Face Reflectance and Geometry Modeling via Differentiable Ray Tracing

Abdallah Dib\textsuperscript{1}, Gaurav Bharaj\textsuperscript{2}, Junghyun Ahn\textsuperscript{3}, Cédric Thébault\textsuperscript{4}, Philippe-Henri Gosselin\textsuperscript{1}, Louis Chevallier\textsuperscript{1}

\textsuperscript{1} InterDigital, Rennes, France
\textsuperscript{2} Technicolor, New York, NY

Introduction. A classic problem in computer vision is to reconstruct 3D faces from monocular image and video data. This needs disentangling scene and face parameters such as light, facial geometry (pose, identity, and expression), and reflectance, while accounting for camera depth ambiguity. Image pixel saturation due to self-shadows and specularity further add to the problem complexity.

To model the problem, optimization based \cite{2} and more recently deep learning based \cite{6} methods have been proposed. A photo-consistency loss difference between a parametric 3D face render and input facial image helps estimate the scene and face parameters. Often, the optimization cost is regularized with statistical priors and mathematical models of the parameters. \cite{7} train a deep neural net to reconstruct faces and rely on high-quality training data obtained via a photometric camera rig. Both \cite{7} and \cite{3} present reflectance models with diffuse and specular albedo. They do not handle self-shadows explicitly.

We present a novel method to model 3D faces from monocular images, with parameterized ray traced scene lights(s), and explicit disentangling of facial geometry (pose, identity and expression), reflectance (diffuse and specular albedo), and self-shadows (see Figure 2). We take inspiration from \cite{1} and model the scene light as a virtual light stage with pre-oriented area lights. This setup is used in conjunction with a differentiable Monte-Carlo ray tracer \cite{4}, to optimize the scene and face parameters. Due to the disentangling of the parameters, we can not only obtain robust results, but also gain explicit control over them, with several practical applications. For example, we can change facial expression with accurate resultant self-shadows or relight the scene and obtain accurate specular reflection.

Face Modeling. Similar to \cite{2}, our method follows an optimization-based face reconstruction methodology. \cite{2, 6} model scene light via spherical-harmonics, and reflectance via the Lambertian (diffuse albedo only) model. Consequently, they cannot handle self-shadows and specularities. Image pixels with saturated color values due to self-shadows or specularity can lead to an underconstrained photo-consistency loss based optimization. Here, face geometry vertices corresponding to saturated pixel color values can deform the geometry unnaturally. We model reflectance via the Cook-Torrance BRDF (diffuse and specular albedo) \cite{5}, and use physically accurate ray tracing to handle parameterized area lights and resultant self-shadows.

In order to render physically accurate images via ray tracing, scene geometry and light must be known. To this end, we create a parameterized virtual light stage with \( N \) fixed oriented (pointing towards origin) area lights placed on vertex positions of an isotropically sampled hemisphere (see Figure 1). Parameterized face geometry is placed at the light stage’s origin, facing the camera. We then employ \cite{4}’s differentiable ray tracing method to render an image. Finally, during face reconstruction optimization, backpropagation automatically calculates the gradient for the various parameters w.r.t photo-consistency cost loss between the input and rendered images.

The scene light parameters are the color intensities \( \{ I_i \} \in \mathbb{R}^{20} \) and \( \mathbf{I} \in \mathbb{R}^{3} \). A light-off (Fig 1) implies intensity \( 0^3 \). The face geometry parameters are \( \alpha, \beta, \gamma \in \mathbb{R}^{80} \) which control a 3D model morphable based facial identity. \( \mathbf{R} \in \mathbb{R}^{20} \) controls facial bend shape expressions, and \( \mathbf{F} \in \mathbb{R}^{20} \) the geometry transform. The skin reflectance model is given by Cook-Torrance reflectance \( f_r = k_d \text{Lambertian}(\mathbf{I}, \mathbf{R}) + k_s \text{Cook-Torrance}(\mathbf{I}, \mathbf{Spec}, \mathbf{N}) \). Here, \( \delta, \mathbf{R} \in \mathbb{R}^{32} \) are the morphable model’s vertex color albedo, \( \mathbf{R}(\alpha, \beta, \gamma, \mathbf{t}) \) the normal vector and \( \mathbf{Spec} \in \mathbb{R}^{32} \) are the pixel values of a vectorized gray-scale specular texture map.

Optimization. The photo-consistency optimization is given by:

\[
\text{argmin} \quad E_{\text{data}} + w_1 E_{\text{prior}} + w_2 E_{\text{light}},
\]

where \( E_{\text{data}} = \sum_{i \in \mathcal{I}} ||\mathbf{p}_i^T - \mathbf{p}_i^\rho||_2^2 \) and \( E_{\text{light}} = \sum_{j} ||\mathbf{I}_j - \mathbf{m}_j||_2^2 \). The vectors \( \mathbf{p}_i^T, \mathbf{p}_i^\rho \) are ray traced and real image (\( \mathcal{I} \)) pixel color values, respectively.