

Single-Shot Facial Capture using Polarized RGB Sinusoidal Illumination

ARVIN LIN, Imperial College London, United Kingdom and Lumirithmic, United Kingdom

ABHIJEET GHOSH, Imperial College London, United Kingdom

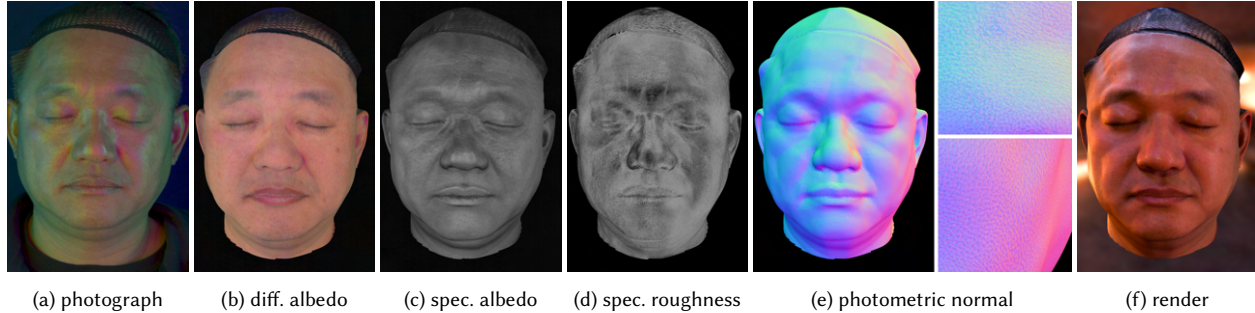


Fig. 1. Our single-shot method employs RGB sinusoidal illumination (a) to capture: diffuse albedo (b), specular albedo (c), specular roughness (d), and photometric specular normals (e). The captured appearance maps enable high-quality realistic rendering in novel lighting environment (Grace cathedral) (f).

We present a novel single-shot method for capturing high-quality facial appearance that enables per-pixel estimation of diffuse albedo, specular albedo, specular roughness, and photometric (specular) normals using only linearly polarized RGB illumination and consumer cameras. Our approach leverages color-multiplexed sinusoidal lighting encoded across RGB channels, allowing phase-based decomposition of reflectance parameters per view, without time-multiplexing or iterative refinement. Unlike prior works that jointly optimize appearance and geometry, we decouple the estimation process – separately recovering spatially-varying specular albedo and roughness for the first time in a single-shot capture setup. Additionally, we estimate high-frequency specular photometric normals independently using view-dependent specular phase cues, obtaining accurate surface mesostructure. We demonstrate our system using a practical monitor-based capture setup with 15 polarized DSLRs, producing detailed reflectance maps suitable for photorealistic rendering. Our approach achieves higher accuracy in reflectance separation and more accurate estimation of fine-scale surface details compared to previous single-shot methods.

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

Additional Key Words and Phrases: Passive Computational Photography, Appearance Capture

ACM Reference Format:

Arvin Lin and Abhijeet Ghosh. 2025. Single-Shot Facial Capture using Polarized RGB Sinusoidal Illumination. 1, 1 (September 2025), 11 pages. <https://doi.org/10.1145/3757377.3763942>

1 Introduction

Acquisition of high-fidelity facial reflectance is essential for photorealistic rendering of digital humans. A complete appearance model captures facial geometry as well as spatially-varying diffuse albedo,

specular albedo, specular roughness, and detailed photometric normal that encodes mesoscopic skin details. Accurately capturing these attributes is essential for realistic and physically accurate rendering of a human face.

Established techniques capture high-fidelity reflectance maps with active illumination, such as polarized spherical gradient illumination [Ghosh et al. 2009, 2011; Ma et al. 2007]. These methods enable robust separation of diffuse and specular responses while capturing detailed photometric normals. Despite their effectiveness, these methods require time-multiplexed lighting. This constraint limits active illumination methods to capturing static subjects.

To overcome this limitation, recent efforts have focused on passive or single-shot alternatives [Gotardo et al. 2018; Ha et al. 2024; Riviere et al. 2020]. These approaches leverage inverse rendering to infer reflectance parameters through joint optimization. While these methods do not require active illumination, joint estimation often results in entangled parameters, especially between specular albedo and roughness. Furthermore, single-shot inverse rendering methods require regularization and heuristic-based priors to converge to a plausible solution.

In this work, we present a novel single-shot method for high-quality facial appearance capture that combines polarization-based separation with color-multiplexed sinusoidal illumination. Our system enables the direct estimation of spatially-varying reflectance parameters, including the disentanglement of specular albedo and roughness as well as estimation of specular photometric normals, without requiring temporal multiplexing or joint-optimization over the reflectance parameters.

Authors' Contact Information: Arvin Lin, Imperial College London, London, United Kingdom and Lumirithmic, London, United Kingdom, arvin.lin19@imperial.ac.uk; Abhijeet Ghosh, Imperial College London, London, United Kingdom, abhijeet.ghosh@imperial.ac.uk.

Our key contributions are:

- (1) **First single-shot method to independently estimate spatially-varying specular albedo and specular roughness.** We resolve the under-constrained nature of roughness estimation by analyzing per-channel phase responses from colored sinusoidal illumination, enabling direct parameter separation under a single illumination.
- (2) **Photometric specular normal estimation via direct optimization.** In contrast to prior single-shot approaches that jointly optimize normals with reflectance parameters leading to a “mixed” normal estimate, we recover photometric specular normals independently using view-dependent specular phase cues. This leads to sharper, detail-preserving normal maps that capture fine-scale skin mesostructure.

We evaluate our method using a practical monitor-based illumination setup. Our results across multiple subjects with age, gender, and skin tone variation demonstrate high-quality reflectance capture and rendering with our single-shot approach.

2 Related Work

Facial appearance capture plays an important role in enabling photorealistic digital human rendering. Research methods span a wide range, from high-quality studio captures [Ghosh et al. 2011; Riviere et al. 2020] to practical methods with mobile devices [Dib et al. 2021; Han et al. 2024; Rainer et al. 2023]. The targeted appearance properties also range from standard SVBRDFs to biophysical properties [Gitlina et al. 2020; Ha et al. 2024; Li et al. 2024]. More recently, neural representations have also been used in appearance models [Bi et al. 2021, 2020; Saito et al. 2024]. In this section, we focus our discussion on high-quality shape and SVBRDF acquisition methods.

Static appearance capture: High-fidelity facial appearance has traditionally been captured with active illumination setups. The introduction of the light stage [Debevec et al. 2000] enabled static captures with various time-multiplexed illumination patterns. Ma et al. [2007] utilized polarized spherical gradient illumination to acquire specular and diffuse normal maps and albedos. Their method is later extended to multiview capture [Ghosh et al. 2011] and roughness capture [Ghosh et al. 2009]. Notably, most facial capture systems only capture skin appearance, where eyes and teeth are usually treated separately [Bérard et al. 2014; Li et al. 2022; Velinov et al. 2018; Wu et al. 2016].

More recent work aims to reduce the number of required captures or speed up the capture sequence. Fyffe et al. [2009] fit a cosine-lobe reflectance function using colored gradient illuminations, reducing the necessary number of captures to two images. Similarly, Kampouris et al. [2018] capture facial reflectance with two color-multiplexed binary spherical gradients. On the other hand, Fyffe et al. [2016] uses a setup with multiple DSLR cameras and flashes to achieve near-instant reflectance capture by sequentially triggering the flashes. However, these methods are still fundamentally limited to static captures due to their reliance on time-multiplexed lighting patterns and synchronized hardware.

Dynamic appearance capture: Dynamic photometric captures have relied on high-speed cameras with motion-compensating algorithms [Fyffe et al. 2011; Lattas et al. 2022a; Wenger et al. 2005; Wilson et al. 2010]. One major drawback with these approaches is that the actual capturing framerate is much lower than the cameras employed, since multiple frames are required to produce a single reflectance map. Fyffe and Debevec [2015] proposed a polarized color gradient illumination, which enables single-shot performance photometric facial capture. However, their method requires complex illumination with both parallel and cross-polarized lights, and fails to capture spatially varying roughness.

On the other hand, passive facial acquisition systems do not rely on specifically designed lighting patterns. These methods often require only uniform illumination, acquiring high-quality facial geometry with multiview stereo [Beeler et al. 2010, 2011; Bradley et al. 2010; Cao et al. 2015]. While these methods provide geometry with fine details, they do not measure the spatially varying reflectance of the human face. Reflectance estimation with passive systems is challenging due to difficulties in disambiguating different lighting effects. Gotardo et al. [2018] utilizes inverse rendering with a pre-estimated albedo sub-space to jointly optimize for high-quality surface normal, diffuse albedo, and specular intensity. This method is later improved with the inclusion of polarization to separate the diffuse and specular reflection [Riviere et al. 2020] and with more accurate lighting models [Xu et al. 2022]. These methods do not separate the spatially varying specular albedo and roughness.

More recently, Ha et al. [2024] utilized more sophisticated polarization cues and multispectral imaging to acquire the subsurface scattering properties as well as specular intensity and roughness during performance capture. However, their method requires specialized polarization cameras as well as multi-spectral imaging techniques. Their reflectance parameters are also jointly optimized instead of being separately measured, which gives the risk of converging to a local optimum instead of the ground truth reflectance maps. In our approach, we show that with a proper illumination design, single-shot facial appearance capture can be achieved with only linearly polarized light sources and regular consumer cameras. We capture per-view diffuse albedo, specular albedo, and specular roughness, and only optimize for photometric normals across views.

Deep learning and data-driven methods: Recently, multiple methods have been proposed for the estimation of facial appearance under unconstrained lighting conditions. Dib et al. [2021] uses differentiable rendering to estimate face appearance and environmental lighting based on a pre-learned 3DMM [Paysan et al. 2009]. Rainer et al. [2023] use tiny MLPs as neural shading fields to disambiguate albedo from unknown shading with inverse rendering. Facial appearance has also been directly estimated with deep neural networks [Han et al. 2025; Huynh et al. 2018; Lattas et al. 2022b; Saito et al. 2017]. These work aim to predict plausible appearance maps instead of physically accurate measurements. Additionally, 3D neural representations such as LitNeRF [Sarkar et al. 2023] have shown high-quality facial rendering results. However, the results of such methods are difficult to integrate into traditional rendering pipelines.

3 Capture Setup

Our approach assumes a lighting system capable of emitting linearly polarized RGB illumination. Such configurations are prevalent in light stage setups, which enable effective separation of diffuse and specular reflection components via polarization [Debevec et al. 2000; Ghosh et al. 2011; Kampouris and Ghosh 2018; Ma et al. 2007]. Furthermore, our design is easily realizable using standard LCD monitors, as most displays are inherently linearly polarized. The empirical results presented in subsequent sections were acquired using a practical monitor-based setup shown in Figure 2 (a). The capture apparatus comprises 15 DSLR cameras arranged into five vertical columns, with each column consisting of three vertically stacked cameras. Each camera is equipped with a polarization filter. For each column, the center camera is cross-polarized with respect to the illumination to capture diffuse reflection, while the top and bottom cameras are parallel-polarized to capture the combined diffuse and specular reflections. This polarization-based separation strategy follows prior art, where specular reflections preserve the state of polarization of the incident lights, while diffuse reflections are depolarized [Debevec et al. 2000]. A set of sample photographs that are captured using our setup is shown in Figure 3.

3.1 RGB Sinusoidal Illumination

We illuminate the subject using a sinusoidal lighting pattern that varies with angle and phase along the longitudinal direction, with a fixed angular frequency $f = 3$. Each RGB color channel produces a sinusoidal pattern with a different phase shift:

$$I(\vec{\omega}) = \begin{cases} R: & \cos(3\phi) + 1 \\ G: & \cos(3\phi + 120^\circ) + 1 \\ B: & \cos(3\phi - 120^\circ) + 1, \end{cases} \quad (1)$$

where $\vec{\omega} = (\phi, \theta)$ is the incident lighting direction in spherical coordinates. We chose $f = 3$ as this is the lowest frequency for effective diffuse estimation (see section 4.2), and higher frequencies will complicate the specular estimation. A mirrorball visualization of the colored sinusoidal illumination is shown in Figure 2 (b). This structured illumination allows per-channel encoding of phase-shifted sinusoids, which are critical for frequency-based reflectance separation [Lin et al. 2023].

The observed radiance at the camera is an integral over the spectral power distribution (SPD) of the light source $P(\lambda)$ and the spectral sensitivity of the camera $S_i(\lambda)$. We perform spectral calibration such that each camera channel corresponds uniquely to its respective illumination band. This can be achieved by calculating the color correction matrix $C \in \mathbb{R}^{3 \times 3}$, whose entries $c_{i,j}$ encode the spectral overlap between the i^{th} color channel of the camera sensor and the illumination channel j :

$$c_{i,j} = \int_{\Lambda} P_j(\lambda) S_i(\lambda) d\lambda. \quad (2)$$

Captured RGB values $\vec{c}$ are then corrected via inverse transformation:

$$\vec{c}' = (C^{-1})^T \vec{c}. \quad (3)$$

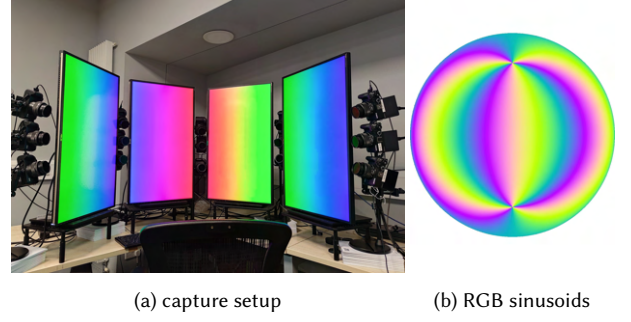


Fig. 2. Experimental results in this paper are obtained with a monitor-based setup similar to that described in Lattas et al. [2022a] (a). Our method utilizes polarized RGB sinusoidal illumination for reflectance capture (b).

In practice, the color correction matrix C is directly measured for each camera using a color chart for better estimation.



Fig. 3. Sample photographs captured with our 15 DSLR cameras. The top and bottom rows are parallel-polarized to the sinusoidal lighting to retain specular responses, while the center row is cross-polarized to isolate the diffuse color.

3.2 Geometry Reconstruction and Preprocessing

To reconstruct base facial geometry, we employ NeuS [Wang et al. 2021] trained on the 15 captured photographs. While the raw RGB images encode photometric information of the geometry, we found that using grayscale images with less lighting variation yield superior geometry, with fewer artifacts.

After geometry reconstruction, we perform UV unwrapping to parameterize the surface in a 2D texture space with minimum distortion [Rabinovich et al. 2017]. The multi-view photographs are projected into this UV space, establishing pixel-wise correspondence across different viewpoints.

For diffuse and specular separation, we first compute a view-independent diffuse color by pooling observations from the five

cross-polarized cameras. Specular-only UV images for the parallel-polarized cameras are then computed by subtracting the diffuse color from each respective camera’s measurement. This results in a shared diffuse color and ten view-dependent specular images encoded in the UV domain.

4 High Quality Appearance Maps

After estimating the base geometry along with diffuse and specular texture maps, we compute high-quality spatially-varying appearance attributes. This section details the estimation of diffuse albedo, specular albedo, specular roughness, and photometric normals, which capture mesoscopic surface details beyond the resolution of the base geometry.

4.1 Preliminaries

Under phase-shifted sinusoidal illuminations $I(\vec{\omega}) = \cos(f\phi + \varphi) + 1$, Lamond et al. [2009] showed that the reflected radiance E is a shifted cosine function that depends on the phase shift φ . They leveraged this result to separate diffuse and specular albedo with three distinct phase-shifted sinusoids.

More recently, Lin et al. [2023] extended this approach by additionally estimating the maximum phase-shift angle $\hat{\varphi}$. Their analysis showed that three phase-shifted observations are sufficient to recover the radiance function E under phase-shifted sinusoidal illumination as:

$$E(\varphi) = \rho + A \cos(\varphi - \hat{\varphi}), \quad (4)$$

where ρ is the specular albedo modulated by Fresnel effects, A the modulation amplitude, and $\hat{\varphi}$ the phase offset indicating the dominant reflection direction.

4.2 Diffuse Albedo Estimation

Assuming a convex surface, diffuse albedo estimation under full-spherical sinusoidal illumination is straightforward since the irradiance for sinusoidal illuminations with odd angular frequencies greater than one averages to zero [Ramamoorthi and Hanrahan 2001]. Applying this principle to Equation (1), the irradiance for each color channel c_i simplifies to:

$$\begin{aligned} c_i &= \int_{\Omega} \frac{\rho_{d,i}}{\pi} [\cos(3\phi + \varphi_i) + 1] \cos \hat{\theta} d\omega \\ &= \rho_{d,i} + \int_{\Omega} \frac{\rho_{d,i}}{\pi} (\cos(3\phi + \varphi_i) \cos \hat{\theta} d\omega) \\ &= \rho_{d,i}, \end{aligned} \quad (5)$$

where $\rho_{d,i}$ denotes the diffuse albedo for channel $i \in \{R, G, B\}$, and $\hat{\theta}$ is the angle between the surface normal and the lighting direction. The second term integrates to zero due to the orthogonality of the sinusoidal basis over the hemisphere. This result, similar to Lamond et al. [Lamond et al. 2009], implies that cross-polarized imagery directly yields the diffuse albedo.

In practice, however, limited zonal illumination—as in our monitor-based setup—cannot emulate a full-spherical field. Under these conditions, diffuse captures are modulated by colored sinusoidal shading. To correct this, we simulate the lighting setup virtually and

render the reconstructed geometry under uniform albedo to compute the shading factor. The corrected diffuse albedo is obtained by dividing the measured cross-polarized color by this estimated shading.

4.3 Specular Albedo and Roughness Estimation

Conventional single-shot methods treat specular roughness estimation as under-constrained [Ha et al. 2024; Riviere et al. 2020]. Here, we resolve this by exploiting two key observations: (1) human skin behaves as a dielectric material with largely achromatic specular reflection; (2) colored sinusoidal illumination encodes three independent sinusoidal patterns via RGB phase shifts $\varphi \in [0, 120^\circ, -120^\circ]$, allowing us to fit a reflected radiance function E for phase-shifted sinusoidal illumination following Lin et al. [2023].

Distinct from Lin et al. [2023], our formulation benefits from polarization-based separation of diffuse and specular components. Consequently, ρ directly measures the specular albedo. For rough surfaces, we empirically observe $A \leq \rho$, with the ratio $\gamma = A/\rho$ dependent on surface roughness σ and the polar angle θ_r of the reflection vector $\vec{r}$.

To recover specular roughness σ , we construct a precomputed 2D lookup table $\sigma(\gamma, \theta_r)$ via rendering a reference sphere under colored sinusoidal lighting with varying roughness. We adopt a coordinate system where the polar angle θ_r ranges between -90 and 90 degrees, with $\theta_r = 0$ when the reflection vector points along the equator. The sphere is rendered with Blender using a micro-facet BSDF with GGX distribution. Figure 4 plots the 2D lookup table at selected reflection polar angles θ_r . The final algorithm proceeds as follows:

- (1) Fit Equation (4) at each pixel to estimate ρ , A , and $\hat{\varphi}$ from the RGB specular images.
- (2) Remove Fresnel gain by compensating for view angle using the estimated surface normals (section 4.4), yielding view-independent specular albedo ρ_s .
- (3) Use the ratio γ and reflection polar angle θ_r to query the lookup table for the corresponding roughness σ .

Figure 5 presents the error map for our roughness estimation, assuming a measurement noise level of approximately 1 bit in an 8-bit image. The estimated roughness remains reliable for ground-truth values in the range $[0.1, 0.7]$, with a wider range at a lower polar reflection angle. In practice, our system records 16-bit raw images with lower sensor noise, so the captured roughness maps exhibit less error compared to the synthetic error bounds.

4.4 Photometric Specular Normal Estimation

In addition to specular albedo and roughness, the recovered phase angle $\hat{\varphi}$ encodes directional information about the specular lobe. Although $\hat{\varphi}$ correlates with the reflection vector, it alone is insufficient to determine the surface normal from a single camera. We present two methods for reconstructing detailed photometric normals, applicable under differing illumination coverage.

Full spherical illumination. Under ideal full-spherical illumination, the subject appears well illuminated across all views, with $\hat{\varphi}$ well measured across all cameras. The photometric normal $\vec{n}$ at each surface point can be estimated by minimizing:

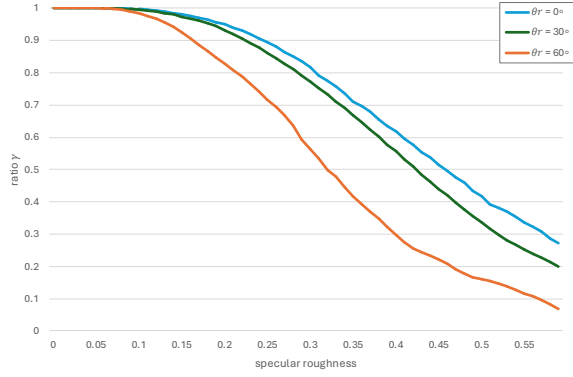


Fig. 4. Plot of the ratio $\gamma = A/\rho$ with respect to different roughness values, under different reflection angle θ_r .

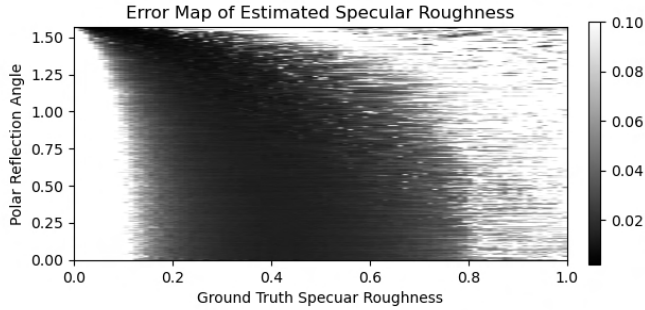


Fig. 5. Error map for roughness estimation as a function of ground-truth roughness and polar reflection angle. The map assumes a measurement error of 0.039 in pixel intensities (≈ 1 bit in an 8-bit image), with errors above 0.1 clipped for visualization. Estimation is generally reliable for roughness values in the range $[0.1, 0.7]$, while highly reflective surfaces exhibit greater variance.

$$\min_{\vec{n}} \sum_c (\Phi(\vec{n}, \vec{v}_c) - \phi_c)^2, \quad (6)$$

where $\vec{v}_c$ and ϕ_c are the view vector and phase angle measured at camera c , and $\Phi(\vec{n}, \vec{v}_c)$ is the theoretical phase shift for a given normal. Solving Equation (6) fuses angular measurements from multiple views, yielding robust, high-fidelity normal estimation.

Limited zonal illumination. Under a limited zonal illumination setup, as illustrated in Figure 2 (a), the estimated phase-shift map $\hat{\phi}$ is compromised due to insufficient angular coverage and near-field lighting artifacts. These limitations degrade the accuracy of multi-view optimization, often resulting in noisy photometric normal maps. To address this, we adopt an alternative approach that directly embosses high-frequency detail from the observed measurements onto the base geometry, yielding significantly cleaner photometric normal reconstructions. Let $\vec{r} = (\phi_r, \theta_r)$ denote the reflection vector map derived from the base geometry and viewing direction, empirical observations reveal two key properties:

- (1) The phase-shift map $\hat{\phi}$ exhibits high-frequency variations that are strongly correlated with those of ϕ_r , the azimuthal component of the reflection vector.
- (2) Due to the limited vertical coverage of the illumination field, the measured specular albedo ρ is attenuated when the reflection vector deviates from the equatorial plane. This falloff can be approximated by:

$$\rho = \rho_s \cos(\theta_r), \quad (7)$$

where ρ_s is the ground-truth specular albedo, and ρ is its observed counterpart under the constrained lighting configuration.

Assuming ρ_s is locally uniform, small-scale variations in ρ can be attributed to local changes in θ_r . Differentiating Equation (7) yields:

$$d\hat{\rho}_s = -\rho_s \sin(\theta_r) d\theta_r, \quad (8)$$

which links the observed high-frequency fluctuations in specular albedo to variations in the polar angle of the reflection vector.

Based on these observations, we estimate the refined photometric normal map for each parallel-polarized view using the following procedure:

- (1) For each camera, compute the initial reflection vector $\vec{r} = (\phi_r, \theta_r)$ from the known view direction and the base surface geometry.
- (2) Enhance ϕ_r by embedding the high-frequency variations of the measured phase-shift map $\hat{\phi}$, producing a refined azimuthal component ϕ'_r .
- (3) Using Equation (8), infer the differential $d\theta_r$ from local variations in ρ , and apply it to obtain the enhanced polar component θ'_r .
- (4) Construct the high-frequency reflection vector $\vec{r}' = (\phi'_r, \theta'_r)$, and compute the corresponding photometric normal as the half vector between the reflection vector and the view vector.

This method effectively recovers high-fidelity normal detail by leveraging phase and intensity cues from the measured appearance, even under spatially limited illumination conditions. Furthermore, we also obtain a corrected specular albedo estimation that accounts for limited zonal illumination using Equation (7).

After estimating the appearance parameters for each camera view—specifically specular albedo and specular roughness, and under limited illumination, also the photometric normal—a texture map is generated by aggregating data from multiple views. Our entire process yields UV texture maps for: (1) diffuse albedo, (2) specular albedo, (3) specular roughness, and (4) photometric normals.

5 Experimental Results

In this section, we present the appearance maps and geometry reconstructed using our method. We also highlight how fine-scale details captured in the photometric normals, spatially-varying specular albedo, and roughness maps contribute to high-fidelity facial renderings.

5.1 Synthetic Data

We begin with a quantitative and qualitative evaluation using synthetic data rendered from the Digital Emily model [Alexander et al. 2010], under distant full-spherical illumination and our monitor-based illumination. Photometric normals are simulated using the provided displacement map, while spatially-varying roughness is emulated by scaling the specular albedo map, resulting in the majority of skin roughness values ranging from 0.3 to 0.5.

We compare our results against those produced by the state-of-the-art single-shot facial capture method introduced by Riviere et al. [2020]. We adhere to their acquisition protocol by illuminating the subject under uniform white light. As no open-source implementation of their method is available, we replicate their pipeline as faithfully as possible, with hyperparameter tuning adapted to our synthetic data. Both methods use identical mesh geometry, isolating the comparison to appearance map estimation.

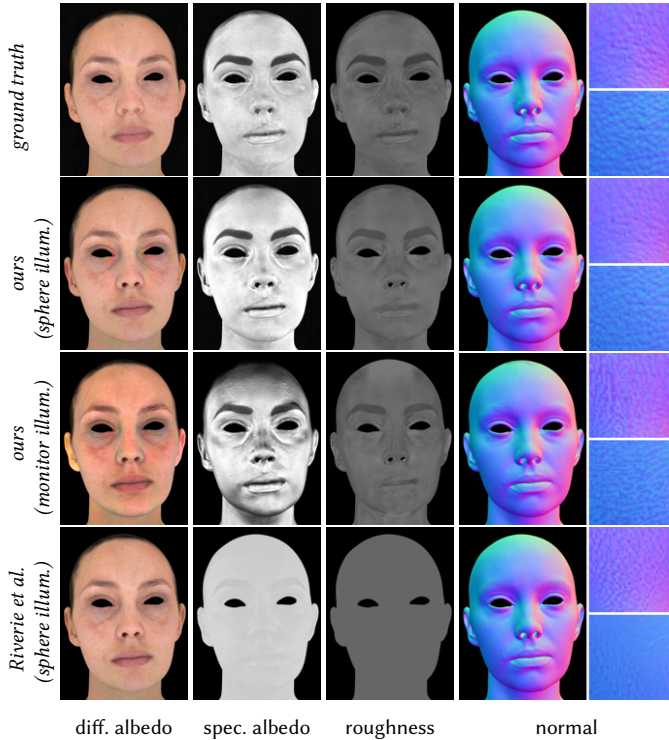


Fig. 6. Comparison between our method and Riviere et al. [2020] with digital Emily rendered under full spherical illumination and our method under monitor-based illumination. Our method better captures the spatially-varying specular albedo and roughness, while having a more photometrically accurate normal map.

Figure 6 shows the results alongside the ground truth. Under full spherical illumination, our reconstructed specular albedo and roughness maps closely match the ground truth. In contrast, Riviere et al.’s method estimates a single specular intensity map without separating the effects of specular roughness and albedo. Although both methods recover high-frequency normal maps, our method

benefits from directly optimizing normals using the measured phase-shift angles, yielding more photometrically accurate results under full spherical lighting. This is in contrast to Riviere et al.’s joint optimization strategy. Quantitative comparisons in Table 1 further show that our captured specular maps and normals are of higher quality.

Under the monitor-based setup with limited illumination, we observe a slight color shift in the estimated diffuse albedo, which we attribute to subsurface scattering effects in skin. Reliable estimation of specular albedo and roughness is restricted to facial regions that receive sufficient incident light. Consequently, areas such as the top of the head and the underside of the chin remain under-illuminated, leading to increased mean square error in the reconstructed appearance maps (see Figure 6 and Table 1). These observations indicate that higher-quality scans can be achieved in practice with a dedicated light stage setup that provides full spherical illumination.

Table 1. Comparison of the average mean square error (MSE) of the acquired appearance maps. While Riviere et al. [2020] obtains better diffuse albedo under white illumination, our specular maps and normals are more accurate, which is essential for realistic renderings. The acquired appearance maps under limited illumination exhibit higher error, as facial area with extreme angles receive insufficient light. Interestingly, the estimation of specular roughness remains more robust in this setting, since it depends on parameter ratios rather than absolute values.

Method	diff. albedo (10^{-3})	spec. albedo (10^{-2})	roughness (10^{-3})	normal (10^{-3})
ours	1.607	2.048	1.212	0.668
Riviere et al.	1.422	2.831	-	1.324
ours (monitor)	6.212	5.767	1.744	2.956

Aside from dedicated studio captures, more recent differentiable rendering methods have estimated reflectance under unknown illumination using neural fields. However, their achieved quality still cannot match that of a dedicated setup. For example, while Rainer et al. [2023] use neural shading fields to separate diffuse and specular albedo from unknown lighting, their specular albedo lacks spatial variation, and their model does not provide an explicit estimate of normals and specular roughness (see Figure 8). Approaches based on statistical face models [Dib et al. 2021] further suffer from identity preservation and loss of details as shown in Figure 7. We do not evaluate more advanced methods that cannot be applied in single-shot captures (e.g. methods that require co-located flash illumination [Han et al. 2024]).

5.2 Real Captures

This section presents the acquired reflectance maps on real subjects with our monitor-based capture setup. Our method produces high-quality appearance maps that encode mesoscopic geometric variations via fine-scale normal maps. Figure 9 shows both the reconstructed results and corresponding reference photographs of the scanned subjects. The photometric normals capture additional high-frequency lighting effects, such as skin pores and fine wrinkles, which are not represented in the base geometry. This leads to more realistic rendering of skin microstructure. Although we do not

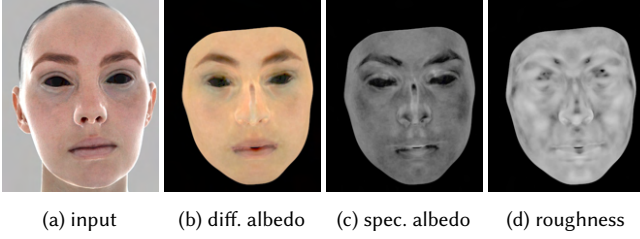


Fig. 7. Face reflectance maps obtained using NextFace [Dib et al. 2021]. The result is obtained using 15 rendered images from the same viewpoints as our setup, with full-on spherical lighting. While statistical differential rendering methods are robust to unknown illumination, the results are sub-optimal compared to dedicated capture methods.



Fig. 8. While Rainer et al. [2023]’s neural shading field estimates diffuse and specular albedo, their specular albedo lacks spatial variation, and their model does not provide explicit estimate of normals and specular roughness (image re-produced from Rainer et al. [2023] for comparison purposes since their code is not available online).

present performance capture results due to current hardware limitations, our method can be directly applied to dynamic sequences by independently processing each video frame.

Our separated specular albedo and roughness maps enable BRDF estimation of non-skin facial features such as teeth and eyes—an area often neglected in prior work. Figure 11 illustrates the captured spatially varying roughness obtained with our method. Our method thus can support facial appearance acquisition beyond skin-only regions. Additional results demonstrating generalization across a range of subjects with varying facial expressions, ethnicities, ages, and genders are shown in Figure 10 and in supplemental material.

For validation, we compare the rendered results of the subject under novel illumination and compare the rendering with the ground truth photographs captured with our monitor setup. Sample results are shown in Figure 12. Our specular renderings show details that closely match those of a real photograph. However, consistent with our synthetic analysis, we observe a slight change in diffuse color with our limited zonal illumination setup. We note that this has a minimal impact on the visual quality of the renderings, and the color shift will not appear in captures with a full spherical illumination setup, as demonstrated in section 5.1.

Figure 13 presents a visual comparison between our method and that of Riviere et al. [2020], using data captured with our monitor-based illumination setup. Our approach produces sharper normals

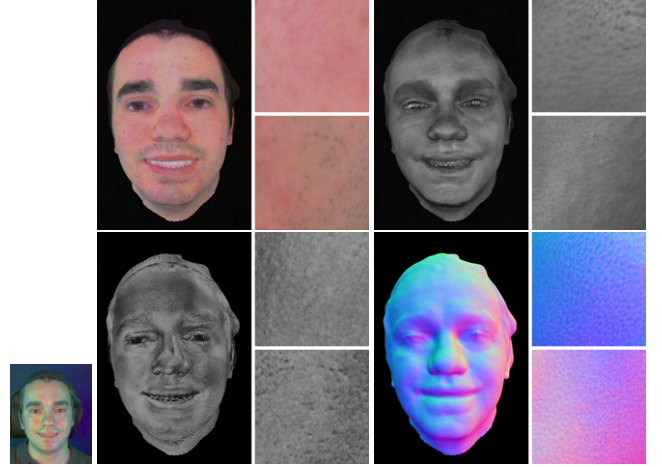


Fig. 9. Our obtained high-quality appearance maps. *Left most*: reference photograph, *Top left*: diffuse albedo, *Top right*: specular albedo, *Bottom middle*: specular roughness, *Bottom right*: photometric normal. Zoomed-in patches show the captured fine detail, including mesoscopic structures in the normal and skin pores in the specular maps.

and specular maps. This is due to our specular phase-based analysis with colored sinusoids, enabling separate estimation of the reflectance maps. In contrast, jointly optimizing reflectance maps through inverse rendering yields entangled maps that are suboptimal.

Additionally, we observe that our re-implementation of Riviere et al.’s method, when applied to data captured using our monitor-based setup, produces softer geometry than the results reported in their original publication. This discrepancy persists even after extensive tuning of regularization parameters. We attribute this to the near-field lighting effects inherent to our monitor-based setup, which are not adequately modeled in their formulation.

5.3 Ablation Study

Figure 14 shows the obtained photometric normal and specular albedo maps with direct optimization and normal embossing. The directly optimized normal for captures under near-field limited zonal illumination produces noisy results, while the specular albedo exhibits lighting variation caused by the limited illumination. In contrast, Our approach—based on embossing high-frequency features—yields photometric normals that retain fine details and exhibit significantly less noise, and our corrected specular albedo has less shading. These embossed maps enables realistic rendering with less noise and biases.

5.4 Limitations

Despite its effectiveness, our method has limitations. Most notably, our method has slight color shifts in the diffuse albedo when employing limited zonal illumination. This issue does not arise in full spherical illumination settings. We do not explicitly model subsurface scattering effects. However, our captured data—comprising triplets of one cross-polarized image and two parallel-polarized

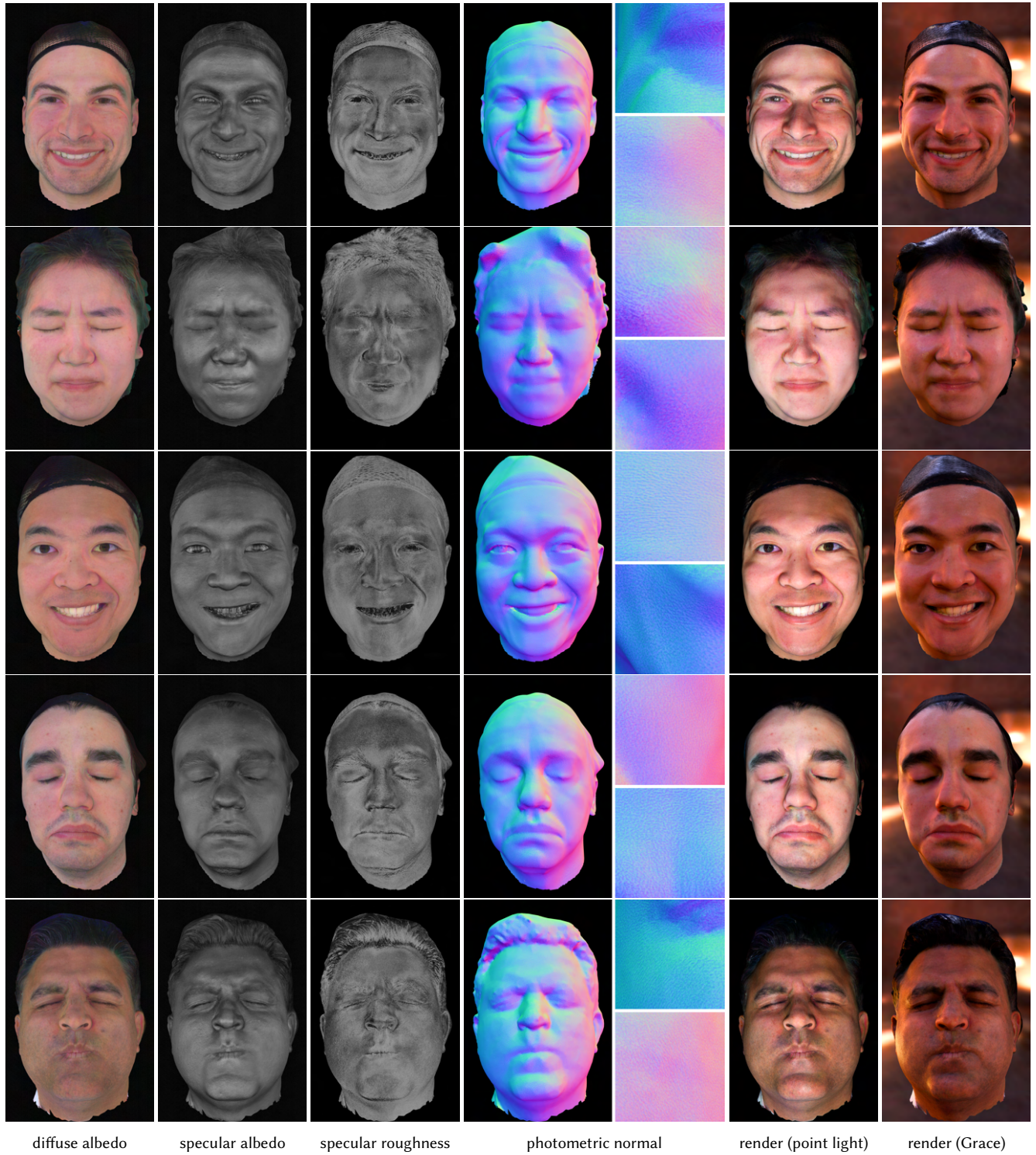


Fig. 10. Our single-shot method captures high-quality appearance maps including: diffuse albedo, specular albedo, specular roughness, and photometric normal with mesoscopic surface detail. The appearance maps enable realistic renderings of skin, eye, and teeth.

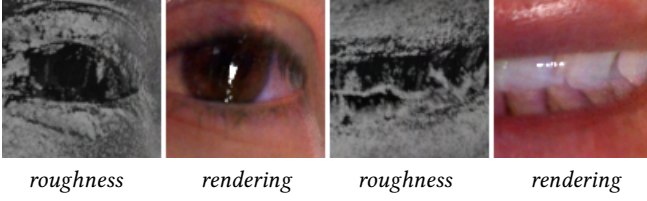


Fig. 11. Our method decouples the estimation of specular albedo and roughness, allowing independent estimation of the spatially-varying roughness map. This enables appearance estimation of non-skin area within the same capture system. Here we show a zoomed-in view of our measured specular roughness and the resulting renderings with spatially-varying specular roughness.

images-can serve as input to inverse rendering methods such as Riviere et al. [2020]. Integrating such methods into our framework remains an avenue for future work. Limited brightness of our setup’s screen illumination also limits us currently to static expression scans. However, the method should naturally be applicable for dynamic capture in brighter illumination setups.

6 Conclusion

We present a novel single-shot method for capturing high-quality facial appearance, leveraging sinusoidal illumination and polarization-based separation to recover spatially-varying diffuse albedo, specular albedo, specular roughness, and photometric specular normals. Our approach requires only linearly polarized light sources and cameras, making it highly practical and readily deployable using commodity hardware such as LCD monitors and DSLRs.

By encoding phase-shifted sinusoids across color channels, our system effectively disentangles reflectance properties in a single capture per view. Coupled with geometry reconstruction and UV-based reflectance pooling, we achieve dense and detailed appearance maps that rival those from traditional multi-shot active illumination systems, while remaining orders of magnitude more efficient in acquisition time and setup complexity.

In summary, this work highlights the feasibility of compact, high-quality facial appearance capture using structured RGB lighting and polarization cues, opening new possibilities for practical deployment in consumer-grade systems and dynamic settings.

Acknowledgments

This research was partly supported by the EPSRC IAA Grant EP/R511547/1. We gratefully acknowledge the contributions of all participants involved in the data acquisition process.



Fig. 12. We validate our method by rendering the captured subject under white illumination and compare the rendering to ground truth photographs. Our specular renders exhibits high frequency details that closely match that of the real photograph. While we observe a shift in color-tone for the diffuse albedo, this is an effect caused by the near-field zonal illumination and will not appear under full-spherical illumination.

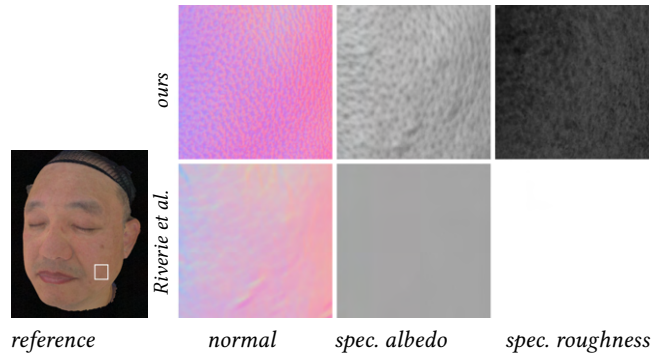


Fig. 13. Our method produces appearance maps that better capture the fine-details of skin-pores compared to Riverie et al. [2020]. Images from left to right: reference image, normal, specular albedo, specular roughness.

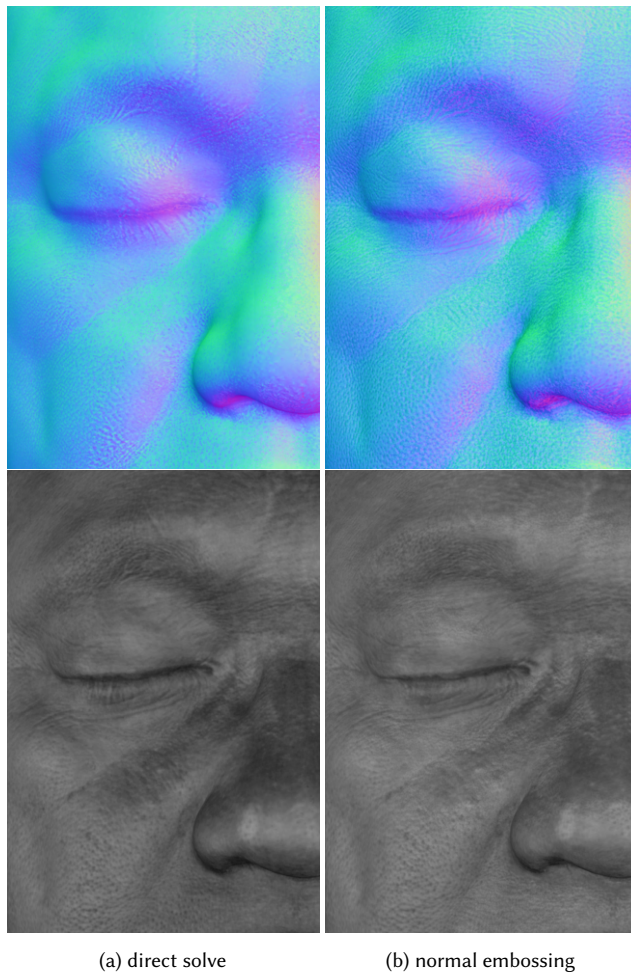


Fig. 14. Under limited zonal illumination, our embossing approach generates less noisy normals compared to direct optimization. Our corrected specular albedo also exhibits less shading that is visible in the directly captured specular normal.

References

- 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. *IEEE Computer Graphics and Applications* 30, 4 (2010), 20–31. doi:10.1109/MCG.2010.65
- Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, and Markus Gross. 2010. High-quality single-shot capture of facial geometry. *ACM Trans. Graph.* 29, 4, Article 40 (July 2010), 9 pages. doi:10.1145/1778765.1778777
- Thabo Beeler, Fabian Hahn, Derek Bradley, Bernd Bickel, Paul Beardsley, Craig Gotsman, Robert W. Sumner, and Markus Gross. 2011. High-quality passive facial performance capture using anchor frames. *ACM Trans. Graph.* 30, 4, Article 75 (July 2011), 10 pages. doi:10.1145/2010324.1964970
- Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. 2014. High-quality capture of eyes. *ACM Trans. Graph.* 33, 6 (2014), 223–1.
- Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. 2021. Deep relightable appearance models for animatable faces. *ACM Transactions on Graphics (ToG)* 40, 4 (2021), 1–15.
- Sai Bi, Zexiang Xu, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. 2020. Deep reflectance volumes: Relightable reconstructions from multi-view photometric images. In *Proceedings of the 27th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*. Springer, 294–311.
- Derek Bradley, Wolfgang Heidrich, Tiberiu Popa, and Alla Sheffer. 2010. High resolution passive facial performance capture. In *ACM SIGGRAPH 2010 papers*. 1–10.
- Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time high-fidelity facial performance capture. *ACM Trans. Graph.* 34, 4, Article 46 (July 2015), 9 pages. doi:10.1145/2766943
- 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 (SIGGRAPH '00)*. ACM Press/Addison-Wesley Publishing Co., USA, 145–156. doi:10.1145/344779.344855
- Abdallah Dib, Gaurav Bharaj, Junghyun Ahn, Cédric Thébaud, Philippe Gosselin, Marco Romeo, and Louis Chevallier. 2021. Practical face reconstruction via differentiable ray tracing. In *Computer Graphics Forum*, Vol. 40. Wiley Online Library, 153–164.
- Graham Fyffe and Paul Debevec. 2015. Single-Shot Reflectance Measurement from Polarized Color Gradient Illumination. In *2015 IEEE International Conference on Computational Photography (ICCP)*. 1–10. doi:10.1109/ICCPHOT.2015.7168375
- Graham Fyffe, Paul Graham, Borom Tunwattapong, Abhijeet Ghosh, and Paul Debevec. 2016. Near-Instant Capture of High-Resolution Facial Geometry and Reflectance. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 353–363.
- Graham Fyffe, Tim Hawkins, Chris Watts, Wan-Chun Ma, and Paul Debevec. 2011. Comprehensive facial performance capture. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 425–434.
- Graham Fyffe, Cyrus A. Wilson, and Paul Debevec. 2009. Cosine Lobe Based Relighting from Gradient Illumination Photographs. In *2009 Conference for Visual Media Production*. 100–108. doi:10.1109/CVMP.2009.18
- Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. 2009. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Proceedings of the Twentieth Eurographics Conference on Rendering (Girona, Spain) (EGSR'09)*. Eurographics Association, Goslar, DEU, 1161–1170. doi:10.1111/j.1467-8659.2009.01493.x
- Abhijeet Ghosh, Graham Fyffe, Borom Tunwattapong, Jay Busch, Xueming Yu, and Paul Debevec. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Trans. Graph.* 30, 6 (Dec. 2011), 1–10. doi:10.1145/2070781.2024163
- Yuliya Gitlina, Giuseppe Claudio Guarnera, Daljit Singh Dhillon, Jan Hansen, Alexander Lattas, Dinesh Pai, and Abhijeet Ghosh. 2020. Practical measurement and reconstruction of spectral skin reflectance. In *Computer graphics forum*, Vol. 39. Wiley Online Library, 75–89.
- 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 (Dec. 2018), 13 pages. doi:10.1145/3272127.3275073
- Hyunho Ha, Inseung Hwang, Nestor Monzon, Jaemin Cho, Donggun Kim, Seung-Hwan Baek, Adolfo Muñoz, Diego Gutierrez, and Min H. Kim. 2024. Polarimetric BSSRDF Acquisition of Dynamic Faces. *ACM Trans. Graph.* 43, 6, Article 275 (Nov. 2024), 11 pages. doi:10.1145/3687767
- Yuxuan Han, Junfeng Lyu, Kuan Sheng, Minghao Que, Qixuan Zhang, Lan Xu, and Feng Xu. 2025. Facial Appearance Capture at Home with Patch-Level Reflectance Prior. *ACM Transactions on Graphics (TOG)* 44, 4 (2025), 1–16.
- Yuxuan Han, Junfeng Lyu, and Feng Xu. 2024. High-Quality Facial Geometry and Appearance Capture at Home. *CVPR*.
- 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8407–8416.

- Christos Kampouris and Abhijeet Ghosh. 2018. ICL multispectral light stage: building a versatile LED sphere with off-the-shelf components. In *Proceedings of the Eurographics 2018 Workshop on Material Appearance Modeling* (Karlsruhe, Germany) (EG MAM '18). Eurographics Association, Goslar, DEU, 1–4. doi:10.2312/mam.20181190
- Christos Kampouris, Stefanos Zafeiriou, and Abhijeet Ghosh. 2018. Diffuse-specular separation using binary spherical gradient illumination. In *Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations* (Karlsruhe, Germany) (SR '18). Eurographics Association, Goslar, DEU, 1–10. doi:10.2312/sre.20181167
- Bruce Lamond, Pieter Peers, Abhijeet Ghosh, and Paul Debevec. 2009. Image-based separation of diffuse and specular reflections using environmental structured illumination. In *IEEE International Conference on Computational Photography*. IEEE San Francisco, CA, 1–8.
- Alexandros Lattas, Yiming Lin, Jayanth Kannan, Ekin Ozturk, Luca Filipi, Giuseppe Claudio Guarnera, Gaurav Chawla, and Abhijeet Ghosh. 2022a. Practical and Scalable Desktop-Based High-Quality Facial Capture. In *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI* (Tel Aviv, Israel). Springer-Verlag, Berlin, Heidelberg, 522–537. doi:10.1007/978-3-031-20068-7_30
- Alexandros Lattas, Stylianos Moschoglou, Stylianos Ploumpis, Baris Gecer, Abhijeet Ghosh, and Stefanos Zafeiriou. 2022b. AvatarMe++: Facial Shape and BRDF Inference With Photorealistic Rendering-Aware GANs. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 44, 12 (Dec. 2022), 9269–9284. doi:10.1109/TPAMI.2021.3125598
- Gengyan Li, Abhimitra Meka, Franziska Mueller, Marcel C Buehler, Otmar Hilliges, and Thabo Beeler. 2022. Eyenerf: a hybrid representation for photorealistic synthesis, animation and relighting of human eyes. *ACM Transactions on Graphics (ToG)* 41, 4 (2022), 1–16.
- Xiaohui Li, Giuseppe Claudio Guarnera, Arvin Lin, and Abhijeet Ghosh. 2024. Practical measurement and neural encoding of hyperspectral skin reflectance. In *2024 International Conference on 3D Vision (3DV)*. IEEE, 1301–1309.
- Arvin Lin, Yiming Lin, and Abhijeet Ghosh. 2023. Practical acquisition of shape and plausible appearance of reflective and translucent objects. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, e14889.
- 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 *Rendering Techniques*, Jan Kautz and Sumanta Pattanaik (Eds.). The Eurographics Association. doi:10.2312/EGWR/EGSR07/183-194
- Pascal Paysan, Reinhard Kothne, Brian Amberg, Sami Romdhani, and Thomas Vetter. 2009. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 296–301. doi:10.1109/AVSS.2009.58
- Michael Rabinovich, Roi Poranne, Daniele Panozzo, and Olga Sorkine-Hornung. 2017. Scalable Locally Injective Mappings. *ACM Trans. Graph.* 36, 2, Article 16 (April 2017), 16 pages. doi:10.1145/2983621
- Gilles Rainer, Lewis Bridgeman, and Abhijeet Ghosh. 2023. Neural shading fields for efficient facial inverse rendering. In *Computer Graphics Forum*, Vol. 42. Wiley Online Library, e14943.
- Ravi Ramamoorthi and Pat Hanrahan. 2001. On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object. *Journal of the Optical Society of America A* 18, 10 (2001), 2448–2459.
- Jérémy Riviere, Paulo Gotardo, Derek Bradley, Abhijeet Ghosh, and Thabo Beeler. 2020. Single-shot high-quality facial geometry and skin appearance capture. *ACM Trans. Graph.* 39, 4, Article 81 (Aug. 2020), 12 pages. doi:10.1145/3386569.3392464
- Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 130–141. doi:10.1109/CVPR52733.2024.00021
- Shunsuke Saito, Lingyu Wei, Liwen Hu, Koki Nagano, and Hao Li. 2017. Photorealistic facial texture inference using deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5144–5153.
- Kripasindhu Sarkar, Marcel C. Buehler, Gengyan Li, Daoye Wang, Delio Vicini, Jérémy Riviere, Yinda Zhang, Sergio Orts-Escolano, Paulo Gotardo, Thabo Beeler, and Abhimitra Meka. 2023. LitNeRF: Intrinsic Radiance Decomposition for High-Quality View Synthesis and Relighting of Faces. In *SIGGRAPH Asia 2023 Conference Papers* (Sydney, NSW, Australia) (SA '23). Association for Computing Machinery, New York, NY, USA, Article 42, 11 pages. doi:10.1145/3610548.3618210
- Zdravko Velinov, Marios Papas, Derek Bradley, Paulo Gotardo, Parsa Mirdehghan, Steve Marschner, Jan Novák, and Thabo Beeler. 2018. Appearance capture and modeling of human teeth. *ACM Transactions on Graphics (ToG)* 37, 6 (2018), 1–13.
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. *arXiv preprint arXiv:2106.10689* (2021).
- Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. 2005. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 756–764.
- 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), 1–11.
- Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus H Gross, and Thabo Beeler. 2016. Model-based teeth reconstruction. *ACM Trans. Graph.* 35, 6 (2016), 220–1.
- Yingyan Xu, Jérémy Riviere, Gaspard Zoss, Prashanth Chandran, Derek Bradley, and Paulo Gotardo. 2022. Improved lighting models for facial appearance capture. *EG 2022-Short Papers* (2022), 5–8.