

房蔚. 基于各向异性球面高斯分布的三维人脸渲染方法 [J]. 智能计算机与应用, 2026, 16(1): 133-141. DOI: 10.20169/j.issn.2095-2163.24032101

基于各向异性球面高斯分布的三维人脸渲染方法

房蔚

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 自神经辐射场被提出以来, 从多视图图像中渲染三维人脸已经取得了巨大进展。然而, 由于漫反射和镜面反射率是随空间变化的, 且人脸面部几何特征较复杂, 如何有效地恢复三维人脸已经成为一个具有挑战性任务。针对这些问题, 提出了一种具有各向异性反射率建模的可微逆渲染模型。该算法使用了各向异性球面高斯(ASGs)混合函数来表示环境照明和双向反射分布函数(BRDFs)。首先, 为了估计渲染方程中的镜面反射积分项, 在特征空间中对其进行编码, 有效地解决了直接预测积分结果会忽略丰富信息的问题。其次, 对于漫反射积分项, 将其视为一个球面高斯(SG)和一个ASG的卷积。最后, 使用有符号距离函数(Signed Distance Function, SDF)作为形状表示。实验结果表明, 所提出算法相较于现有人脸重建算法具有更优越的性能。

关键词: 神经辐射场; 三维人脸渲染; 各向异性球面高斯分布; 有符号距离函数

中图分类号: TP391

文献标志码: A

文章编号: 2095-2163(2026)01-0133-09

3D neural face rendering with Anisotropic Spherical Gaussians mixtures

FANG Wei

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Since the advent of neural radiance fields, significant progress has been made in rendering 3D faces from multi-view images. However, effectively recovering 3D faces remains a challenging task due to the complex facial geometry and the difficult-to-model spatially varying specular and diffuse reflectance. To address these issues, this paper proposes a differentiable inverse rendering model with anisotropic reflectance modeling. The framework employs Anisotropic Spherical Gaussian mixtures(ASGs) to represent environmental illumination and Bidirectional Reflectance Distribution Function(BRDF). Firstly, to estimate the specular integral term in the rendering equation, a feature space encoding method is introduced, effectively addressing the problem of neglecting rich information when directly predicting integral results. Then, for the diffuse integral term, this paper treats it as a convolution of an SG (Spherical Gaussian) and an ASG. Finally, a well-established method is utilized to represent facial geometry as a signed distance function. Experimental results demonstrate the superior performance of the proposed approach compared to state-of-the-art methods.

Key words: neural radiance fields; 3D face rendering; Anisotropic Spherical Gaussian mixtures; Signed Distance Function

0 引言

三维成像和计算机图形学的快速发展, 引发了人们对实现精确且高效的三维人脸重建和渲染的极大兴趣。这项研究的应用有着巨大潜力和价值, 包括增强现实、虚拟现实体验、电话会议、游戏和生物识别^[1-4]。然而, 人脸固有的独特性, 以复杂的几何形状、复杂的纹理变化、以及复杂的镜面和漫反射为特征, 使其成为该领域备受关注的热点挑战之一。

解决这些挑战需要创新的方法来捕捉微妙的细节, 并忠实地重建人脸的几何形状和外观。

本文的工作集中在人脸几何建模和纹理研究。人脸几何建模的目的是生成人脸网格的顶点的位置。在已有的研究中, 最初的方法依赖于线性模型^[5], 后来扩展到多线性模型^[6], 就是分别编码不同的模式。然而, 由于线性模型的有限表示, 很难得到复杂表示, 如夸张的表情。在深度学习中, 研究人员使用卷积神经网络或图神经网络探索了二维图

基金项目: 国家自然科学基金(62273239)。

作者简介: 房蔚(1999—), 男, 硕士研究生, 主要研究方向: 三维人脸重建。Email: 212240437@st.usst.edu.cn。

收稿日期: 2024-03-21

像^[7]或三维网格^[8]的非线性模型,实现了性能的显著提升。然而,离散数据表示限制了分辨率,不能充分地捕捉面部先验,并由此造成形状细节损失。最近,一些研究^[9]表明,连续深度隐式函数能够实现精确的三维几何结构建模,优于体素、点云和网格等离散表示,并且还进一步观察到了其在形状重建^[10]和表面配准^[11]方面的好处。这些发现突出了神经隐式函数在推进三维面部重建、减少重复和提高模型保真度方面的潜力。受上述方法的启发,本文的模型也采用了深度隐式函数来建模面部几何细节。

最近,可微渲染器在通过照明等因素来识别图像和几何图形之间的联系方面取得了进步,并可应用于人脸渲染^[12]。Blinn-Phong 阴影模型^[13]虽然解释了镜面反射,但缺乏真正的物理基础,通常被视为一种经验方法。于是,这将前沿推向了基于物理的渲染(PBR)^[13],其中 BRDFs^[14]作为模拟环境光和场景之间交互的整体方法得到了广泛应用。尽管 PBR 取得了显著的进展,但在捕捉面部皮肤微妙的光学特性方面仍存在问题,由于面部皮肤其半透明的性质和亚表面散射的相互作用,将很难用目前的许多模型准确描述。

随着神经辐射场(NeRFs)^[15]的引入,神经渲染领域的成果也不断涌现。最初提出的 NeRF 已经彻底改变了 3D 场景表示和渲染领域。这些方法^[16]将一个场景表示为一个连续的 5D 函数,将三维空间位置映射到 RGB 的颜色和密度值,从而渲染场景的新视图。然而,却仍然有 2D 到 3D 的模糊性问题。此外,在基于 NeRFs 的模型中,传统的球面高斯(SG)^[17]表示环境照明和双向反射率分布函数,也显示出了局限性,特别是在捕捉高频细节方面。这推动了向各向异性球面高斯(ASG)^[18]的转变,旨在克服 SG 的不足。

面临上述挑战和需求,本文构建了一个基于 ASGs 混合函数的物理渲染模型,该模型可以联合表示多视图面部图像的几何形状和复杂的皮肤反射率。具体地说,采用有符号的距离函数(SDFs)作为形状表示的一种手段。本文框架的一个关键研发涉及到利用 ASGs 来近似照明和 BRDF。与 SG 相比,ASG 能够有效地捕捉各向异性反射,能够细粒度、精确建模面部复杂特征细节,特别是在面对不同的光照条件和视角时。此外,渲染方程的精确解计算复杂度,如蒙特卡罗采样^[19]和积分。在本研究中,分别通过估计渲染方程的镜面积分项和扩散积分项来降低计算复杂度,从而得到一个估计的输出

辐射。此外,对于多层面部皮肤的情况,采用体积渲染技术来实现最终累积颜色值的合成。

1 相关工作

1.1 隐式神经表示

近年来,隐式神经函数在三维视觉应用中取得了不错的效果,因其具有良好的离散数据存储能力和自身的高阶导数性质,现已成为一种连续、而非离散的预测方法。SIRENs^[20]使用周期激活函数进行隐式神经表征,目前已被证明在捕获复杂的自然信号方面是非常有效的。同时,也已证明多层感知器(MLP)可以通过一个简单的傅里叶特征图传递输入点来学习低维向量中的高频函数,但却可能会引入噪声。此外,MLP 还可以通过将三维坐标映射到相应的有符号距离场(SDF)^[21]来表示三维形状。但是,相邻的坐标点将被映射到相似的函数值中,这使其很难捕获具有足够平滑度的高频细节。Wang 等学者^[22]提出了一种新的位置编码方案,称为样条位置编码,该方案提出在将坐标输入 MLP 之前,将其映射到高维空间。为了从有噪声的点云中提取一个更好的表面,Ma 等学者^[23]提出通过训练一个神经网络,使用预测的 SDF 和查询位置的梯度将查询点变换至表面上最近的点。然而,由于三维点云的离散性或在以往的工作中缺乏多视图图像的先验,结果仍然不确定。Ma 等学者^[24]引入了一个基于神经拉的水平集对齐损失,将所有水平集对齐到零水平集,这显著提高了 SDF 预测的精度,但是对于人脸复杂的几何特征,所提方法能否有效地恢复人脸表面尚未得到仿真验证。

1.2 几何细节恢复

人脸几何重建中最经典的方法之一是三维变形模型(3DMM)^[27],模型假设每个三维人脸都可以由数据集中所有人脸组成的基向量空间来表示。Romdhani 等学者^[26]基于图像特征对 3DMM 进行了优化,以提高图像的精度和鲁棒性。然而,在之前的工作中,人工 3D 模型效果很好,但需要大量的人工工作,一般的 3D 模型效果很容易与图像匹配,但缺乏复杂的细节。为了解除这些限制,一个参数模型 FLAME^[27]集成了 3 个数据库来弥补数据缺陷,并分别表示了人脸的身份、姿势和表情。有时线性 3DMM 拟合不能产生令人满意的结果,为了克服这一问题,Luan 等学者^[28]提出了一种利用编码器来估计投影、形状和纹理参数的非线性方法,将这些参数解码为三维点和纹理贴图,随后计算了渲染图像

与原始图像之间的重建损失。最近,一种基于 SDF 编码的非线性隐式表示 3DMM 模型^[29]通过 2 个解纠缠变形场进行人脸身份和表达式变形,增强了模型捕捉人脸表情的能力。除了基于图像的方法外,也可以从视频序列中对三维面孔进行重建。Bindita 等学者^[30]通过设计 2 个网络对身份相关信息和不相关信息分别进行预测,来学习不同表情下的动态形状特征和纹理特征。3DMM 方法通常假设光照条

件是恒定的,这意味着对光照条件的变化很敏感,当面部处于不同的光照条件时,可能无法准确地重建面部的纹理。与之前的工作相比,本文的方法可以在模拟照明条件下获得更真实的效果。

2 方法

本文提出了 ASG-Face 算法(如图 1 所示),这是一个基于 ASG 混合函数的端到端物理渲染模型。

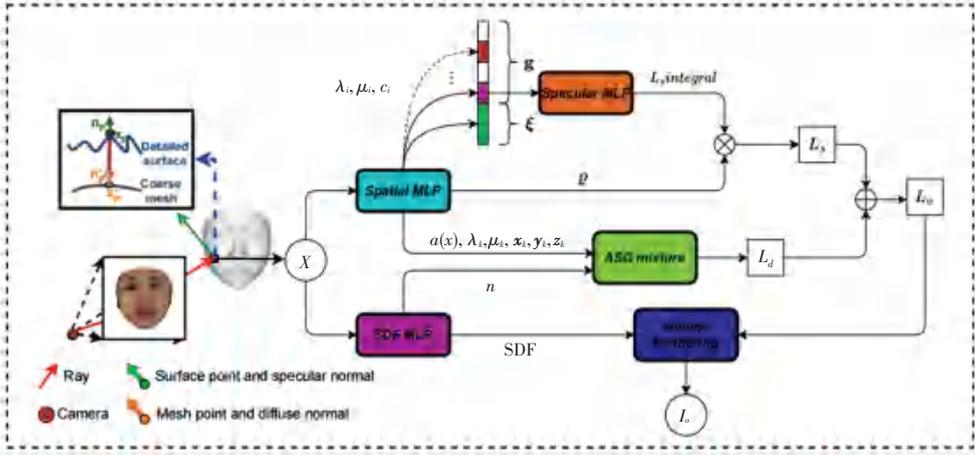


图 1 ASG-Face 总体框架

Fig. 1 Overall framework of ASG-Face

2.1 渲染方程理论

在介绍本文的方法之前,需要探讨一下在照明模型中基于物理的分析模型 BRDF。事实上,当人们看到一个面部表面时,来自周围环境的光线会照射到脸上,同时这个表面会将部分光线反射到观察者的眼睛里。BRDF $f_r(\omega_o, \omega_i; \mathbf{x})$ 描述了入射光与表面反射光的关系,表示反射方向的辐射 $L_o(\omega_o; \mathbf{x})$ 与入射光 E_i 在入射方向的辐照度的比值:

$$f_r(\omega_o, \omega_i; \mathbf{x}) = \frac{dL_o(\omega_o; \mathbf{x})}{dE_i} = \frac{dL_o(\omega_o; \mathbf{x})}{L_i(\omega_i; \mathbf{x}) \cos(\omega_i; \mathbf{n}) d\omega_i} \quad (1)$$

其中, ω_o 表示观察方向; $L_i(\omega_i; \mathbf{x})$ 表示沿 ω_i 方向的入射光线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ 中采样得到的位置 \mathbf{x} 处的入射光强; $\cos(\omega_i; \mathbf{n})$ 表示 ω_i 与表面法线 \mathbf{n} 之间的夹角的余弦。因此,本文得到了渲染方程,即环境光下反射光的亮度,可以写成不同角度入射光的亮度乘以 BRDF 的积分:

$$L_o(\omega_o; \mathbf{x}) = \int_{\Omega} L_i(\omega_i; \mathbf{x}) f_r(\omega_o, \omega_i; \mathbf{x}) \cos(\omega_i; \mathbf{n}) d\omega_i \quad (2)$$

其中, Ω 表示由入射光形成的半球。

根据之前的工作^[31], $f_r(\omega_o, \omega_i; \mathbf{x})$ 可以分解为

2 个分量的和:

$$f_r(\omega_o, \omega_i; \mathbf{x}) = \frac{\mathbf{a}(\mathbf{x})}{\pi} + \tilde{n} f_s(\omega_o, \omega_i; \mathbf{x}) \quad (3)$$

其中,左边项是一个漫反射分量,而 $\mathbf{a}(\mathbf{x})$ 被认为是在 \mathbf{x} 处的漫反射反照率,是由一个 MLP 映射到颜色向量。另一项是镜面分量,是通过将比例因子 \tilde{n} 和 $f_s(\omega_o, \omega_i; \mathbf{x})$ 相乘得到的。这样,就得到了渲染方程的展开形式:

$$L_o(\omega_o; \mathbf{x}) = \frac{\mathbf{a}(\mathbf{x})}{\pi} \int_{\Omega} L_i(\omega_i; \mathbf{x}) \cos(\omega_i; \mathbf{n}) d\omega_i + \tilde{n} \int_{\Omega} L_i(\omega_i; \mathbf{x}) f_s(\omega_o, \omega_i; \mathbf{x}) \cos(\omega_i; \mathbf{n}) d\omega_i \quad (4)$$

其中,方程左边的漫射项命名为 L_d , 其余的光滑反射项为 L_s 。

2.2 镜面反射渲染

渲染方程中的半球积分一般没有一个封闭形式的表达式,因此需要一种复杂且计算代价高的蒙特卡罗方法来进行数值计算。基于 SGs 的渲染框架,在光滑材料和长距离直接照明的条件下,利用 SGs 有效地逼近积分方程。SGs 是各向同性的,也就是说,其中的叶轴具有圆形对称性,但现实世界中的照明在一定程度上是各向异性的。因此,本文提出在

人脸渲染过程中使用 ASGs 来表示图像的照明和 BRDF。接下来简要介绍 ASG 的定义及一些性质, 这些性质将在以后的研究中使用。

各向异性球面高斯分布 (ASG) 是一种球面函数, 其形式为:

$$G(\mathbf{v}; [\mathbf{x}, \mathbf{y}, \mathbf{z}], [\lambda, \mu], c) = c \cdot S(\mathbf{v}; \mathbf{z}) \cdot e^{-\lambda(v \cdot \mathbf{x})^2 - \mu(v \cdot \mathbf{y})^2} \quad (5)$$

其中, \mathbf{v} 表示函数输入; $\mathbf{x}, \mathbf{y}, \mathbf{z}$ 分别表示切轴和双切轴和叶; λ, μ 分别表示 x 轴和 y 轴的带宽, $\lambda, \mu > 0$; c 表示叶的振幅; 平滑项 $S(\mathbf{v}; \mathbf{z}) = \max(\cos(\mathbf{v} \cdot \mathbf{z}), 0)$, 这里 $(\cos(\mathbf{v} \cdot \mathbf{z}), 0)$ 表示 \mathbf{v} 和 \mathbf{z} 之间的夹角的余弦; 其余部分 $e^{-\lambda(v \cdot \mathbf{x})^2 - \mu(v \cdot \mathbf{y})^2}$ 表示其指数项, 所以很容易认为当 $\mathbf{v} = \mathbf{z}$ 时, ASG 取峰值。为简单起见, 将 $G(\mathbf{v}; [\mathbf{x}, \mathbf{y}, \mathbf{z}], [\lambda, \mu], c)$ 简称为 $G(\mathbf{v})$ 。

此外, ASG 具有一个等价的代数形式:

$$G(\mathbf{v}; \mathbf{A}) = S(\mathbf{v}; \mathbf{z}) \cdot e^{-\mathbf{v}^T \mathbf{A} \mathbf{v}} \quad (6)$$

其中, \mathbf{A} 表示 3×3 对称矩阵。式(6)被认为等同于式(5)。为了证明, 本文对对称矩阵 \mathbf{A} 进行奇异值分解(SVD), 即:

$$\begin{aligned} \mathbf{A} &= [\mathbf{x}, \mathbf{y}, \mathbf{z}] \cdot \text{diag}(\lambda_A, \mu_A, \nu_A) \cdot [\mathbf{x}, \mathbf{y}, \mathbf{z}]^T = \\ &\lambda_A \mathbf{x} \mathbf{x}^T + \mu_A \mathbf{y} \mathbf{y}^T + \nu_A \mathbf{z} \mathbf{z}^T = (\lambda_A - \nu_A) \mathbf{x} \mathbf{x}^T + \\ &(\mu_A - \nu_A) \mathbf{y} \mathbf{y}^T + \nu_A \cdot \mathbf{I} \end{aligned} \quad (7)$$

其中, λ_A, μ_A, ν_A 表示 \mathbf{A} 的 3 个特征值; $\mathbf{x}, \mathbf{y}, \mathbf{z}$ 表示 \mathbf{A} 对应的特征向量, 假设 $\lambda_A \geq \mu_A \geq \nu_A$ 。因此, 可以认为, 式(6)和式(5)是等价的, 同时也满足: $c = \exp(-\nu_A)$, $[\lambda, \mu] = [\lambda_A - \nu_A, \mu_A - \nu_A]$ 这些转换关系。

利用 ASG 的代数形式, 本文将证明 2 个 ASG 的乘积仍然可以很好地被近似。假设有 2 个 ASGs:

$$\begin{aligned} G(\mathbf{v}; \mathbf{A}_1) &= S(\mathbf{v}; \mathbf{z}_1) \cdot e^{-\mathbf{v}^T \mathbf{A}_1 \mathbf{v}}, \\ G(\mathbf{v}; \mathbf{A}_2) &= S(\mathbf{v}; \mathbf{z}_2) \cdot e^{-\mathbf{v}^T \mathbf{A}_2 \mathbf{v}} \end{aligned} \quad (8)$$

两者的乘积可以被写成:

$$G_1(\mathbf{v}) G_2(\mathbf{v}) = S(\mathbf{v}; \mathbf{z}_1, \mathbf{z}_2) \cdot e^{-\mathbf{v}^T (\mathbf{A}_1 + \mathbf{A}_2) \mathbf{v}} \quad (9)$$

其中, $S(\mathbf{v}; \mathbf{z}_1, \mathbf{z}_2) = S(\mathbf{v}; \mathbf{z}_1) \cdot S(\mathbf{v}; \mathbf{z}_2)$ 。通过假设 $\mathbf{A}_3 = \mathbf{A}_1 + \mathbf{A}_2$ 和 \mathbf{z}_3 是 \mathbf{A}_3 的最小特征值, 可以得到:

$$\begin{aligned} G_1(\mathbf{v}) G_2(\mathbf{v}) &= \frac{S(\mathbf{v}; \mathbf{z}_1, \mathbf{z}_2)}{\max(\mathbf{v} \cdot \mathbf{z}_3, 0)} \cdot (\max(\mathbf{v} \cdot \mathbf{z}_3, 0) e^{-\mathbf{v}^T \mathbf{A}_3 \mathbf{v}}) = \\ &\frac{S(\mathbf{v}; \mathbf{z}_1, \mathbf{z}_2)}{\max(\mathbf{v} \cdot \mathbf{z}_3, 0)} \cdot G(\mathbf{v}; \mathbf{A}_3) \approx S(\mathbf{z}_3; \mathbf{z}_1, \mathbf{z}_2) \cdot G(\mathbf{v}; \mathbf{A}_3) \end{aligned} \quad (10)$$

在反向渲染中, $L_i(\boldsymbol{\omega}_i; \mathbf{x})$ 、 $f_s(\boldsymbol{\omega}_0, \boldsymbol{\omega}_i; \mathbf{x})$ 、 \mathbf{n} 都是未知的函数或参数。在求解过程中, 最常见的方

法是使用 MLP 来直接预测积分结果。但是, 这种估计算法可能会忽略渲染方程中丰富的信息, 给人一种直观的感觉, 让 MLP 学习一个非常复杂的函数。MLP 将在网络的最后一层生成方程的参数, 但在此之前, MLP 不能估计输出的辐射方向。相反, 受文献[32]的启发, 在向量特征空间中编码渲染方程, 并让 MLP 从编码的结果中预测本文研究的镜面反射。因此, 本文将 L_s 中的函数作为编码向量特征空间的采样 $\boldsymbol{\omega}_i$, 并使用 ASG 来表示每个 $\boldsymbol{\omega}_i$:

$$\begin{aligned} L_s &= \sum_{i=0}^{N-1} G_i(\boldsymbol{\omega}_i; \mathbf{x}, [\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i], [\lambda_i, \mu_i], c_i) = \\ &\sum_{i=0}^{N-1} c_i S(\boldsymbol{\omega}_i; \mathbf{z}_i) \exp(-\lambda_i(\boldsymbol{\omega}_i \cdot \mathbf{x}_i)^2 - \mu_i(\boldsymbol{\omega}_i \cdot \mathbf{y}_i)^2) \end{aligned} \quad (11)$$

其中, c_i 表示 ASG 振幅的特征向量, $[\mathbf{x}_i, \mathbf{y}_i, \mathbf{z}_i]$ 表示满足 $\mathbf{z}_i \cdot \mathbf{x}_i = \mathbf{z}_i \cdot \mathbf{y}_i = \mathbf{x}_i \cdot \mathbf{y}_i = 0$ 的预定义标准正交轴。然而, 渲染方程编码的一个问题是, 当使用 $\boldsymbol{\omega}_0 = -\mathbf{d}$ 的条件时, ASG 不匹配物理镜面反射的行为, 所以本文重新参数化了视图方向:

$$\boldsymbol{\omega}_0 = 2(\mathbf{d} \cdot \mathbf{n})\mathbf{n} - \mathbf{d} \quad (12)$$

渲染方程编码如图 2 所示, 在一个单位球上采样 $N = 128$, 并根据标准正交性确定切轴和双切轴。在球坐标系 $\mathbf{z}_i = (\theta, \varphi)$ 中, 定义了 $\mathbf{x}_i = (\theta + \pi/2)$, 然后使用四元数使 \mathbf{x}_i 围绕 \mathbf{z}_i 轴旋转 90° 得到 \mathbf{y}_i 。 L_s 可以通过计算编码的特征向量和来做估计, 但却大大减少了特征表示的通道。所以本文巧妙地想到了一种方法: 将编码的特征向量连接成一个全局特征向量:

$$\mathbf{g} = [g_0, g_1, \dots, g_{N-1}] \quad (13)$$

结合空间位置特征向量 $\boldsymbol{\xi}$, 本文应用镜面 MLP 来预测反射项 L_s 。ASG 编码所需的参数通过空间 MLP 进行预测, 其中包括以下参数: $\rho, \boldsymbol{\xi}, c_i, \lambda_i, \mu_i$ 。

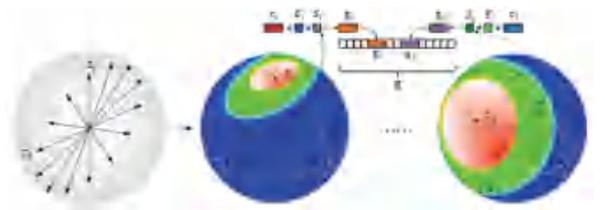


图 2 渲染方程编码

Fig. 2 Rendering equation coding

2.3 漫反射渲染

球面谐波 (SH)^[33] 和 SG 常被用作入射照明的有效表示。本文拟要实现全频信息建模, 但 SH 是带限的; 此外, 还需要基函数的形状可以由参数控

制,但函数有多种形式,所以本文选择使用 ASG 混合函数来表示环境光:

$$L_i(\boldsymbol{\omega}_i; \mathbf{x}) = \sum_k G(\boldsymbol{\omega}_i; A_k) \quad (14)$$

其中, $G(\boldsymbol{\omega}_i; A_k)$ 表示 $G(\boldsymbol{\omega}_i; [\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k], [\lambda_k, \mu_k])$ 的缩写。

为了便于计算,使用 SG 近似表示 $\cos(\boldsymbol{\omega}_i; \mathbf{n})$ 。本文发现,利用一个附加参数 α 的修正公式,能够很容易得到一个具有相似余弦分布的球面高斯分布:

$$\cos(\boldsymbol{\omega}_i; \mathbf{n}) \approx \mu_{\cos} G_{\text{iso}}(\boldsymbol{\omega}_i; \xi_{\cos}, \lambda_{\cos}) - \alpha_{\cos} \quad (15)$$

通过将分布拟合到标准余弦上,参数解析如下:

$\mu_{\cos} \approx 32.7080$, $\xi_{\cos} \approx 0.0315$, λ_{\cos} 被选择为表面法线 \mathbf{n} , $\alpha_{\cos} \approx 31.7003$ 。 G_{iso} 是一个球形高斯分布。

但是,本文发现在表面位置 x 处的漫射辐射 L_d 可以写成一个 SG 和一个 ASG 的卷积。根据 SG 和 ASG 卷积的公式,可以得到 L_d 的表达式:

$$L_d \approx \frac{\mu_{\cos} \cdot \mathbf{a}(x)}{\pi} \sum_k G\left(\xi_{\cos}; [\mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k], \left[\frac{\lambda_{\cos} \cdot \lambda_k}{\lambda_{\cos} + \lambda_k}, \frac{\lambda_{\cos} \cdot \mu_k}{\lambda_{\cos} + \mu_k}\right], \frac{\pi}{(\lambda_{\cos} + \lambda_k)(\lambda_{\cos} + \mu_k)}\right) - \frac{\alpha_{\cos} \cdot \mathbf{a}(x)}{\pi} \sum_k \int_{\Omega} G(\boldsymbol{\omega}_i; A_k) d\boldsymbol{\omega}_i \quad (16)$$

其中,左项中的 ASG 可以简称为 $G_k(\xi_{\cos})$,定义了一个新的参数 η , 定义为:

$$\eta = \pi / \mu_{\cos} - \alpha_{\cos} \quad (17)$$

对于入射辐射 $L_i(\boldsymbol{\omega}_i; \mathbf{x})$, 它们的叶被初始化为在单位球上均匀采样。因此,当带宽 λ_k 、 μ_k 不是特别小时,对单位球上的 ASG 积分的分析可以近似如下:

$$\int_{\Omega} G(\boldsymbol{\omega}_i; A_k) d\boldsymbol{\omega}_i \approx \frac{\pi}{\sqrt{\lambda_k \mu_k}} \quad (18)$$

最后,整个漫反射分量可以有效地写成:

$$L_d \approx \frac{\mathbf{a}(x)}{\eta} \sum_k \left(G_k(\xi_{\cos}) - \frac{\pi}{\sqrt{\lambda_k \mu_k}} \right) \quad (19)$$

因此, L_s 可以由镜面 MLP 计算得到; L_d 可以通过将 $(\rho, \xi, c_i, \lambda_i, \mu_i)$ 和 $(\mathbf{a}(x), \lambda_k, \mu_k, \mathbf{x}_k, \mathbf{y}_k, \mathbf{z}_k)$ 以及表面法线 \mathbf{n} 输入到漫反射渲染模块中计算得到。

2.4 几何建模

在本节中,目标是优化一个有符号距离函数(SDF),以描述基于受试者的图像的人脸重建的详细形状。SDF 是可微的,满足形状和曲面法线之间的约束: SDF 的梯度是曲面的法线。首先,定义 SDF,并解释了如何获得射线和 SDF 的交集。假设 $F_{\theta}: x \rightarrow s$ 可以用来近似隐式曲面 $G_0 = \{x \mid F_{\theta}(x) =$

$0\}$, 其中 x 是面的 3D 点, $F_{\theta}(x)$ 是用 MLP 学习的。SDF 示意如图 3 所示,可以用来表示封闭曲面。为了渲染相机的像素射线的颜色,首先从射线的交点和图像的每个像素,沿着射线通过 Lior 等学者^[34]提出的球形追踪方法找到射线,每一步是当前 SDF 值。因此,将交点的位置 x 和曲面法线 $\mathbf{n} = \nabla_x F_{\theta}$ 分别输入到空间 MLP 和 ASG 混合函数中,计算用于编码渲染方程和 L_d 项的参数。并且梯度必须通过 x 和 \mathbf{n} 传播回 SDF 参数 θ 。最终,为了优化人脸几何,本文使用了一个水平集配准损失来近似查询点和隐式曲面上对应点之间的梯度一致性。

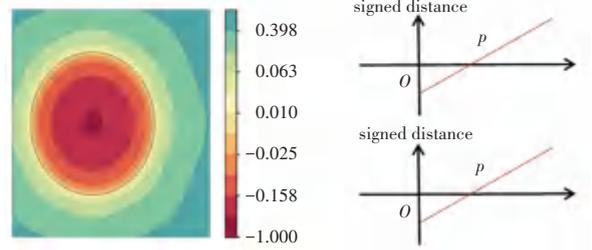


图 3 有符号距离函数

Fig. 3 Signed Distance Function

2.5 体绘制

对于面的多层皮肤,其中沿光的方向有多个表面相交,本文使用 Wang 等学者^[35]的方法在体积上渲染辐射值。从相机中心 \mathbf{o} 沿 $\boldsymbol{\omega}_o$ 方向射线 $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ (t 为步长) 发出的光,总累积颜色值为:

$$C(\mathbf{o}, \boldsymbol{\omega}_o) = \int_{t_1}^{t_2} L_0(\mathbf{r}(t), \boldsymbol{\omega}_o) T(t) \sigma(t) dt \quad (20)$$

其中, $\sigma(t)$ 表示密度函数; $T(t)$ 表示沿射线的累积透过率,并满足:

$$T(t) = \exp\left(-\int_0^t \sigma(u) du\right) \quad (21)$$

为了确保 $\sigma(t)$ 总是非负的,定义了以下不透明的密度函数:

$$\sigma(t) = \max\left(\frac{-\frac{d\Phi_s}{dt}(F(\mathbf{r}(t)))}{\Phi_s(F(\mathbf{r}(t)))}, 0\right) \quad (22)$$

随后,为了自动校准白平衡问题,假设光源下的人脸图像调整为标准光源下的图像,因此在渲染颜色值之前乘以对角矩阵 \mathbf{A}_n :

$$\mathbf{C}_n = \mathbf{A}_n \mathbf{C}(\mathbf{o}, \boldsymbol{\omega}_o) \quad (23)$$

其中, \mathbf{A}_n 表示由 MLP 在第 n 个多视图图像的嵌入上预测的。

2.6 损失函数

(1) 光度损失: 设 N_p 为每次训练迭代时从视图方向 $\boldsymbol{\omega}_o$ 和图像 $\mathbf{C}_n^{\text{ref}}$ 开始的一批像素。在本文之前的

工作中,预测了人脸图像的最终颜色值 $C_n(p)$ ($p \in N_p$),并定义了 $C_n^{\text{gt}}(p)$ 和 $C_n(p)$ 之间的光度损失:

$$\mathcal{L}_{\text{photo}} = \frac{1}{N_p} \sum_{p=1}^{N_p} |C_n^{\text{gt}}(p) - C_n(p)| \quad (24)$$

(2) 表面法线归一化损失。推得的公式如下:

$$\mathcal{L}_{\text{normal}} = \lambda_1 \frac{1}{N} \sum_{i=0}^{N-1} \zeta_i \max(\mathbf{d} \cdot \mathbf{n}_i, 0)^2 \quad (25)$$

其中, λ_1 表示超参数; N 表示样本数; ζ_i 表示颜色的权值。本文引入了一个新的正则化项,能够惩罚脸部远离摄像机的正常方向:当 ζ_i 值过高, \mathbf{n} 和 \mathbf{d} 的点积为正时,将被惩罚。因此,优化后的法线能够计算出精确的反射方向。

(3) 几何损失:为了使 $F_\theta(x)$ 有效地成为一个有符号距离函数,利用自适应正则化推得:

$$\mathcal{L}_{\text{Eikonal}} = \lambda_2 \sum_{x \in \Omega} |\nabla_x F_\theta - 1| \quad (26)$$

此外,通过 SDF MLP 获得 SDF 及其在 x 位置的梯度时,梯度一致性意味着查询点上的梯度和查询点在每个级别集 $G_l = \{x | F_\theta(x) = l\}$ 上投影的梯度应该指向相同的方向。但是,评估每个级别集的一致性需做大量的计算,所以使用 G_0 作为引用,并将所有查询点投影到引用上。水平集配准损失函数的具体形式为:

$$\mathcal{L}_{\text{registration}} = \min_{\theta} \lambda_3 \sum_{x \in \Omega} \beta_x \left(1 - \frac{\nabla F_\theta(x) \cdot \nabla F_\theta(x_0)}{\|\nabla F_\theta(x)\|_2 \cdot \|\nabla F_\theta(x_0)\|_2} \right) \quad (27)$$

其中, λ_3 表示本文损失的平衡权值, β_x 表示一个自适应权重,旨在鼓励优化更多地关注面部表面附近的区域。定义如下:

$$\beta_x = \exp(-\delta \cdot |F_\theta(x)|) \quad (28)$$

综合考虑这 2 个损失,本文得到了几何损失:

$$\mathcal{L}_{\text{geo}} = \lambda_2 \sum_{x \in \Omega} |\|\nabla_x F_\theta\| - 1| + \min_{\theta} \lambda_3 \sum_{x \in \Omega} \beta_x \left(1 - \frac{\nabla F_\theta(x) \cdot \nabla F_\theta(x_0)}{\|\nabla F_\theta(x)\|_2 \cdot \|\nabla F_\theta(x_0)\|_2} \right) \quad (29)$$

综上所述,该方法的全部损失写为:

$$\mathcal{L} = \mathcal{L}_{\text{photo}} + \mathcal{L}_{\text{normal}} + \mathcal{L}_{\text{geo}} + \lambda_4 \|\mathbf{z}\|_2 \quad (30)$$

其中, \mathbf{z} 表示本文所提几何模型的嵌入。

3 实验

本文进行了主观和客观评价实验,并将实验结果与一些相关的三维面部渲染方法进行了比较。

3.1 数据集

本文的目标是从多视图图像中捕捉真实的面部外观和几何形状,所需的数据是具有真实三维模型

的人脸图像。采用了 FaceScape^[36] 三维面部数据集,提供了相机参数、高质量的多视图图像和相应的三维形状。共有 359 个人物和 20 个面部表情,可用图像数量超过 400 K。大部分是亚洲人,并且让每个人做 20 组不同的表情,这样所有的三维模型达到了 18 760 个。该数据集提供了一个包含参与者年龄和性别列表的文本文件,列表中的每一行都从左到右标有身份指数、性别(男性、女性)、年龄和有效性。有效性标签是 [1+4] 的二进制数字,其中第一个数字表示该人的模型是否完整和有效,其余 4 个数字为:obj-模型、mtl-材料、jpg-纹理和 png-dpmap 是否缺失。图 4 为数据集中某一身份微笑表情的多张视图图片。

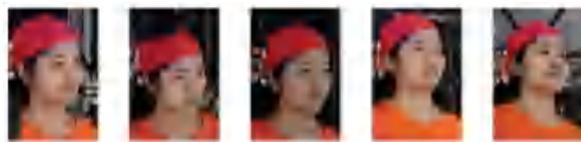


图 4 FaceScape 数据集选取图片

Fig. 4 Pictures selected from FaceScape dataset

3.2 网络结构与实验细节

本文的框架专门使用 MLP 对神经网络进行实例化。镜面 MLP、空间 MLP、SDF MLP、ASG 混合函数 MLP 和光度校准 MLP 的详细结构如图 5 所示。在空间 MLP 中使用了正弦激活函数。

为了获得更好的高频信息,通过正弦位置编码 γ 对 k 个特征点的相对坐标进行编码,写成:

$$\gamma(p) = (\sin(2^0 \pi p), \cos(2^0 \pi p), \dots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p)) \quad (31)$$

其中,实验中的 L 设为 6。

(1) 训练细节:该模型使用 Adam 优化器以端到端方式进行训练,总共为 3 000 个 epochs,初始学习速率设置为 0.000 1,使用 1% 的学习速率进行微调。每 375 个 epochs 的学习率衰减了 0.5 倍,直到训练过程的 75%。在 1 台 NVIDIA RTX 3090 图形处理器上使用 2 048 条射线大约需要 10 h。权衡参数 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 分别设置为 $1.5 \times 10^{-3}, 10^3, 10^{-1}, 10^{-3}$ 。

(2) 体渲染细节:使用球面跟踪技术来加速渲染过程。为了避免在第一步中追踪到面部表面, SDF 值被截断为 $[-5 \text{ cm}, 5 \text{ cm}]$ 。在跟踪过程中,用查询的有符号距离值的 1.4 倍来更新光线。由于几何先验 ImFace 定义了单位球体内的表面,其中后空间中的 SDF 值是负的,假设从负位置进入球面的光线与曲面没有相交(表示为“未命中”)。

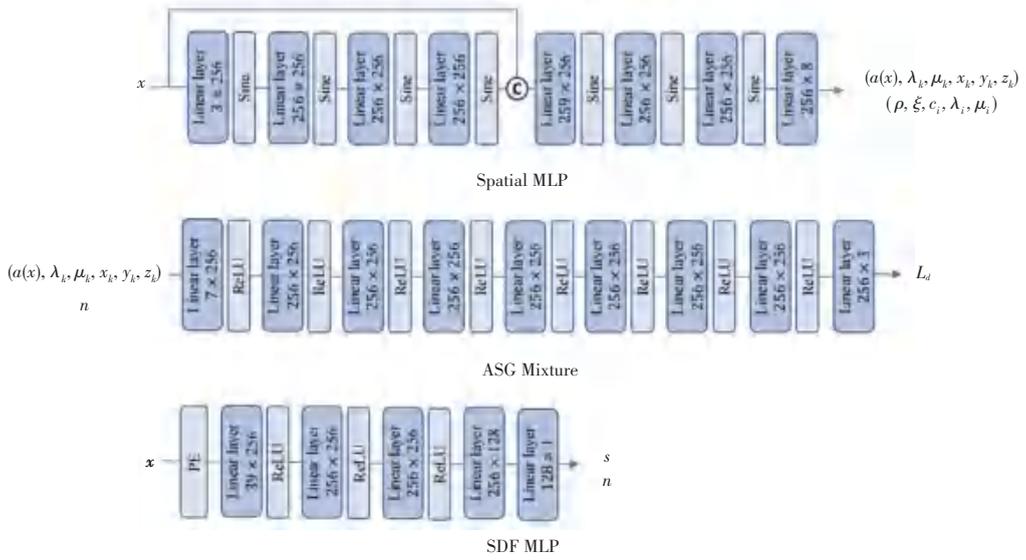


图 5 网络结构图

Fig. 5 Network structure diagram

3.3 实验对比

本文使用 PSNR, SSIM^[37] 和 LPIPS^[38] 作为评价合成质量的指标,而合成结果与真实三维模型的倒角距离被用来衡量几何精度。实验结果如图 6 所示,可以看出 Ref-NeRF^[39] 未能恢复面部外观,这主要是由于人脸具有较强的形状-外观歧义性。PhySG^[40] 是为重建而设计的,因其牺牲了人脸的细节外观质量,导致了模糊的人脸效果。NRFF 和 NeuFace^[41] 都是基于物理的神经渲染方法,旨在实现端到端的外观和几何恢复。然而,受限于解析 BRDFs 的方法不足,也只能恢复具有明显伪影的粗糙面部外观。相比之下,本文的方法以更加丰富的肤色细节和高保真的图像亮度实现了逼真的人脸外观。定量实验结果见表 1,其中 S. N. 表示每条射线采样点数,并在受试者 1 微笑表情上进行了评估。由表 1 可知,本文方法比一般的体积渲染需要更少的采样点。

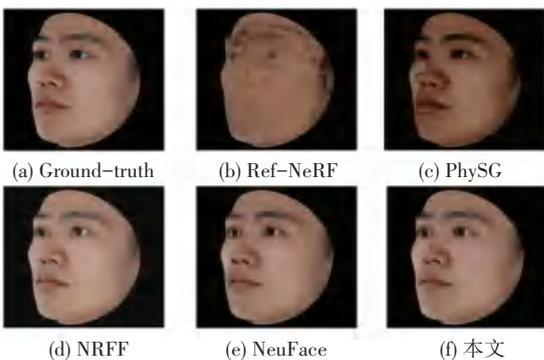


图 6 与一些最新方法实验对比

Fig. 6 Experimental comparison with some of the latest methods

表 1 与一些最新方法的定量比较

Approachs	S. N. ↓	Subject 1			
		PSNR ↑	SSIM ↑	LPIPS ↓	Chamfer ↓
Ref-NeRF	256.0	19.83	0.845	0.113 4	2.042
NRFF	325.0	29.26	0.944	0.057 2	0.628
PhySG	40.2	27.77	0.928	0.061 1	0.864
NeuFace	35.8	31.25	0.958	0.023 7	0.447
本文	35.2	32.34	0.949	0.021 5	0.386

3.4 消融实验

本文的实验是在受试者 1 上进行的,因其人脸图像具有更复杂的亮点。神经 BRDFs 模块是为了对面部皮肤具有更高的表征能力而设计的,本文将替换为 Phong 和 Disney BRDFs^[25] 进行消融研究。特别地,Phong BRDFs 中的镜面反射项由单一的亮度参数控制。DisNey BRDFs 采用 DIFFREC^[13] 提供的代码实现。表 2 揭示了神经 BRDFs 在学习复杂皮肤反射时具有更高的表征能力。

表 2 阴影模型上的消融实验

Shading Model	PSNR ↑	SSIM ↑	LPIPS ↓	Chamfer ↓
ASG+Phong Model	29.88	0.948	0.031 4	0.521
SG+Disney BRDF	30.52	0.956	0.024 5	0.493
SG+Neural Bases	31.05	0.960	0.026 7	0.414
ASG+ Neural Bases	32.18	0.964	0.022 8	0.428

4 结束语

本文提出了基于 ASG 混合函数的 ASG-Face 三

维人脸渲染框架。利用 ASG 来近似照明和 BRDF, 可以联合表示来自多视图面部图像的几何形状和复杂的面部皮肤反射率。其中, 采用 SDFs 表示形状。与 SG 相比, ASG 能够有效地捕捉各向异性反射, 使其成为精确建模面部复杂特征细节的更合适的选择。实验结果证明, ASG-Face 算法在人脸渲染方面具有优越性。但本文同样存在不足, 主要关注复杂面部皮肤的反射建模, 却未能切实解决更具有挑战性的次表面散射问题, 并且只应用了简化的阴影模型, 因此本研究中取得的渲染结果的保真度还有待进一步提升。

参考文献

- [1] AZUMA R, BAILLOT Y, BEHRINGER R, et al. Recent advances in augmented reality[J]. *IEEE Computer Graphics and Applications*, 2001, 21(6): 34-47.
- [2] SLATER M, WILBUR S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments [J]. *Presence: Teleoperators & Virtual Environments*, 1997, 6(6): 603-616.
- [3] KATO H, BILLINGHURST M. Marker tracking and hmd calibration for a video-based augmented reality conferencing system[C]//*Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. Piscataway, NJ: IEEE, 1999: 85-94.
- [4] YANG Mengzhao, WANG Kuanquan, ZHANG Lei. Realistic real-time facial expressions animation via 3D morphing target [J]. *Journal of Software*, 2013, 8(2): 418-425.
- [5] PATEL A, SMITH W A P. 3D morphable face models revisited [C]//*Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2009: 1327-1334.
- [6] BOLKART T, WUHRER S. A groupwise multilinear correspondence optimization for 3D faces [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2015: 3604-3612.
- [7] TRAN L, LIU Feng, LIU Xiaoming. Towards high-fidelity nonlinear 3D face morphable model [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2019: 1126-1135.
- [8] BOURITSAS G, BOKHNYAK S, PLOUMPIS S, et al. Neural 3D morphable models: Spiral convolutional networks for 3D shape representation learning and generation[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2019: 7213-7222.
- [9] GROPP A, YARIV L, HAIM N, et al. Implicit geometric regularization for learning shapes[J]. *arXiv preprint arXiv*, 2002. 10099, 2020.
- [10] GENOVA K, COLE F, VLASIC D, et al. Learning shape templates with structured implicit functions[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2019: 7154-7164.
- [11] LIU Feng, LIU Xiaoming. Learning implicit functions for topology-varying dense 3D shape correspondence[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 4823-4834.
- [12] WU Shangzhe, RUPPRECHT C, VEDALDI A. Unsupervised learning of probably symmetric deformable 3D objects from images in the wild[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(4): 5268-5281.
- [13] PHONG B T. Illumination for computer generated pictures[J]. *Communications of the ACM*, 1975, 18(6): 311-317.
- [14] KAJIYAJ T. The rendering equation[C]//*Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques*. New York: ACM, 1986: 143-150.
- [15] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis [J]. *Communications of the ACM*, 2021, 65(1): 99-106.
- [16] MILDENHALL B, HEDMAN P, MARTIN-BRUALLA R, et al. Nerf in the dark: High dynamic range view synthesis from noisy raw images[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2022: 16190-16199.
- [17] YAN Lingqi, ZHOU Yahan, XU Kun, et al. Accurate translucent material rendering under spherical gaussian lights[J]. *Computer Graphics Forum*, 2012, 31(7): 2267-2276.
- [18] BINGHAM C. An antipodally symmetric distribution on the sphere[J]. *The Annals of Statistics*, 1974, 2: 1201-1225.
- [19] PHARR M, JAKOB W, HUMPHREYS G. *Physically Based Rendering: From theory to implementation* [M]. Cambridge, MA: MIT Press, 2023.
- [20] SITZMANN V, MARTEL J, BERGMAN A, et al. Implicit neural representations with periodic activation functions [J]. *Advances in Neural Information Processing Systems*, 2020, 33: 7462-7473.
- [21] ATZMON M, LIPMAN Y. SAL++: Sign agnostic learning with derivatives[J]. *arXiv preprint arXiv*, 2006. 05400, 2020.
- [22] WANG Pengshuai, LIU Yang, YANG Yuqi, et al. Spline positional encoding for learning 3D implicit signed distance fields [J]. *arXiv preprint arXiv*, 2106. 01553, 2021.
- [23] MA Baorui, HAN Zhizhong, LIU Yushen, et al. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces [J]. *arXiv preprint arXiv*, 2011. 13495, 2020.
- [24] MA Baorui, ZHOU Junsheng, LIU Yushen, et al. Towards better gradient consistency for neural signed distance functions via level set alignment[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ: IEEE, 2023: 17724-17734.
- [25] BURLEY B, STUDIOS W D A. *Physically-based shading at disney* [C]// *Proceedings of the ACM SIGGRAPH*. New York: ACM, 2012: 1-7.
- [26] ROMDHANI V. Efficient, robust and accurate fitting of a 3D morphable model [C]//*Proceedings of the Ninth IEEE International Conference on Computer Vision*. Piscataway, NJ: IEEE, 2003: 59-66.
- [27] LI Tianye, BOLKART T, BLACK M J, et al. Learning a model of facial shape and expression from 4D scans [J]. *ACM Transactions on Graphics*, 2017, 36(6): 194.
- [28] LUAN T, LIU Xiaoming. Nonlinear 3D face morphable model [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2018: 7346-

- 7355.
- [29] ZHENG Mingwu, YANG Hongyu, HUANG Di, et al. Imface: A nonlinear 3D morphable face model with implicit neural representations [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022; 20343–20352.
- [30] BINDITA C, VESDAPUNT N, SHAPIRO L, et al. Personalized face modeling for improved face reconstruction and motion retargeting [C] // Proceedings of 16th European Conference on the Computer Vision. Cham: Springer, 2020; 142–160.
- [31] GOTARDO P, RIVIERE J, BRADLEY D, et al. Practical dynamic facial appearance modeling and acquisition [J]. ACM Transactions on Graphics (ToG), 2018, 37(6): 232.
- [32] HAN Kang, XIANG Wei. Multiscale tensor decomposition and rendering equation encoding for view synthesis [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023; 4232–4241.
- [33] CHEN Wenzheng, GAO Jun, LING Huan, et al. Learning to predict 3D objects with an interpolation-based differentiable renderer [C] // Proceedings of the 33rd International Conference on Neural Information Processing Systems. New York: ACM, 2019; 9609–9619.
- [34] LIOR Y, KASTEN Y, MORAN D, et al. Multiview neural surface reconstruction with implicit lighting and material [J]. arXiv preprint arXiv, 2003.09852, 2020.
- [35] WANG Peng, LIU Lingjie, LIU Yuan, et al. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction [C] // Proceedings of the 35th Conference on Neural Information Processing Systems. NeurIPS, 2021; 27171–27183.
- [36] YANG Haotian, ZHU Hao, WANG Yanru, et al. Facescape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020; 601–610.
- [37] WANG Zhou, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13(4): 600–612.
- [38] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018; 586–595.
- [39] VERBIN D, HEDMAN P, MILDENHALL B, et al. Ref-nerf: Structured view-dependent appearance for neural radiance fields [C] // Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2022; 5481–5490.
- [40] ZHANG Kai, LUAN Fujun, WANG Qianqian, et al. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021; 5453–5462.
- [41] ZHENG Mingwu, ZHANG Haiyu, YANG Hongyu, et al. Neuface: Realistic 3D neural face rendering from multi-view images [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023; 16868–16877.