convolution. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 109.

Ruth Onn and Alfred Bruckstein. 1990. Integrability disambiguates surface recovery in two-image photometric stereo. *International Journal of Computer Vision* 5, 1 (1990), 105–113.

Shunsuke Saito, Lingyu Wei, Liwen Hu, Koki Nagano, and Hao Li. 2017. Photorealistic Facial Texture Inference Using Deep Neural Networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Christoph Schlick. 1994. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum (CGF)* 13, 3 (1994), 233–246.

Scott Shafer. 1985. Using Color to Separate Reflection Components. *Color Research & Application* 10, 4 (1985), 210–218.

Gaurav Sharma and Raja Bala. 2002. *Digital color imaging handbook*. CRC press.

Fuhao Shi, Hsiang-Tao Wu, Xin Tong, and Jinxiang Chai. 2014. Automatic acquisition of high-fidelity facial performances using monocular videos. *ACM Transactions on Graphics (TOG)* 33, 6 (2014), 222.

Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, and Christian Theobalt. 2012. Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting. *ACM Transactions on Graphics (TOG)* 31, 6 (November 2012), 187:1–187:11. https://doi.org/10.1145/2366145.2366206

Javier von der Pahlen, Jorge Jimenez, Etienne Danvoye, Paul Debevec, Graham Fyffe, and Oleg Alexander. 2014. Digital Ira and Beyond: Creating Real-time Photoreal Digital Actors. In *ACM SIGGRAPH 2014 Courses (SIGGRAPH '14)*. ACM, New York, NY, USA, Article 1, 384 pages. https://doi.org/10.1145/2614028.2615407

Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. 24, 3 (2005), 756–764.

Tim Weyrich, Jason Lawrence, Hendrik P. A. Lensch, Szymon Rusinkiewicz, and Todd Zickler. 2009. Principles of Appearance Acquisition and Representation. *Found. Trends. Comput. Graph. Vis.* 4, 2 (Feb. 2009), 75–191.

Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. 2006. Analysis of Human Faces Using a Measurement-based Skin Reflectance Model. *ACM Transactions on Graphics (TOG)* 25, 3 (July 2006), 1013–1024.

Cyrus A Wilson, Abhijeet Ghosh, Pieter Peers, Jen-Yuan Chiang, Jay Busch, and Paul Debevec. 2010. Temporal upsampling of performance geometry using photometric alignment. *ACM Transactions on Graphics (TOG)* 29, 2 (2010), 17.

Chenglei Wu, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. An anatomically-constrained local deformation model for monocular face capture. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 115.

Shugo Yamaguchi, Shunsuke Saito, Koki Nagano, Yajie Zhao, Weikai Chen, Kyle Olszewski, Shigeo Morishima, and Hao Li. 2018. High-Fidelity Facial Reflectance and Geometry Inference From an Unconstrained Image. *ACM Transactions on Graphics (TOG)* 37, 4 (2018).