

中图法分类号: TP399 文献标识码: A 文章编号: 1006-8961(2024)09-2513-28

论文引用格式: Hao C H, Du Y Y, Wang L and Wang B B. 2024. Survey of digital face rendering and appearance recovery methods. Journal of Image and Graphics, 29(09): 2513-2540 (郝琮晖, 杜悠扬, 王璐, 王贝贝. 2024. 数字人脸渲染与外观恢复方法综述. 中国图象图形学报, 29(09): 2513-2540) [DOI: 10.11834/jig.230683]

## 数字人脸渲染与外观恢复方法综述

郝琮晖<sup>1</sup>, 杜悠扬<sup>2</sup>, 王璐<sup>2</sup>, 王贝贝<sup>3\*</sup>

1. 南开大学计算机学院, 天津 300350; 2. 山东大学软件学院, 济南 250101; 3. 南京大学智能科学与技术学院, 苏州 215163

**摘要:** 数字人技术引起了数字孪生、元宇宙等领域的广泛关注, 其中人脸作为数字人的重要构成部分, 其数字化生成和呈现成为人们关注的焦点, 且相关技术已经在电影、游戏等领域得到了广泛应用。人们对实现逼真的人脸效果以及精确恢复人脸的需求日益增长, 但由于人脸的多层材质结构、复杂的半透明皮肤效果以及毛孔、褶皱等微观特征的综合影响, 实现高保真的、高效的人脸渲染一直是领域内的难题。此外, 通过采集设备对人脸的几何和外观进行恢复是构建人脸数据的重要方式, 然而对人脸的高品质恢复也同样受限于高成本的采集设备和相关数据集的不足。本文对数字人脸的渲染与恢复的相关方法进行综述。首先介绍了真实感人脸的渲染方法, 根据其不同的渲染原理, 将它们分为基于扩散近似的渲染方法和基于蒙特卡洛采样的渲染方法, 并着重分析了基于近似扩散渲染方法的发展现状及面临的问题。进一步, 将人脸恢复工作分类为基于专业采集设备的高精度恢复和基于深度学习的低精度恢复。针对高精度人脸恢复, 从主动照明和被动捕获两个分支, 对相应的工作进行了总结。针对结合深度学习的低精度人脸恢复方法, 将其分类为几何细节的恢复、纹理贴图的恢复以及人脸材质信息的恢复3个方面进行介绍。本文系统地论述了各类方法的核心思路, 并进行了横向对比和分析。最后, 对未来人脸渲染及恢复方法的发展趋势进行了展望。希望本文可以为人脸渲染和外观恢复的初学者提供一些背景知识和思路启发。

**关键词:** 人脸真实感渲染; 次表面散射; 人脸逆向恢复; 主动照明; 被动捕获; 深度学习

### Survey of digital face rendering and appearance recovery methods

Hao Conghui<sup>1</sup>, Du Youyang<sup>2</sup>, Wang Lu<sup>2</sup>, Wang Beibei<sup>3\*</sup>

1. College of Computer Science, Nankai University, Tianjin 300350, China; 2. School of Software, Shandong University, Jinan 250101, China; 3. School of Intelligence Science and Technology, Nanjing University, Suzhou 215163, China

**Abstract:** Digital human technology has attracted widespread attention in digital twins and metaverse fields. As an integral part of digital humans, people have started focusing on facial digitization and presentation. Consequently, the associated techniques find extensive applications in film, gaming, and virtual reality. A growing demand for facial realism rendering and high-quality facial inverse recovery has been observed. However, given the complex and multilayered material structure of the face, facial realism rendering presents a challenge. Furthermore, the composition of internal skin chemicals, such as melanin and hemoglobin, highly influences skin rendering. Factors, such as temperature and blood flow rate, may influence the skin's appearance. The semitransparency of the skin introduces difficulties in the simulation of subsurface scattering effects, in addition to the wide presence of microscopic geometric features, such as pores and wrinkles on the face. All the issues mentioned above cause problems in the rendering domain and raise the demand for the quality of facial

收稿日期: 2023-09-25; 修回日期: 2024-04-01; 预印本日期: 2024-04-08

\* 通信作者: 王贝贝 beibei.wang@nju.edu.cn

基金项目: 国家自然科学基金项目(62172220); 新一代人工智能国家科技重大专项(2022ZD0116305)

Supported by: National Natural Science Foundation of China (62172220); National Science and Technology Major Project (2022ZD0116305)

recovery. In addition, as a result of people's exposure to real human faces in daily life, a heightened sensitivity to the texture and details of digital human faces has been observed, and this condition places greater demands on their realism and accuracy. Meanwhile, recovery of facial geometry and appearance is a crucial method for the construction of facial datasets. However, the high costs of acquisition equipment often constrain high-quality facial recovery, and most studies are limited by the acquisition speed for facial data, which result in the challenging capture of dynamic facial appearance. Lightweight recovery methods also encounter challenges related to the lack of facial material datasets. This paper presents an overview of recent advances in rendering and recovery of digital human faces. First, we introduce methods for realistic facial rendering and categorize them based on diffusion approximation and Monte Carlo approaches. Methods based on diffusion approximation, which focus on the efficient achievement of the semitransparency effect of the skin, are constrained by strict assumptions and suffer from certain limitations in precision. However, their simplified subsurface scattering models can render satisfactory images relatively quickly. Dynamic and interactive applications, such as games, often apply these methods. On the other hand, methods based on the Monte Carlo approach yield high precision and robust results via the meticulous and comprehensive simulation of the complex interactions between light and skin but require long computation times to converge. In applications, such as movies, where highly realistic visual effects are needed, they often become the preferred choice. We emphasized the development and challenges of methods based on diffusion approximation and divided them into improvements in the diffusion profiles, with real-time implementation of subsurface scattering, and hybrid methods combined with Monte Carlo techniques for detailed discussion. A recent Monte Carlo research aimed at improving the convergence rate for applications in facial rendering, including zero-variance random walks, next-event estimation, and path guiding. Second, we divided facial recovery work into two categories: high-precision recovery based on specialized acquisition equipment and low-precision recovery based on deep learning. This paper further categorizes the former based on the use of specialized lighting equipment, which distinguishes between active illumination and passive capture techniques, with provided detailed explanations for each category. Active illumination relies on professional lighting equipment, such as the application of gradient lighting to recover high-precision normal maps, to improve recovery quality. Conversely, passive capture methods are independent of professional lighting equipment, and any artificially provided lighting is limited to uniform illumination to reduce the interference of scene lighting on recovery and similar auxiliary roles. The exploration also focuses on low-precision facial recovery methods incorporating deep learning and classifies them into three categories, namely, geometric detail, texture mapping, and facial material information recoveries, to provide in-depth insights into each approach. We discuss a strategy for overcoming the limitations of geometric recovery based on parametric models, introduced refined parametric expressions of models, and predicted a range of maps, including displacement maps, that represent the model surface's geometric details. For texture recovery, we explored the application of deep neural networks in generative tasks in the prediction of high-fidelity and personalized facial skin textures. Comprehensive reviews the various attempts to mitigate the ill-posed problem of separating reflectance information. In addition, we introduce the facial recovery work using multiview images and video sequences. These low-precision facial recovery methods can gain a wide application space given their flexibility and achieve improved recovery results with the rapid development of deep learning technology. Finally, the future trends in facial realism rendering and recovery methods based on the current state of research are outlined. In the realm of facial realism, existing works often represent the material properties of faces using texture maps and neglect the unique principles of skin coloration as a biological material. Furthermore, the rapid development of deep learning technology increases the importance of exploring its integration with currently rendering techniques. In terms of inverse recovery, the lack of high-quality open-source datasets often poses limitations on data-based facial recovery methods. In addition, substantial improvement is needed in modeling and recovering details at the skin pore level. Combination of inverse recovery with text-based generative work also holds enormous potential and application scenarios. Hopefully, this paper can provide novice researchers in facial rendering and appearance recovery with valuable background knowledge and inspiration from harmony and ideas.

**Key words:** facial realism rendering; subsurface scattering; facial inverse recovery; active illumination; passive capture; deep learning

## 0 引言

随着互联网技术的发展以及元宇宙概念的普及,数字人领域受到人们的广泛关注,其相关技术已广泛应用于影视、游戏和虚拟现实等领域。而伴随着数字人技术受众越来越广,各领域对数字人的真实感也提出了更高的要求。

作为数字人外观部分的核心,如何实现数字人脸的真实感是至关重要的。然而由于人脸中的分层结构、复杂的材质和微观特征,实现真实的人脸渲染一直是领域内的一大挑战,而且由于人们在日常生活中接触到众多真实的人脸,因此对于数字人脸的质感和细节有着高度敏感的认知,这对其真实感和准确性的实现提出了更高的要求。此外,对人脸的精确逆向恢复也往往需要昂贵的采集设备,轻量化的方法则受限于图像场景信息难以分离以及相关数据集的匮乏往往准确性较低;如何实现高精度且简便的人脸恢复,同样具有着重要的研究价值以及广泛的应用需求。

本文对现有的数字人脸研究进行了概述与总结,分别从人脸皮肤材质的正向真实感渲染和逆向恢复两个方向展开,深入讨论了相关方法及其中所涉及的问题。在人脸的正向真实感渲染方面,根据渲染流程的实现方式,分为基于扩散近似的方法和基于蒙特卡洛法的方法。对基于扩散近似的发展流程根据扩散剖面的改进、次表面散射的实时实现以及结合蒙特卡洛法的混合方法3个方向进行了详细的梳理,并关注了近年应用于人脸渲染的蒙特卡洛方法的加速收敛研究。在逆向恢复人脸方面,将高精度人脸捕获根据是否使用专门的照明设备分为主动照明和被动捕获两类分别介绍。在被动捕获的范畴下,基于深度学习的低精度恢复技术是一个重要分支,本文对该部分单独进行整理,该部分可进一步划分为几何细节的恢复、纹理贴图的恢复以及人脸材质信息的恢复这3个方向。最后,深入讨论了当前数字人脸研究中所面临的难点,以及未来该领域的发展趋势。需要明确的是,本文重点关注人脸渲染的真实感相关内容,风格化人脸等计算机视觉领域的工作不在本文的讨论范围之内。

## 1 人脸的真实感渲染方法

真实感渲染模拟了光线在给定几何、材质及光照等场景信息下的传播过程,旨在生成高度逼真的图像,以尽可能接近现实世界,产生令人信服的视觉效果。人脸的真实感渲染较难实现,这是因为其主体的皮肤材质与常见的金属、塑料等材质不同,具有复杂的光学性质,难以使用传统的双向反射分布函数(bidirectional reflectance distribution function, BRDF)直接模拟,这也是本节讨论的重点。

皮肤作为一种半透明的参与介质,当光线与其相互作用时,除了在表面发生反射外,光线还会折射进入皮肤内部,在内部进行多次散射后随机游走,最终再次从物体表面射出,这一过程称为次表面散射(subsurface scattering)。由于次表面散射在皮肤的光照效果中起主要作用(Hanrahan和Krueger,1993;Gosselin,2004),且皮肤表面的反射效果通常可以使用Torrance-Sparrow模型(Torrance和Sparrow,1967)模拟(Debevec等,2000;Weyrich等,2006),因此皮肤渲染的关键在于如何实现次表面散射。

### 1.1 背景知识

皮肤等半透明材质的次表面散射可以使用双向散射表面反射分布函数(bidirectional scattering surface reflectance distribution function, BSSRDF)进行描述(Nicodemus等,1977)。BSSRDF定义为

$$S(x_i, \omega_i; x_o, \omega_o) = \frac{dL(x_o, \omega_o)}{d\Phi_i(x_i, \omega_i)} \quad (1)$$

式中, $L(x_o, \omega_o)$ 为出射点 $x_o$ 沿 $\omega_o$ 方向的出射辐射度, $\Phi_i(x_i, \omega_i)$ 为入射点 $x_i$ 沿 $\omega_i$ 方向的入射通量。给定材质的BSSRDF后,可通过对出射点散射范围 $A$ 内的入射辐射度 $L_i$ 进行积分得到最终的出射辐射度 $L_o$ ,具体为

$$L_o(x_o, \omega_o) = \int_A \int_{\Omega} S(x_i, \omega_i; x_o, \omega_o) L_i(x_i, \omega_i) (N \times \omega_i) d\omega_i dA(x_i) \quad (2)$$

式中, $N$ 为入射点对应的法向量。

光线从表面进入物体后,仅经历一次散射或折射,就出射到物体表面的部分称为单次散射;而在内部经历不少于两次的散射或折射事件后,才最终出射到物体表面的部分称为多次散射。

基于皮肤与光线之间的复杂作用过程,起初研

究人员通常使用蒙特卡洛方法来模拟参与介质的次表面散射(Wang等, 1995; Lafortune和Willems, 1996; Jensen和Christensen, 1998; Pauly等, 2000)。此类方法精度较高,但对于皮肤等高度散射材质则需要花费很长时间才能收敛,因此无法得到广泛应用。Stam(1995)基于辐射传输方程引入了扩散近似理论,以模拟物体内部的多次散射效果。基于扩散近似的相关方法相较于蒙特卡洛方法,精度较低但速度更快,同时也是皮肤渲染中的主要分支。本节分别对近年来应用于皮肤真实感渲染的基于扩散近似的方法和基于蒙特卡洛方法进行介绍。

### 1.2 基于扩散近似的皮肤渲染方法

Jensen等人(2001)首先基于扩散近似理论对BSSRDF进行简化,将BSSRDF转化为

$$S(x_i, \omega_i; x_o, \omega_o) = \frac{1}{\pi} F_t(\eta_o, \omega_o) R(\|x_o - x_i\|) F_t(\eta_i, \omega_i) \quad (3)$$

式中,  $F_t(\eta_o, \omega_o)$  与  $F_t(\eta_i, \omega_i)$  分别代表光线折射进入物体内部以及从物体内部出射时的菲涅尔透射项,  $\eta_o$  与  $\eta_i$  分别对应出射与入射时的相应折射率;  $R$  为扩散剖面函数,该函数用于描述光线在物体内部多次散射后的分布情况。Jensen等人(2001)还通过偶极子模型近似,给出了  $R$  的解析数学表达式,以对物体内部的多次散射进行求解。需要注意的是,次表面散射中的单次散射部分仍然需要进行额外的蒙特卡洛路径追踪计算,而且在随后的许多基于扩散近似的工作中,这种额外的单次散射计算也都是必要的。

该方法可以实现半透明材质的效果,图1展示了分别使用该方法与传统BRDF方法渲染皮肤的结果。但该方法的渲染结果呈现一种蜡质感,模糊了皮肤特征。该方法基于扩散近似理论进行推导,推导过程假设了物体表面为厚度无限的均匀同介质平面,且光线垂直入射物体表面;这些假设导致该方法在渲染人脸时存在一定误差,且在曲率较大和厚度较薄的区域尤为明显。

此后的扩散近似工作基于该方法进行改进,可以分为3个方向:扩散剖面的改进、次表面散射的实时实现以及结合蒙特卡洛方法的混合方法。

#### 1.2.1 扩散剖面改进

Jensen等人(2001)基于扩散近似,提出扩散剖



图1 使用传统BRDF以及偶极子的人脸渲染结果  
(Jensen等,2001)

Fig. 1 A face rendered using the BRDF model and the dipole model (Jensen et al., 2001)  
(a) BRDF model; (b) dipole model

面  $R$  对光线在物体内部的多次散射情况进行描述,得到了BSSRDF的数学解析表达式以实现次表面散射。但由于他们的方法基于扩散近似假设产生了明显的误差,为了实现更好的皮肤次表面散射效果,需要进一步对扩散剖面进行改进。

但由于偶极子近似模型在推导过程中将物体视做厚度无限的均匀同介质平面,无法模拟光线穿过介质或多层参与介质叠加在一起的情况,也就难以拟合皮肤的多层材质结构。Donner和Jensen(2005)针对此提出了多极子近似来模拟多层半透明材料的次表面散射:该方法使用多个偶极子来模拟光线在单层薄板内部与表面间的反射、透射效果,其与偶极子的比较示意如图2所示。对每两层相邻薄板,将其反射和透射分布函数傅立叶变换到频域中,基于Kubelka-Munk理论(Kubelka和Munk, 1931)进行卷积和修正,进而得到两层薄板叠加后的光线扩散效果。通过重复地进行这一过程,可以得到多层半透明材料的光线扩散效果。

之后,Donner和Jensen(2006)结合多极子近似,

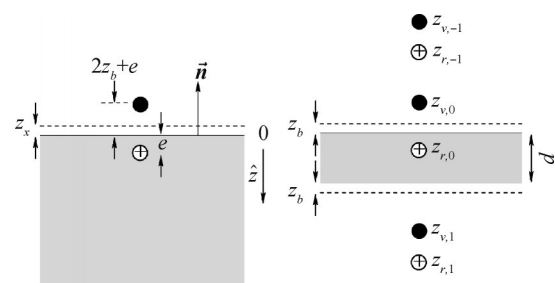


图2 半无限几何的偶极子配置和薄板的多极子配置  
(Donner和Jensen, 2005)

Fig. 2 Dipole configuration for semi-infinite geometry and the multipole configuration for thin slabs  
(Donner and Jensen, 2005)

提出了一个光谱皮肤渲染模型。该模型可通过调控皮肤内黑色素总量、黑色素类型、血红蛋白含量和表面油脂度的比例渲染对应的皮肤效果。该方法还可以通过反照率贴图来模拟皮肤的细节。

Donner 等人(2009)基于大规模的蒙特卡洛模拟,提出了一个使用椭圆坐标表示的分析函数来拟合次表面散射中的出射光分布。该分析函数的参数分别由次表面散射的控制参数:入射角、出射位置、反照率、相位函数和折射率计算得到,因此可以使用一张函数的参数表来记录各种次表面散射参数设置下的拟合结果。在渲染时,通过对表内记录数据进行插值可以得到任意光照和材质条件下的函数参数,代入分析函数,得到次表面散射的出射光分布。需要注意的是,该方法仅使用蒙特卡洛模拟用于生成数据,渲染过程也仍基于以往的扩散近似理论,并没有对人脸的渲染效果作出很大改进。

d'Eon 和 Irving(2011)提出了新的扩散理论以及一种新的量化扩散剖面,通过对扩散函数进行时间量化,将其分解为多个高斯函数,从而实现了一个分析的、可扩展的、多层的 BSSRDF 模型。通过使用该方法进行皮肤渲染可以产生逼真且富有细节的皮肤材质效果。但由于该方法的扩散剖面函数计算过程比较复杂,可能会导致渲染结果不稳定。且该方法仍然对光照条件进行简化,并且在渲染一些具有复杂形状和表面细节的半透明物体时,该方法需要额外的多次采样。随后,d'Eon(2012)同样依照新的扩散理论对偶极子方法(Jensen 等,2001)的公式进行改进,尽管新的偶极子模型无法达到此前量化扩散模型(d'Eon 和 Irving,2011)的精度,但模型的计算流程更为简洁。而相较于传统的偶极子模型,新的模型能够更好地适应不同材质的折射率,并在远离光源的情况下表现出更高的精度。

针对以往工作均对入射光线方向进行限制的问题,Frisvad 等人(2014)提出了定向偶极子模型。不同于传统的偶极子模型使用两个点光源拟合半透明材质内的次表面散射效果,该方法使用两个方向光,即折射后的入射光线和在边界处反射的虚拟光线进行拟合。该方法提出了第1个能够考虑入射和出射光线方向的 BSSRDF 模型。此外,通过引入 delta-Eddington 近似(Joseph 等,1976),该方法不再需要额外进行蒙特卡洛光线追踪以模拟单次散射。但该方法仍然假设物体表面为均匀介质的半无限平面。

对所有基于扩散近似推导的次表面散射渲染方法,Christensen 和 Burley(2015)提出的扩散剖面是目前最接近蒙特卡洛路径追踪拟合效果的。该研究中,Burley 基于 Schlick 近似(Schlick,1994)使用简单函数对复杂光学函数进行逼近的思路,使用两个指数函数之和除以散射距离作为扩散剖面对次表面散射进行拟合。通过与蒙特卡洛路径追踪模拟以及其他基于扩散近似方法的模型进行比较,该方法的拟合效果优于此前所有基于扩散近似的方法。同时,该方法的参数十分简洁,仅包含反照率以及平均自由程,并且该方法的拟合结果中同样包含单次散射,不需要额外计算。

### 1.2.2 次表面散射实时实现

尽管基于扩散近似的方法相较蒙特卡洛方法在速度上有了很大提高,但仍然无法满足游戏等应用场景中的实时皮肤渲染需求,需要进一步改进渲染方法。

Jensen 和 Buhler(2002)为了加快渲染速度,将渲染流程分为两个步骤进行,在第1步中将许多采样点随机分布在物体的表面,并在每个点预计算对应的辐照度;在第2步中使用分层积分技术快速查找预计算结果进行次表面散射拟合。这种方法相较于 Jensen 等人(2001)此前的工作,渲染速度提高了2~3个数量级;但这种方法一方面需要额外的内存开销去记录辐照度的预计算结果;另一方面采样点的分布也会直接影响渲染结果的精度,可能引起噪声。此外,该方法的渲染速率仍不支持次表面散射的实时实现。

Borshukov 等人(2005)提出了一种简单的经验方法以快速渲染皮肤材质。他们将漫反射辐照度光栅化为纹理图,并通过对纹理图进行卷积来模拟次表面散射的效果;但该方法只是基于经验性的模糊,不能模拟多层皮肤材质的物理特性。d'Eon 等人(2007)在此基础上,使用多个高斯函数的加权和拟合多极子模型中皮肤材质的扩散剖面,通过在纹理空间上进行高斯卷积拟合了次表面散射的效果,大幅提高了渲染效率。但由于该方法的模糊过程是在纹理空间进行的,并没有考虑到遮挡以及像素的背面剔除,导致了一定的开销浪费。Jimenez 等人(2009)针对这个问题作出改进,将次表面散射的模拟过程从纹理空间转换到屏幕空间中进行。他们将次表面散射的近似过程简化为对渲染的后处理,在

渲染后的图像中进行卷积模糊,并根据图像中人脸的深度和深度梯度调整卷积核的大小。这种方法不再需要额外计算辐照度贴图,即使在屏幕中同时出现多张人脸,也可以保证实时渲染。

但由于这些方法使用多个高斯函数对皮肤的扩散剖面进行拟合,在实时渲染的实现过程中需要多次卷积,仍然需要一定的计算开销。Jimenez 等人(2015)提出了一种预积分的可分离核,可以将次表面散射拟合过程进一步简化为两次一维卷积,再次提高了皮肤次表面散射的实现效率。

### 1.2.3 结合蒙特卡洛方法的混合方法

基于扩散近似的方法始终受限于其严格的假设,保证效率的同时却难以达到蒙特卡洛方法的精度。研究人员尝试将两种方法混合在一起,同时结合蒙特卡洛方法的准确性与扩散近似方法的效率。基于这一想法,Li 等人(2005)提出将蒙特卡洛路径追踪与偶极子模型相结合,在渲染时使用蒙特卡洛路径追踪方法来模拟前几次散射事件,当光线路径追踪到半透明物体表面下一定深度后,切换为偶极子扩散近似来加速收敛。

Donner 等人(2008)将光子追踪技术与扩散近似结合。在渲染多次散射效果时,通过光子追踪将光子分布在半透明材料内,并记录光子第一次发生散射的位置,之后通过扩散近似方法计算这些光子在物体表面处的多次散射效果。除了偶极子和多极子外,Donner 等人(2008)还引入了一种新的四极子模型,用于模拟光线在直角边缘处的扩散效果,如图3所示。该方法可以处理倾斜光照乃至全局光照。

但上述方法仍受限于结合的不准确的扩散模型影响,无法处理多层材料等情况。为此,Habel 等人

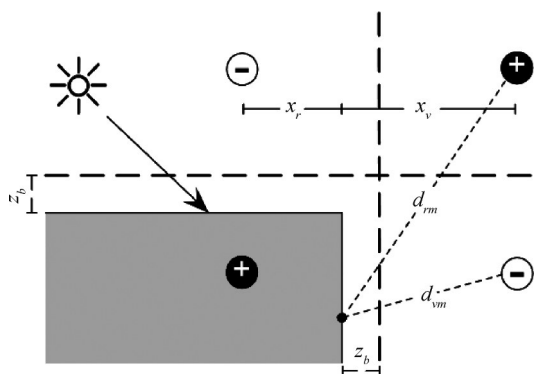


图3 垂直相邻面的四极子配置(Donner等,2007)

Fig. 3 Quadpole configuration for vertical adjacent faces  
(Donner et al., 2007)

(2013)采用了 d'Eon 和 Irving(2011)提出的改进扩散模型,并用蒙特卡洛积分来求解扩散方程,避免了量化和高斯拟合带来的问题。他们还提出了一个新颖且精确的多次散射项,可以与多次散射同时积分。实际计算时,该方法将改进后的扩散模型与光子束追踪结合,其过程与 Donner 等人(2008)的方法类似,但使用的是连续的光子束离散的光子。

基于扩散近似的方法还可以通过相对较简单的参数直观地调控渲染结果,如 Christensen 和 Burley(2015)的工作仅使用反照率以及平均自由程两个参数控制渲染效果。Chiang 等人(2016)希望将蒙特卡洛路线追踪的精确性与扩散近似方法使用的直观参数相结合,他们基于传统的蒙特卡洛路径追踪,利用 Christensen 的公式推导计算单次散射反照率以及消光系数,保证可以通过反照率贴图以及平均自由程直观地控制渲染的效果。该方法在光线追踪过程的距离采样中,根据光线的不同波长进行多重重要性采样,将各波长光线的吞吐量纳入采样计算,当某一波长的吞吐量衰减为0时停止采样,从而有效降低了采样过程中的噪声情况并提高了采样效率。Wrenninge 等人(2017)同样希望能够利用直观的参数控制蒙特卡洛路径追踪的效果,通过使用数值反转的方法来根据期望的表面颜色反推出散射反照率,并用多项式和指数多项式来拟合反转函数。此外,他们还使用了 Dwivedi 采样(Křivánek 和 d'Eon, 2014)和多重重要性采样等技术来改善路径追踪过程中的收敛速度和方差。

### 1.3 基于蒙特卡洛方法的皮肤渲染方法

如前文介绍,相较于严格假设对次表面散射过程进行模拟的扩散近似方法,蒙特卡洛方法具有更高的精度。然而,传统的蒙特卡洛方法(Wang等,1995; Lafortune 和 Willems, 1996; Jensen 和 Christensen, 1998; Pauly等,2000)在处理高度散射的半透明材质时收敛速度非常慢,这一问题在渲染皮肤时同样非常明显。本文重点关注近年来关于提高蒙特卡洛方法收敛速度的研究,更多关于蒙特卡洛方法的内容没有展开,可以参考 Wu 等人(2022)的工作。

目前用于加快蒙特卡洛收敛速度的方法主要采用以下3种方式:零方差随机游走方法、下一次事件估计方法以及参与介质路径指导方法。

#### 1.3.1 零方差随机游走

蒙特卡洛渲染收敛速度较慢的一个主要原因在

于许多输入介质内部的光线在多次散射后截止,仅其中一部分光线可以最终离开介质边界。Křivánek和d'Eon(2014)为此提出了Dwivedi采样技术,利用半无限平板来近似介质的几何形状,并根据平板法线沿入射点方向偏置采样分布,使得采样路径更容易离开介质,从而加快收敛速度并降低方差。但该方法无法解决较薄介质处的透射情况。Meng等人(2016)进一步将偏置方向改成入射照明偏置以及最近点偏置,解决了这一问题。以上方法受限于零方差随机游走的实现难度,仅使用于高度散射的各向同性介质,与皮肤材质间存在一定差异。

### 1.3.2 下一次事件估计

下一次事件估计常用于蒙特卡洛渲染中,在每次光线反弹过程中通过对光源采样以获取直接光照来减少方差,但该方法并不适用于具有折射边界的参与介质,因此无法应用于皮肤渲染。针对这一问题,Hanika等人(2015)基于流形探索的理论,提出了一种通过多个折射表面连接到光源的方法,但该方法依赖于高质量的微分几何框架,对于复杂或者细分的曲面可能存在数值不稳定或者边界顶点求解失败的问题。Koerner等人(2016)通过在体积内部搜寻满足费马原理的点对这一方法作出改进。Weber等人(2017)提出了多点下一次事件估计,显著降低了噪声并进一步提高了收敛的效率,但是该方法不适用于具有边界的参与介质。

### 1.3.3 参与介质路径指导方法

路径指导方法近年来得以应用于表面渲染,通过引导采样过程以加速收敛。Herholz等人(2019)提出了一种结合零方差路径采样理论的路径指导方法,将此类技术扩展到参与介质的体渲染中。该方法的关键在于利用伴随传输解决方案的缓存估计来指导体渲染中的所有采样决策。其中伴随传输解决方案的缓存估计使用k-d树来表示空间分布,同时使用基于vMF(von Mises-Fisher)分布的参数混合模型来表示方向分布。渲染过程中,根据透射率和伴随传输解的乘积对散射距离进行采样,并使用增量采样策略来避免昂贵的累积密度函数采样。他们还根据相位函数和入射辐射率估计的乘积对散射方向进行采样。该方法相较于传统路径追踪具有更高的收敛速度,该方法渲染结果与标准采样方法间的比较如图4所示。

Deng等人(2020)同样将路径指导方法应用于体渲染中。该方法使用SD-tree表示介质中的入射辐射度分布,并结合入射辐射度分布和相位函数,用一种重采样方法来高效地采样出射方向,仅需要几个样本就可以近似出射辐射分布。该方法同样加快了收敛速度,并适用于任何均匀介质,对皮肤渲染也有很好的效果。

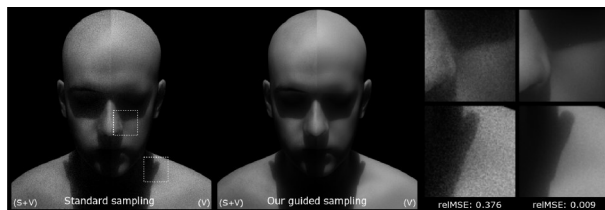


图4 标准采样与路径指导间的比较(Herholz等,2019)

Fig. 4 Comparison between standard sampling and path guiding (Herholz et al., 2019)

## 2 人脸的逆向恢复方法

人脸逆向恢复,又称为人脸捕获,指通过使用相机等拍摄设备采集人脸,并根据相应的采集人脸结果(视频、图像等)对人脸的几何信息和外观进行恢复的过程。其中几何信息通常指人脸的显式几何模型。外观指存储皮肤颜色信息的纹理贴图,这一概念后来被扩展为支持人脸渲染流程所需的全部图像数据,如法线贴图、漫反射反照率以及镜面反射反照率等,又称为材质信息。

对人脸的高精度恢复方法往往依赖于专业的拍摄设备(如高速相机)。这些方法按照是否依赖于专门的照明设备,可进一步划分为主动照明和被动捕获。其中主动照明依赖于专业的照明设备以提高恢复质量,如应用梯度照明来恢复高精度的法线贴图;相对地,被动捕获方法则不依赖于专业照明设备,即使人为提供照明,也仅限于提供均匀光照以减少场景光照对恢复干扰等辅助作用。为摆脱昂贵的拍摄设备以及复杂的恢复流程,基于深度学习的低精度恢复技术作为被动捕获的一个分支,近年来也愈发受到人们的关注。

### 2.1 高精度人脸恢复

#### 2.1.1 主动照明

基于主动照明的方法可以对人脸进行精确恢

复,其特点是通过使用设定的照明设备为采集过程提供光照。这些方法通常需要高精度的摄像设备和复杂、耗时的恢复流程。近年来,该领域的研究除了关注如何提高人脸的捕获精度,还致力于简化捕获设备、缩短捕获时间,并将传统方法与深度学习技术相结合,从而以更低的成本对人脸进行精确恢复。

Debevec 等人(2000)首先提出了 Light Stage,如图 5 所示,将主动照明引入人脸捕获,利用专门设计的照明系统为脸恢复提供光照。该方法通过调整由双轴旋臂控制的聚光灯的位置和方向来照亮人脸,并使用相机从左右两侧捕捉人脸在 2 048 个不同光照方向下的图像。通过每个像素在不同光照条件下的反射亮度和色彩信息构建了人脸的反射函数。



图 5 Light Stage 人脸捕获设备(Debevec 等,2000)  
Fig. 5 Light Stage facial capture system  
(Debevec et al. , 2000)

此外,该方法对捕获的各个光照方向下的信息进行线性组合以表示任意光照,进一步生成相机捕获角度下不同光照条件的人脸重建图像。为进一步生成不同视角下的人脸恢复图像,该方法利用结构光技术获取了人脸的几何模型,并将原始视角下的外观信息投影到几何模型上。结合皮肤反射模型,利用颜色空间分析技术对反射函数中的镜面反射和次表面散射成分进行分离,并据此估计相应的法线方向。最后,根据新的观察角度和法线方向调整镜面反射效果,实现了新视角下的人脸重建。

该方法的照明系统难以模拟局部或近距离光照下的人脸,且由于照明系统中仅包含一个聚光灯,为记录不同光照方向下的人脸,需要等待聚光灯不断

变换位置和角度,整个采集流程的时间接近 1 min。此外,尽管该方法的光照模型中考虑了次表面散射分量,但并未对扩散参数进行建模,因此在高频光源下无法正确恢复次表面散射。

Weyrich 等人(2006)针对 Light Stage(Debevec 等,2000)捕获时间长以及皮肤建模不精确的问题进行改进。他们使用 16 台数码相机和 150 个 LED(light emitting diode)光源构建捕获系统,如图 6 所示,将整个人脸捕获流程所需时间缩短至 25 s。



图 6 包含 16 台数码相机和 150 个 LED 灯的人脸扫描球  
(Weyrich 等,2006)

Fig. 6 The face-scanning dome consists of 16 digital cameras and 150 LED light sources (Weyrich et al. , 2006)

受该方法启发,后续的人脸重建工作普遍采用了多个 LED 灯组成的照明系统。该团队还专门设计了一种接触式的传感测量设备,即皮肤接触探针,用于直接扫描皮肤并基于偶极子扩散近似皮肤的半透明参数,以更准确地恢复皮肤的次表面散射效果。具体的人脸恢复过程中,他们根据不同照明条件下的人脸捕获结果恢复皮肤的总反射率。通过光度立体法(Woodham, 1980),Weyrich 等人(2006)细化了几何形状并计算了高分辨率的面部网格和法线贴图,利用 Lensch 等人(2001)提出的 lumitexels 在 3D 几何结构中记录人脸的反射测量结果。此外,基于人脸镜面反射在某些观察角度上消失的现象,Weyrich 等人(2006)使用 lumitexel 样本的稳定最小值来估计皮肤的漫反射率,并进而分离镜面反射分量,使用非线性优化对镜面反射参数进行拟合,结合皮肤探针恢复的次表面散射参数,实现了对人脸的捕获。尽管该方法的采集时间相较于 Debevec 等人(2000)的方法已经有所缩短,但整个采集过程仍需要 25 s,在这种时间的曝光下,人会不可避免地移动,导致数据

记录错误。

为了进一步减少数据采集时间和数据量,并实现更高精度的人脸恢复, Ma等人(2007)提出了一项重要的方法,即偏振球形梯度照明方法。该方法仅需在4种梯度照明下,根据两种偏振状态对人脸进行采集,仅8幅图像即可有效地分离物体表面的漫反射和镜面反射,并计算出相应的法向量。结合手动选择的粗糙度参数以及全光模式下拍摄的人脸图像作为漫反射和镜面反射率,实现对面脸的重建。在法向量的恢复过程中,该方法同时恢复了用于计算漫反射的漫反射法向量和用于计算镜面反射的镜面反射法向量,并将漫反射法向量分解成RGB三个通道,以更精确地适应复杂散射情况下不同颜色通道的不均匀反射。该方法显著减少了所需的采集数据量和采集时间,并且结合结构光照明技术,实现了高精度的几何细节恢复,整个过程仅需要拍摄13幅图即可使人脸的恢复效果能够达到毛孔级别的介观尺度。然而,该方法仍存在问题:在线性偏振模式下分离漫反射和高光反射时,需要单独调整每个光源上的线性偏振器以确保所有光线经人脸反射到摄像机后都是竖直偏振的。导致仅能在单视角下对人脸进行采集,提供了有限的人脸覆盖范围。

此外,该方法并未对次表面散射效果进行建模。Ghosh等人(2008)为了解决皮肤反射模型过于简化的问题,根据皮肤的3层结构,将人脸的反射效果分为镜面反射、单次散射、浅层皮肤次表面散射以及深层皮肤次表面散射,并分别使用微表面模型、Hanrahan-Krueger模型(Hanrahan和Krueger, 1993)、多极子模型以及偶极子模型进行建模,以更好地拟合皮肤的多层材质效应。但该方法同样采用了Ma等人(2007)的方法对人脸几何进行重建,使用单视角相机对人脸进行采集,影响了反射参数的估计和渲染结果的质量。针对这一问题, Ghosh等人(2011)提出了一种新的经纬偏振模式,将球形梯度照明技术(Ma等, 2007)的采集从单视角扩展到多视角。具体而言,该方法将照明装置的LED灯分为两组交替排列的灯组,一组安装垂直线性偏振片,一组安装水平线性偏振片,保证两种偏振模式在球面上局部正交且沿LED球的Y轴对称。这种对称性允许从接近XZ平面的LED球赤道附近的任何视点对人脸进行采集,从而将采集过程推广到多视角应用。通过使用梯度和额外的多视角采集, Ghosh等人

(2011)的方法提高了采集数据的信噪比。然而,该方法需要将相机水平放置在球面采集系统的赤道上,因此仍然对采集视角有一定的限制。此外,在多视图设置中,完美的分离假设并不成立,该方法相较于Ma等人(2007)的方法在镜面反射和漫反射的分离效果方面存在一定程度的退化。

Graham等人(2012)注意到尽管现有的人脸恢复方法能够捕获亚毫米级的毛孔、折皱等面部特征,但无法从微米级别对皮肤表面的粗糙情况进行捕获,这会对皮肤恢复的镜面反射产生影响。为了解决这一问题并实现微米级精度的皮肤几何恢复,他们首先利用偏振梯度照明,以10微米级的精度对面脸的额头、太阳穴、脸颊、下巴和鼻子等5个区域进行近距离扫描,记录微观细节。随后,他们采用了基于图像类比的约束纹理合成方法(Hertzmann等, 2001),分别将采集到的微观细节与使用多视角球形梯度照明(Ghosh等, 2011)恢复的几何结果相结合,从而将亚毫米级的面部细节增强到了微米级水平。此外,该方法还利用线性插值来平滑合成后的人脸各个区域,并结合一个两叶瓣的镜面BRDF来精确模拟采集样本的高光效果。

Kampouris等人(2018)提出了一种利用二元球面梯度光照对漫反射和镜面反射进行分离的方法,替代了此前工作中使用的偏振技术。他们观察到当利用二元球面梯度光照对面脸进行照明时,面向半球光源的一侧会同时产生漫反射和镜面反射,而背对半球光源的一侧则产生纯漫反射效果。这是因为漫反射较宽的半球积分瓣,即使在皮肤等介电材质的表面背对半球光源时,仍会产生一部分漫反射效果。基于上述观察结果, Kampouris等人(2018)将采集的数据转换到HSV(hue, saturation, value)颜色空间,并假设面向半球光源采集的图像在减去镜面反射分量后,颜色饱和度与纯漫反射的饱和度一致。据此将反照率成功分离为漫反射和镜面反射分量。与Ghosh等人(2011)的方法相比,该方法采集人脸所需的照片更少,多视图采集设置更为简单,预计可以应用于动态人脸捕捉。然而,该方法的镜面反射分离效果实际弱于偏振分离方法,这是因为与偏振球形梯度照明相比,该方法捕获的镜面反射信号较弱。

尽管上述的研究已经可以快速地对人脸进行精确捕捉,但仍无法捕获动态的面部情况。Fyffe

(2010)针对此点作出改进,对以往光度立体法对人脸恢复时的恒定色度限制(Woodham, 1980;Klaudiny等, 2010;Hernández和Vogiatzis, 2010),引入光谱复用技术,利用3个不同光谱分布的光源对场景进行照明,捕获多光谱不同颜色通道下的人脸图像。他们将每个像素的光谱响应结果表示为表面法向和反射率的双线性方程组,并通过引入一个低维线性基来降低方程组的自由度,使用归一化迭代算法对方程组进行求解,实现了单张照片的光度立体恢复。该方法不同于以往的工作,只需要单次拍摄就能完成对人脸的恢复,使得动态人脸捕捉成为可能。然而,该方法仅能恢复法向量和反照率,并没有考虑人脸的反射信息。在随后的研究中,Fyffe和Debevec(2015)进一步将光谱复用技术与极化球形梯度照明技术相结合,搭建了一个偏振RGB的LED球用于提供照明,利用彩色渐变偏振光在单次拍摄中实现了对人脸漫反射、镜面反射和表面法线的恢复。该方法采集时间同样非常短,每帧的曝光时间只有12.5 ms,因此能够实现对人脸表情的动态捕捉。但由于采用了Ghosh等人(2011)的偏振模式,该方法在表面法线恢复方面可能会产生伪影。

除上述研究外,还有一部分研究人员致力于简化人脸采集设备,采用消费级别的设备进行人脸捕获。Fyffe等人(2016)首先提出了一种低成本、简单且快速的人脸捕捉方法,使用现有的摄影硬件进行静态面部外观捕捉。整个捕获系统仅包括24个摄像头和6个闪光灯。该方法通过多视角的颜色空间分析,对漫反射和镜面反射进行分解,并结合光度立体法对法线分别进行恢复。通过使用商业软件PhotoScan,基于多视图立体构建面部的初始几何形状,并根据漫反射法线和镜面法线对几何细节进行细化。由此,该方法能够以低成本的硬件设置,在几乎瞬时的时间内捕获高分辨率的面部反照率和几何,获得了亚毫米级的人脸介观结构细节。但该方法使用的硬件设置需要控制闪光灯在几毫秒内按顺序触发,仍有一定难度。此外,不同于Fyffe之前的单次拍摄捕获方法(Fyffe, 2010;Fyffe和Debevec, 2015),该方法无法实现对动态面部外观的捕捉。

Lattas等人(2020)也在设备方面取得了重要的简化,仅使用4个桌面LCD(liquid crystal display)显示器和8台数码相机就实现了高质量的人脸捕捉。该方法采用了类似于Kampouris等人(2018)的二元

照明技术,但基于所使用的照明实现设备,可将恢复人脸所需捕获的照片减少至4张。此外,他们还结合了光谱复用技术并进行改进,将所需的图像捕捉数量进一步减少到两幅,并且恢复精度几乎与4张照片的恢复结果一致。具体而言,Lattas等人(2020)发现相机的绿色通道对显示屏发出的红色照明相较敏感,并通过光谱编码解决了由此导致的光谱串扰问题,从而实现了高质量的光度测量。此外,该方法通过光谱复用技术,采用了归一化颜色作为新的度量标准,根据逆渲染流程对空间变化的镜面粗糙度进行了恢复。该方法还将相机与显示屏的刷新频率同步,结合光谱复用技术,保证了30帧的有效捕获速率,成功实现了对动态面部的捕捉。

随着深度学习的高度发展和广泛应用,研究人员也开始尝试将深度学习技术与人脸捕获相结合,以简化设备需求、提高采集效率和恢复质量。

Li等人(2021)首次将矫正的高精度多视角人脸图像与深度学习方法相结合,提出了ToFu(topologically consistent face from multi-view)。与传统的多视角重建方法和面部捕捉技术不同,ToFu在深度学习的支持下能够以交互级别的速度进行恢复。在几何恢复方面,Li等人(2021)采用了一种正交均匀网格的采样方法和概率预测法,将几何恢复分为全局粗糙恢复和局部精细化两个连续的过程。全局粗糙恢复使用正交均匀采样网格涵盖整个头部,对于网格内的每个小格,都可以映射到多视角图像的相应特征图中提取特征值,然后通过概率估计网络分析最有可能产生顶点的位置;局部精细化将这一采样网格缩小到每个顶点周围,利用多轮迭代法,每一轮精细优化的基础都是上一轮精细优化的模型表面细分后的结果,采样方式与全局方法相同。在反射贴图恢复方面,ToFu采用了图像转换网络,因为多视角图像本身已经去除了高光,所以从图像中提取的完整贴图可视为漫反射反照率贴图,之后根据漫反射反照率贴图来预测镜面反射反照率贴图和置换贴图。

Liu等人(2022)提出的ReFA(recurrent feature alignment)在几何恢复方面则考虑直接从一个平均模型出发进行多轮迭代,并采用了与ToFu相似的正交网格采样方法,以从多个视角的图像中获取顶点的特征信息。ReFA还使用了GRU(gated recurrent units)网络结构来存储多轮迭代中的共享信息。在

反射贴图恢复方面, ReFA 同样采用了图像转换的网络, 但与 ToFu 不同之处在于, ReFA 考虑了从特征信息贴图来预测漫反射反照率贴图等反射贴图。

Azinović 等人(2023)也探索了将深度学习与主动捕捉人脸相结合的方法, 借助手机这一简易设备以及偏振片技术, 实现了镜面反射与漫反射的分离。他们充分利用了所有的采集数据, 包括照片和视频, 进行几何重建, 以获得更加稳健的结果, 并且在纹理重建方面专注于使用照片, 因为照片能够捕捉更多的细节。此外, 他们结合可微渲染, 采用由粗到细的优化策略, 先后对漫反射和镜面反射进行优化, 以进一步提高重建结果的清晰度。

### 2.1.2 被动捕获

不同于需要高精度设备以及复杂恢复流程的主动照明, 被动捕获通常不需要严格的光照条件, 并且设备要求简单, 所需的成本较低。随着多视角立体重建技术、数据驱动的图像一致性和几何表面平滑等研究的推进, 被动捕获的质量得以提升。其根据所采集人脸数据的形式, 可进一步分为静态和动态捕获分别展开分析。

1) 静态被动捕获。Beeler 等人(2010)首次将被动捕获引入人脸捕获技术, 并将被动捕获的几何质量提升到了与主动捕获相媲美的水平。此前基于多视角重建技术的三维物体重建在细节上有所欠缺, 而该方法在多视角立体重建技术的基础之上, 提出了一种粗糙到精细的迭代优化法。该迭代优化法利用相邻的相机捕获结果, 结合图像一致性约束和表面平滑度约束, 对人脸表面的褶皱细节进行雕刻。然而, 该方法的几何重建过程需要耗费近 20 min, 且该方法仅能恢复皮肤的纹理信息, 不能恢复皮肤的反射信息, 在有高光的情况下重建结果会被扭曲。

由于被动捕获不能控制光照条件, 而漫反射和镜面反射在不利用特殊光照投射的情况下难以分离, 基于被动捕获的人脸反射信息恢复的研究在很长一段时间没有进展。直到 Gotardo 等人(2018)利用皮肤漫反射反照率随时间变化时的一些特殊性质, 建立了皮肤漫反射反照率的子空间, 从而首次利用被动捕获方法恢复了人脸反射信息。他们注意到了血红蛋白含量随时间的变化规律, 发现皮肤漫反射反照率在黑色素浓度固定的前提下, 颜色信息随时间变化数据可在 LAB (lightness, a (green-red

axis), b (blue-yellow axis)) 颜色空间中投影到 LB (lightness, b (blue-yellow axis)) 平面的一条直线上, 并由此建立了皮肤漫反射反照率的子空间, 成功分离了漫反射和镜面反射。

Riviere 等人(2020)则将偏振片分离漫反射和镜面反射的技术引入被动捕获, 通过多个视角下的相机进行单次拍摄图像, 并应用可微渲染, 成功地获得了人脸反射材质贴图。该方法更进一步地使用了可微次表面散射的渲染流程, 使得恢复的反射信息更加真实。

2) 动态被动捕获。Bradley 等人(2010)提出了一种高速的被动捕获方法, 能够实现 30 帧/s 的捕获速率。该方法通过约束相机的拍摄位置, 使用 7 对相机分别拍摄人脸的 7 个区域, 每个区域分别进行高分辨率的几何重建, 并且额外使用一台相机来拍摄人脸的正脸作为参考以更好地恢复几何。为了提供均匀光照, 本方法中还提出了一种人工提供人脸均匀光照的结构化光照设计, 但是该恢复方法并不依赖于该结构化光照。此外, 他们的方法还可以重建毛孔级别的高分辨率皮肤纹理贴图, 捕捉由于演员面部出汗或脸红等原因引起的皮肤外观变化。该方法适用于捕获动态人脸, 但与 Beeler 等人(2010)的方法相比, 重建的几何细节和纹理真实度都有所缺失。Beeler 等人(2011)后续将他们基于静态被动捕获的方法扩展到了动态面部的被动捕获, 利用一个锚点帧串联其他帧在几何上的联系。

后续的方法进一步简化了拍摄条件, Valgaerts 等人(2012)成功地将捕获动态几何模型所需的设备精简至单个双目摄像机。Garrido 等人(2013)则利用个性化的混合形状模型, 对于单目拍摄的录像中的人脸几何进行捕获和跟踪。Shi 等人(2014)将被动捕获技术扩展到自然拍摄的非约束人脸录像, 如网络上的人脸录像。Cao 等人(2015)将单目拍摄的人脸录像的几何捕获速率提高到了实时水平。但這些被动捕获的工作关注几何模型在视频序列上的跟踪, 并不涉及皮肤纹理或材质信息的恢复, 故在此不做过多介绍。

## 2.2 低精度人脸恢复

被动捕获的恢复方法仍然需要一定的设备和拍摄方法。然而, 在现实世界中, 存在大量无拍摄约束的自然拍摄的人脸图像或视频。因此, 该领域的研究重心逐渐转向了研究如何从非约束人脸图像中提

取人脸几何模型和相关的外观信息。基于非约束人脸图像的人脸恢复可以视为被动捕获方法的延伸,其恢复过程不再需要预先知晓光照情况、捕获视角,不再依赖于具体的捕获设备,甚至高清晰度拍摄结果。通常,恢复所需的捕获信息只包括单视角人脸图像、多个视角的人脸图像或一个视频序列,而拍摄环境也不需要受到严格的约束。

Blanz 和 Vetter(1999)对于如何预测基于单幅人脸图像恢复人脸几何模型与皮肤纹理这一问题交出了一份令人印象深刻的答案,他们收集了200幅人脸的几何模型和皮肤纹理的扫描结果,并制成数据集。通过使用主成分分析法(principal component analysis, PCA)将数据集中的人脸几何模型和纹理贴图分别投影到了两个线性空间中,从而可以使用较低维度的向量来表示任何一个人脸几何模型,大大减少了需要的信息表示空间,同时确保了每个向量都对应一个正确且完整的人脸几何模型和纹理。具体而言,对于一个人脸的几何模型  $S$  和面部纹理  $T$ ,它们可以在线性空间中由一对  $m$  维和  $n$  维的向量  $\mathbf{a}$  和  $\mathbf{b}$  表示,其形式为

$$S = \bar{S} + B_s \times \mathbf{a}, T = \bar{T} + B_t \times \mathbf{b} \quad (4)$$

式中,  $S$  代表经 PCA 分析后得到的平均面部模型顶点数据,  $\bar{T}$  是平均纹理贴图,  $B_s$  是面部模型顶点的改变量基底,  $\mathbf{a}$  是每一个改变量的改变程度,  $B_s \times \mathbf{a}$  为相对于平均面部模型顶点的变化情况;类似地,  $B_t \times \mathbf{b}$  表示相对于平均纹理贴图的变化情况。由于该技术可以通过改变参数以对应改变人脸模型的形状和纹理, Blanz 和 Vetter(1999)将此技术称为三维可形变模型(3D morphable model, 3DMM),如图7所示。

在寻找和匹配图像中的人脸方面, Blanz 等人(1999)提出的 3DMM 技术需要依赖于使用者对人脸位置有一个初步的估计,再通过密集地点到点优化来实现。然而随着计算机视觉领域的人脸图像检测的发展,后续的工作可以直接利用人脸图像识别与 landmark 检测技术来辅助参数化模型到图像中的人脸的匹配过程。对于特定的 landmark 位置规范,参数化模型会提供相应位置的顶点的索引。比对从人脸图像检测出的 landmark 的位置和参数化模型投影到图像后的对应 landmark 的位置之间的距离关系,极大地提高了参数化模型匹配的准确性。

随着深度学习的发展,一些研究尝试通过网络

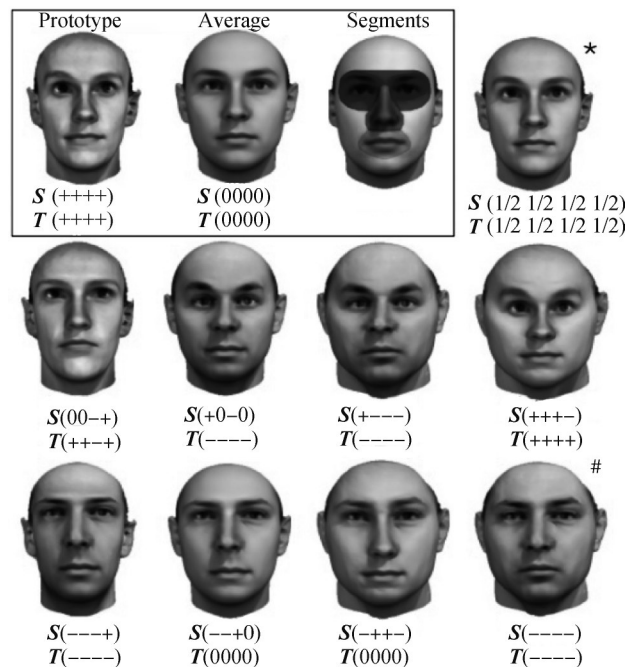


图7 三维可形变模型示意图(Blanz 和 Vetter, 1999)

Fig. 7 Overview of 3DMM (Blanz and Vetter, 1999)

回归的方式直接预测 3DMM 的参数。Tuan Tran 等人(2017)提出使用 101 层的残差网络来直接从非约束的人脸图像中获取 3DMM 的参数。首先通过传统的方式获取非约束人脸图像的相应 3DMM 参数,然后利用这些参数进行高层次的有监督训练。Tewari 等人(2017)提出了一种自监督的网络架构 MoFA(model-based deep convolutional face autoencoder),用于直接从非约束的单视角人脸图像中预测几何模型、反射信息以及光照。他们从非约束图像中预测参数化模型的参数信息和球谐光照的参数等场景参数,并设计了一个可微的参数信息解码模块,将场景参数转化为合成图像,通过比较最终合成图像与真实图像之间的各种损失来实现自监督学习。Genova 等人(2018)在自监督训练中引入了人工合成的人脸图像进行预测,并将预测参数与合成图像所使用的真实参数进行损失计算,从而训练得到更准确的参数预测。同时,他们还考虑了身份信息的差距,使得预测的人脸更加真实。InverseFaceNet(Kim 等, 2018)则考虑基于完全合成的图像进行训练,并提出了一种让合成图像集与真实人脸集分布更加相近的方法。此外,Deep3D(Deng 等, 2019)引入了人脸区域的注意力机制,将非人脸区域的差距从损失计算中剔除。他们还引入了置信度的概念,使用一个置信度网络来衡量重建几何形状的质量,并利用置信

度网络的预测结果对同一个人的多视角图像的重建结果进行整合,从而通过多视角图像来训练网络。

尽管深度学习的引入使得从非约束人脸中恢复几何模型和皮肤纹理变得更加高效,但基于线性空间的 3DMM 模型表示仍然存在一些限制。由于 PCA 本身的降维表达限制,其会过滤掉高频信息,导致其难以很好地拟合特定的人脸细节。为了改善 3DMM 在细节恢复方面的不足,后续研究主要分为 3 个方向:1)更具细节的几何;2)更具细节的皮肤纹理;3)恢复皮肤反射信息。

### 2.2.1 几何细节的恢复

3DMM 这类基于 PCA 的建模方式对于几何细节的恢复限制主要存在两个原因:一是利用 PCA 分析几何模型线性空间所采用的几何模型数据集的规模与多样性总是有限的;二是 PCA 分析本身就会剔除高频的信息,使得诸如毛孔这样的介观尺度细节将不能被保留。本文简称基于 PCA 分析的参数化面部模型为 PCA 参数化模型。

第 1 种基于数据集的限制可以通过更大的真实扫描的几何数据集量、更多的顶点、更细粒度的参数化控制和更完整的几何模型结构来缓解。Paysan 等人(2009)提出了 BFM(basel ace model),从 200 个具有正常表情的个体的头部扫描模型中提取了面部的 PCA 参数化模型。Brunton 等人(2014a)也同样从 100 个人的扫描数据中构建了面部的 PCA 参数化模型。Booth 等人(2016)在数据集大小问题上迈出关键一步,提出了 LSFM(large scale facial model),从 9 663 个人的头部扫描数据中构建了 PCA 参数化模型。除此之外,该工作通过观察高维空间中人脸特征的聚集情况,证明了人脸几何在诸如年龄、性别、人种等特征上具有显著的聚集效应。另外,Cao 等人(2014)提出了 FaceWarehouse,提出了通过额外的表情参数化来控制模型的表情。FaceWarehouse 的数据集包含 150 个不同的个体,每一个个体又有 20 个不同的表情。Brunton 等人(2014b)、Bolkart 和 Wuhner(2015)、Huber 等人(2016)和 Gerig 等人(2018)也相继提出了受形状和表情两种参数控制的 PCA 参数化模型。Dai 等人(2017)提出了 LYHM(liverpool-york head model),从 1 212 个个体的扫描数据中提取了完整的头部 PCA 参数化模型,但是只有形状参数,并且对于眼球也没有进行建模处理。

同年, Li 等人(2017)提出了 FLAME(faces

learned with an articulated model and expressions), 将其参数控制扩展为形状、表情和姿态 3 个参数,如图 8 所示。其中姿态参数控制的是头部、下颏、脖子等关节的相对位置和旋转关系,同时考虑了眼球的建模。Li 等人(2020)提出 ICT-FaceKit,其不仅对完整的头部进行了建模,还对口腔内部结构如牙齿、牙龈,眼睛结构如眼球、睫毛进行了建模。考虑到先前的 PCA 参数化模型在脖子上的拟合仍然具有缺陷,Zhang 等人(2023b)提出了 HACK(head-and-neck),对于脖子长度、旋转等姿态做了细致的建模。

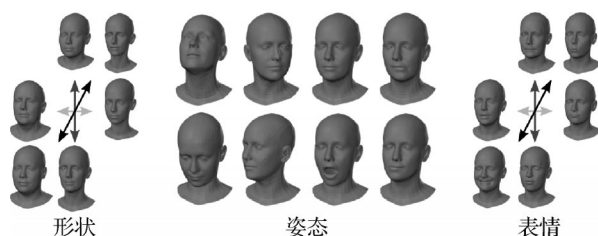


图 8 FLAME 的参数控制示意图(Li 等, 2017)

Fig. 8 Overview of the parameterization of FLAME  
(Li et al., 2017)

虽然基于 PCA 参数化的人头建模方法已经得到了长足的发展,但是由于 PCA 方法本身过滤细节这一缺陷,PCA 参数化人头模型很难表示类似皱纹、痘、毛孔等个性化且细粒度的细节。于是,图形学工作者也开始尝试引入置换贴图来提升几何信息恢复的质量。置换贴图是图形学中经常用来表示几何表面细节的贴图,其通过存储几何表面沿法线的位移来表示几何表面的凹凸程度。

一部分学者基于“暗色即是深,亮色即是浅”的经验性结论,尝试通过深度学习的方法从图像角度进行置换贴图的预测和合成。Huynh 等人(2018)提出了一种恢复几何细节的方法,通过在恢复的基础模型上添加具有毛孔尺度细节的置换贴图来弥补基础模型在细节上的不足。该方法训练了一个从高质量漫反射反照率贴图预测介观尺度几何细节置换贴图的神经网络,如图 9 所示。Chen 等人(2019)基于 Huynh 等人(2018)将从人脸图像预测细节置换贴图这一想法实现在非约束拍摄的人脸图像的恢复工作上。考虑到人脸面部的几何细节包含皱纹这样的寻常细节和毛孔与凸起这样的特征化细节,该方法将几何细节的恢复分为由 PCA 分析控制的线性空间寻常细节恢复和非线性的特征化细节恢复,并分别

通过两个网络对 PCA 参数和细节置换贴图进行预测。FaceScape (Yang 等, 2020) 额外考虑了表情变化时置换贴图在细节上的改变。根据表情变化时细节变化的规律, Yang 等人 (2020) 提出了静态细节和动态细节的区分。静态细节是指人脸毛孔、痣和毛发等不随表情变化而变化的细节; 动态细节则是皱纹等由于表情变化时皮肤收缩和舒张时会动态出现和消失的细节。FaceScape 利用形变贴图记录了某一个表情的几何的顶点相对于正常表情下的几何的相

同纹理坐标的顶点在 xyz 三个坐标轴上的偏移量, 并伴随漫反射反照率贴图传入图像转化网络。Lei 等人 (2023) 沿袭并细化了从图像中预测几何细节的思想。其认为基于 PCA 的模型本身就无法对几何的轮廓做出很好的拟合, 故额外提出了一个恢复几何轮廓细节的办法, 并将几何的恢复分为 3 个阶段: PCA 参数化拟合的基础模型、用形变贴图控制的顶点的移动的轮廓细节以及最终置换贴图控制的微尺度细节。

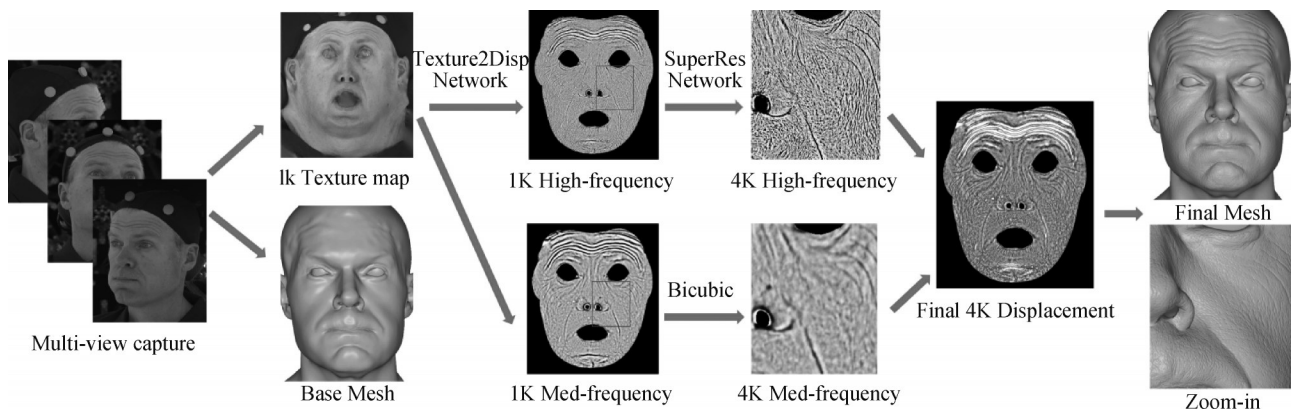


图9 置换贴图的恢复流程示意图 (Huynh 等, 2018)

Fig. 9 Pipeline of displacement map recovery (Huynh et al., 2018)

虽然基于漫反射反照率贴图或皮肤纹理贴图预测几何细节置换贴图可以恢复出毛孔层面的细节, 但是这种方法通常要求网络输入的贴图是完整的, 相反, 如果非约束拍摄图像的人脸区域出现了遮挡物, 皮肤纹理贴图不再完整, 那么预测的几何细节置换贴图也就不再正确。同时, 训练图像转换网络难免离不开真实的数据, 意味着较高的训练成本, 使得该网络很难被其他图形学工作者自行训练。于是, 一些学者尝试研究如何通过自监督来从图像中恢复几何细节。Guo 等人 (2019) 利用优化的策略计算出了人脸深度图的每个像素的深度变化量。虽然不是传统意义上的置换贴图, 但是该工作证明了基于纯优化可以得到包含皱纹的表面细节信息。

Feng 等人 (2021) 提出了一个基于非约束拍摄的人脸图像数据集上的自监督训练架构 DECA (detailed expression capture and animation)。DECA 可以从非约束的人脸图像中直接预测相机参数、基于球谐函数的光照信息、FLAME 模型的形状、表情、姿态参数、反照率贴图编码和置换贴图编码。考虑到表情和姿态对细节置换贴图的动态影响, DECA

将预测的表情参数和姿态参数, 与置换贴图编码一同传入反卷积网络, 最终得到随表情变化的细节置换贴图, 如图 10 所示。随着可微渲染器的实现, DECA 得以实现将图像空间的损失梯度回传到参与渲染的所有三维场景信息, 实现自监督的训练。然而, DECA 对于细节的预测过于粗糙, 将编码解码为图像这种隐式过程使得网络训练存在不确定性, 实验结果中也表明 DECA 恢复的细节置换贴图在不同的人脸上总是预测相似的结果。Chai 等人 (2023) 对于 DECA 的架构做出改进, 提出了 HiFace。HiFace 回到了基于 PCA 的方法, 不同的是, 该方法参照 Raman 等人 (2023) 的方法, 对于表情变化时的几何的拉伸和收缩进行了定量的计算。他们将静态细节置换贴图、皮肤收缩下置换贴图的改变量以及皮肤舒张下置换贴图的改变量分别进行了 PCA 分析。HiFace 的精确细节恢复流程大大增强了几何细节恢复的稳定性和多样性, 但是引入基于 PCA 分析的方法也意味着数据集成本的增加。此外, HiFace 利用 Wood 等人 (2021) 提出的技术生成了大量高真实度的合成图像进行网络训练。

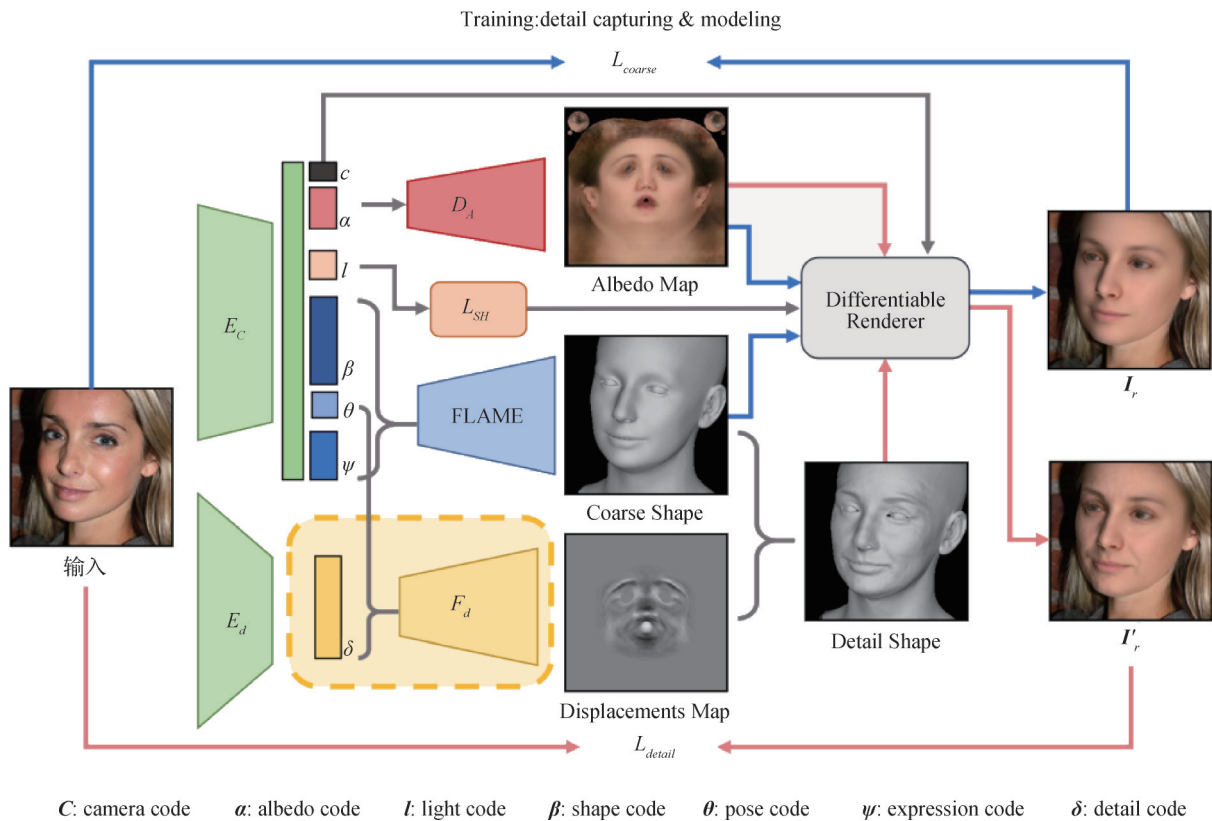


图 10 DECA 的自监督流程示意图(Feng 等, 2021)

Fig. 10 Overview of self-supervision training pipeline of DECA(Feng et al. , 2021)

### 2.2.2 纹理贴图的恢复

非约束单视角图像不能提供人脸面部完整的外观信息,意味着从图像中直接提取的皮肤纹理贴图总是不完整的。图像中的皮肤纹理信息对于恢复高频皮肤纹理贴图不可或缺,因此,一部分学者将目光投向了如何利用图像中直接提取的不完整纹理信息。Saito 等人(2017)同时考虑了 PCA 贴图的完整性和从图像直接提取的不完整的皮肤纹理的高频信息,提出了一种恢复具有毛孔级别高频信息的架构。该方法从不完整的皮肤纹理中提取特征参数,并且与提前构建的完整人脸贴图数据集中的特征参数进行匹配,最终得到完整的特征参数,并将特征参数转化为细节信息附在 PCA 贴图,得到最终的贴图。该方法可以提供很高精度的恢复质量,但是其特征匹配过程需要很大的计算开销。Hu 等人(2017)沿用了 Saito 等人(2017)的补全贴图的方法,用于将图像中被遮挡部分人脸纹理的补全,并提出了一种数字化发型的方法。Tewari 等人(2017)在 MoFA 的框架内添加了一种网络(Tewari 等, 2018),通过从图像中预测一些额外的细节参数,并将这些参数利用一

个网络非线性地映射为反射贴图和几何的细节信息,用于对 PCA 所不能表示的高频信息进行补充。

Deng 等人(2018)则基于一个不同的思路,提出 UV-GAN,希望利用深度生成网络,根据一个残缺的纹理贴图生成完整的纹理贴图。该方法采用了生成对抗网络(generative adversarial network, GAN)(Goodfellow 等, 2014)的训练方式,通过收集真实的 UV 皮肤纹理贴图数据集,随机剔除掉某个视角下的遮挡信息,并将这个残缺的贴图经过一个 U-Net 风格的生成器预测完整的皮肤纹理贴图,并通过全局对抗、局部对抗和身份信息损失来训练生成网络。然而在一些情况下 UV-GAN 生成的皮肤纹理和原始的皮肤纹理之间会有明显的撕裂感。Lin 等人(2020)提出了一种基于 GAN 的更轻量级的恢复架构。该架构额外训练了一个直接从非约束人脸图像中提取高频人脸信息的网络,通过自编码解码网络与粗糙皮肤纹理直接相加后得到最终的纹理恢复结果。为了防止自监督学习过程中粗糙重建的网络与细节预测的网络之间的耦合,该方法利用对偶训练,同时对粗糙重建后的人脸图像和精细重建后的人脸

图像进行对抗训练。Kim 等人(2021)则对于 Deng 等人(2020)的 UV-GAN 的问题进行了改进。该团队注意到 UV-GAN 的 GAN 训练对于撕裂问题不敏感的原因在于判别器对于 UV 贴图图中的撕裂不敏感。其次,UV-GAN 的局部对抗只截取了面部,使得贴图的其他部分被疏忽。得益于可微渲染器, Kim 等人(2021)将对抗训练迁移到渲染结果之中,先前的撕裂问题在渲染结果中会被大幅度放大,判别器对于这些不真实的情况更加敏感。其次,该方法将局部对抗训练从单一的面部区域截取改为了 UV 空间的随机片段截取,使得局部的真实感得以提高。同年, Gecer 等人(2021)借助在人脸图像生成上的优越效果的 StyleGAN2(Karras 等, 2020), 提出了 OSTeC (one-shot texture completion)。其考虑基于优化策略得到 StyleGAN2 的隐空间编码, 利用 StyleGAN2 生成 5 个视角下的人脸图像, 并将多个视角的皮肤纹理信息通过图像混合, 将各个视角下提取的皮肤纹理的正确区域混合到同一张贴图中作为最终的输出。

Gecer 等人(2019)改变了 GAN 在之前的工作中只用来补全贴图的做法, 而将 GAN 作为一种全新的纹理空间承载工具, 提出了基于预训练 GAN 的单视角图像逆向恢复框架 GANFIT (generative adversarial network fitting), 如图 11 所示。GANFIT 摒弃了利用 PCA 线性空间表示皮肤纹理的方法, 转而采用 GAN 所训练得到的隐空间表示皮肤纹理。与 PCA 相似, GAN 也可以将一个向量唯一地映射为一幅图像, 本

文通常称这个向量为隐空间编码; 与 PCA 不同的是, 相比于线性空间通过线性插值得到数据集中不存在的图像, GAN 可以通过隐式地学习图像的分布, 生成数据集中所不包含的图像。而且实验表明, GAN 生成的图像往往更具有细节, 这是由于 GAN 训练的特点导致的, 并且 GAN 还支持根据给定的图像逆向推断隐空间编码, 这就使得 GAN 成为替代 PCA 的更好的方法。相比于先前工作中只考虑像素级别的损失, GANFIT 同时引入了身份损失来增强恢复的人脸外观和提供的图像的一致性, 使得恢复的纹理更加逼真。Gecer 等人(2022)随后又提出了 Fast-GANFIT, 改善了先前 GANFIT 拟合时出现的容易陷入局部最优解导致的不稳定现象。Fast-GANFIT 添加了 3 个深度神经网络用于从传入的图像直接预测 GANFIT 结果, 既可以作为 GANFIT 优化结果的直接估计, 又可以作为 GANFIT 优化的初始值, 使得待优化的参数更容易且更快地收敛到全局最优解。GANFIT 的训练数据集限制了 GANFIT 的拟合范围, 使得 GANFIT 总是预测出偏暗的皮肤纹理。受到 GANFIT 和 OSTeC 二者的启发, Bai 等人(2023)提出了 FFHQ-UV (Flickr-faces-HQ-UV), 期望通过 FFHQ (Karras 等, 2019) 这一庞大的高质量自然拍摄的人脸图像数据集, 制作相应的高质量皮肤纹理贴图数据集 FFHQ-UV, 并训练出类似于 GANFIT 的拟合网络。得益于 StyleFlow (Abdal 等, 2021) 与 InterFaceGAN (Shen 等, 2020) 的提出, FFHQ-UV 得以通过隐

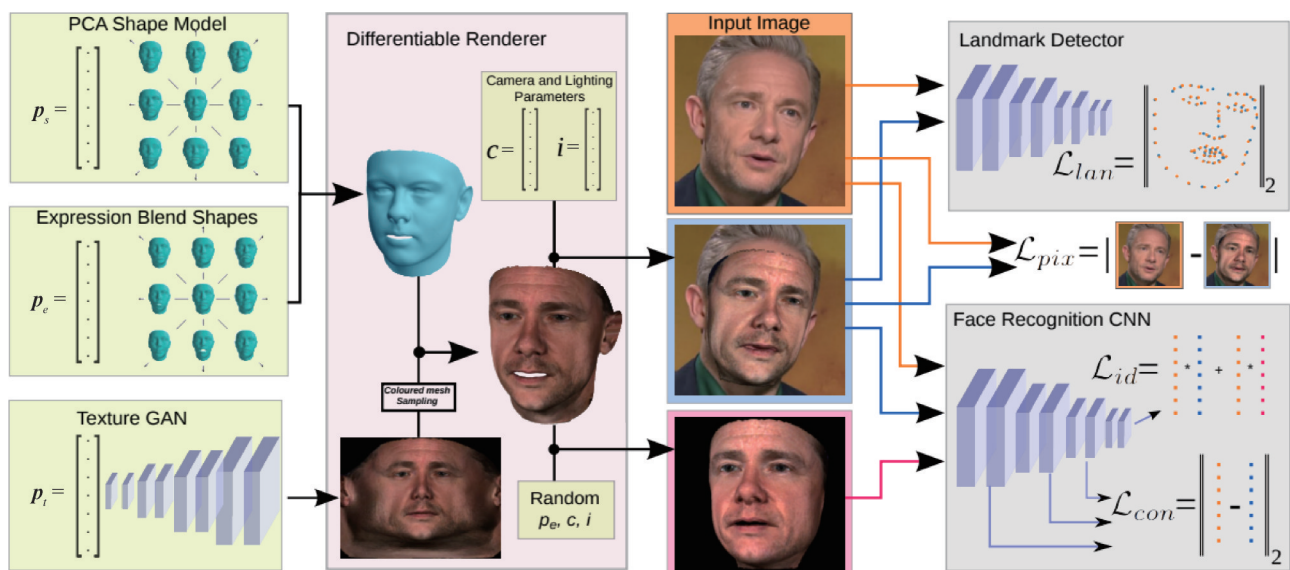


图 11 GANFIT 恢复流程示意图(Gecer 等, 2019)

Fig. 11 Overview of GANFIT recovery pipeline(Gecer et al. , 2019)

空间编辑的方式,利用StyleGAN2直接生成多个固定视角下的相同人脸的图像。通过固定的3个视角的人脸图像,FFHQ-UV可以提取并得到面部的皮肤纹理贴图,并通过与一个模板完整漫反射反照率贴图进行泊松图像编辑(Pérez等,2003),最终得到完整且色调一致的反照率贴图。

### 2.2.3 人脸材质信息的恢复

大量工作研究了从非约束单视角人脸图像恢复高质量皮肤纹理的方法,并且在渲染管线中充当漫反射反照率数据进行渲染。但是皮肤纹理富含光照效果、如着色效果、高光、次表面散射等,并非正确的漫反射反照率信息。随着基于图像恢复人脸模型与纹理的技术的不断发展,图形学领域的学者们也不再满足于只恢复人脸的皮肤纹理,而是想要恢复皮肤的反照率信息等反射信息。反射信息的恢复是比较困难的,这是由于恢复反射信息是一个病态问题。通常,人们希望恢复的人脸皮肤反射信息包括漫反射反照率贴图、镜面反射反照率贴图、细节法向量贴图、粗糙度贴图等,统称为反射材质贴图,这些贴图通常与主流的实时渲染管线相兼容。

许多图形学工作者采用真实数据的有监督训练来使得网络最终能够预测正确分布的反射贴图。Yamaguchi等人(2018)提出了一种从非约束图像中恢复人脸几何模型、反射材质贴图以及两种不同层次的细节置换贴图的框架。其基于PCA参数化模型拟合人脸几何模型,并根据拟合的模型与视角将图像中的人脸纹理提取到UV材质贴图中。该工作的核心思想是采用图像到图像的转化,将提取到的图像中的人脸纹理转化为想要的反射贴图和置换贴图。他们基于Ghosh等人(2011)提出的主动捕获技术,构造了一个人脸反射贴图数据集,并预训练了4个网络,利用皮肤纹理分别对3张反射贴图和1张置换贴图进行预测。ICT-FaceKit(Li等,2020)同样沿袭了图像空间转换的思想,根据真实的反射数据训练了从漫反射反照率贴图预测其他反射材质贴图和置换贴图的网络。Lattas等人(2020)提出了AvatarMe,其同样采纳了图像到图像的转化的思想,希望能够利用面部纹理贴图,通过图像转化的网络如Pix2PixHD(Wang等,2018),从GANFIT得到的皮肤纹理直接推断人脸反射材质贴图。Lattas等人(2021)进一步提出了AvatarMe++,将先前AvatarMe中的4个图像转换网络合并为一个图像转换网络,

从而保证4种反射贴图的一致性。AvatarMe++中额外提出了一个基于Pytorch3D(Johnson等,2020)的支持次表面散射的可微渲染器,并利用网络预测自遮挡阴影来弥补PyTorch3D可微渲染器不支持自遮挡渲染的空缺。

Lattas等人(2023)随后提出了FitMe,借鉴了MaterialGAN的结构,提出了一种人脸皮肤反射材质的生成—拟合网络:利用一个预训练的BRDF-GAN将一个隐空间编码映射为皮肤反射材质贴图。整个架构想要恢复的是隐空间编码、PCA参数化模型的形状参数与表情参数,以及相机参数与光照参数,并通过正向可微渲染得到人脸图像,与目标图像进行损失计算,从而优化出上述5个目标参数。该方法基于先前的AvatarMe++框架,将MimicMe数据集(Papaioannou等,2022)中的皮肤纹理贴图转化为相应的反射贴图,并通过基于直方图匹配的图像算法增广了漫反射贴图的色调,使得数据集更加具有多样性,训练得到的BRDF-GAN能更加胜任多种皮肤色调的生成工作。

对于从非约束人脸图像中恢复材质信息的病态问题,Smith等人(2020)提出了AlbedoMM,给予了解决该问题的另一个思路。其沿袭了3DMM参数化几何模型的思想,提出了构建参数化漫反射反照率贴图和参数化镜面反射反照率贴图的完整方法论。由于不同模型的纹理映射规则不同,不同艺术家设计的模型之间的贴图是无法相互使用的,而该工作的方法论适用于将开源数据集如3DRFE(3D relightable facial expression)(Stratou等,2011)中的固定视角下的漫反射反照率信息和镜面反射反照率信息,或是按照Smith等人(2020)提出的方法自行采集的固定视角反射信息,提取到预期的参数化的材质空间中,并且给出了填补贴图上由于遮挡问题导致的空缺的方法。参数化的方法是传统的PCA方法。参数化漫反射反照率贴图和镜面反射反照率贴图,可以约束漫反射反照率和镜面反射反照率信息在合理的范围之内,虽然PCA分析会导致高频信息的丢失,但是AlbedoMM可以用于粗糙程度的反射信息恢复,作为后续精细化恢复高频漫反射信息和镜面反射信息的基础。

早年的图形学研究者曾经尝试过不利用任何反射信息上的先验来通过纯优化进行人脸的逆向渲染。Aldrian(2013)提出了一种基于纯优化的方法,

创造性地将 RGB 颜色映射到 SUV 颜色空间(Zickler 等,2008)中,在知道高光颜色的前提下,这一映射可以使得 UV 通道中的颜色信息在视觉上只包含漫反射信息。虽然 Aldrian 利用这一映射方法得以实现漫反射和镜面反射的粗糙分离,但是该映射方法是一个纯数学的映射,只是在视觉上剔除了高光。随着 AlbedoMM 的提出,Dib 等人(2021a)提出了基于可微光线追踪渲染器的纯优化逆向恢复框架。该逆向恢复架构分为两个阶段,阶段一首先配合 Landmark 信息优化相机参数和初始的 PCA 参数化模型参数,随后微调 PCA 参数化模型参数和相机参数,并同时优化 AlbedoMM 粗糙的反射信息和光照信息;阶段二采用可微光线追踪渲染器,取消参数化反射信息的限制,基于阶段一恢复的反射信息结果优化反射材质贴图。光线追踪渲染器相比于 PyTorch3D 等基于可微光栅化的渲染器,可以更好地处理阴影问题。Dib 等人(2021b)利用图像编码网络预测 PCA 参数化模型参数、光照参数、相机视角以及 AlbedoMM 参数,并将这些参数转化为 3D 资产后经过可微光线追踪渲染器渲染为人脸图像,与目标图像中的人脸区域进行损失计算以实现自监督训练流程。Dib 等人(2023)的这项工作证明了从非约束人脸图像预测贴图高频细节信息的可能性,并提出了更加完善的自监督恢复框架 S2F2 (self-supervised high fidelity face reconstruction)。

注意到先前的工作没有考虑 PCA 参数化模型在高频几何信息上的不足,S2F2 增加了从图像中预测几何高频信息的自编码解码网络。除了在细节上更进一步的恢复设计,S2F2 还提出多阶段训练和对偶训练来对自监督训练的稳定性进行优化,S2F2 的恢复结果与真实图像十分相近。

另外,从非约束人脸图像中恢复人脸外观信息的工作经常会出现预测的皮肤纹理或漫反射反照率色调与真实数据不匹配。Hu 等人(2013)提出可以利用人脸的反射信息的对称性这一先验信息来防止极端高光情况下的光照对于人脸造成影响,但是仅仅通过对称性无法解决环境光颜色带来的均匀的颜色扰动。Egger 等人(2018)则通过学习一个环境光的先验信息来约束逆向恢复过程中的环境光颜色,但是仍然会出现将皮肤颜色恢复到光照颜色上的问题。Feng 等人(2022)针对皮肤色调和光照颜色的分离问题,提出了 TRUST (towards racially unbiased

skin tone estimation) 框架。TRUST 框架假设同一幅图像中的光照信息低频且一致,场景中的光照信息、不同人物之间的光照信息应该是一致的,那么就可以同时考虑非人脸区域和纯人脸面部光照信息提取光照信息。

一些图形学者将从图像恢复的皮肤外观信息从传统的反射信息扩展到皮肤的生物物理信息。Aliaga 等人(2023)从一个新颖的角度出发,将逆向渲染扩充到了基于皮肤的生物物理特性的渲染流程中,并尝试这类流程的参数的逆向恢复。基于皮肤生物物理特性的渲染方法是一种体渲染方法,对于皮肤内的结构进行建模,并将其中化学物质的体积分数和比率作为逆向渲染恢复的目标皮肤属性。该方法利用一个全连接的神经网络替代传统的体渲染流程,从输入的皮肤 RGB 颜色值直接预测自遮挡信息和皮肤属性,以及超光谱反射信息。该网络的训练数据由传统体渲染流程制备,从而实现有监督的学习。

#### 2.2.4 基于深度学习从多视角图像或视频序列中恢复人脸外观

从多视角人脸图像中恢复人脸外观信息相较于从单视角人脸图像中恢复要简单得多,这是由于多视角下的人脸图像提供了足够的几何信息,通常在多个视角的人脸轮廓的拟合中就可以恢复出高真实度的人脸几何模型。同时多视角的图像也可以覆盖完整的头部外观信息,进而获取到完整的皮肤纹理贴图。同时,由于多视角图像在光照信息上或视角上的变化能更好地避免镜面反射信息的耦合问题,在材质的逆向恢复上具有优势。

Ichim 等人(2015)提出了第 1 个基于智能手机拍摄的录像恢复完整人头几何模型与漫反射反照率贴图的完整框架。视频提供的多视角人脸图像使得 Ichim 等人(2015)采用更高精度的多视角立体重建法进行头部模型的重建。在漫反射反照率贴图的恢复方面,该研究通过计算光照效果,从提取的人脸皮肤纹理贴图中剥离光照效果等着色信息,并利用一个模板纹理来填充缺失的贴图部分,从而得到近似的漫反射反照率贴图。由于 Ichim 等人(2015)提出的方法基于纯计算的形式,时间开销大,并且拍摄过程中人头的意外移动很容易导致重建结果出现错误。Piotraschke 和 Blanz(2016)随后提出了一种利用 3DMM 在同一个人的多个不同时间地点拍摄的单

视角图像恢复几何模型与外观的方法。该方法提出了一种筛选重建质量的策略,筛选出重建质量较好的结果。其将人脸几何模型分为眼部、鼻子、嘴部与其余部分,并将筛选出来的结果在每个部分分别进行整合,最终得到一个高质量的结果。皮肤纹理贴图则是基于最终得到的几何重建结果进行图像空间的提取。借助卷积神经网络和可微渲染器,Wu等人(2019)提出了MVF-Net(multi-view 3D face-Net)。MVF-Net从同一个人的多个视角的图像预测一致的形状与表情参数,以及各自的视角参数。皮肤纹理则是从多个视角下提取得到。该方法借助300W-LP数据集对卷积网络进行预训练,然后再基于可微渲染器,通过自监督训练进一步优化网络参数。Bao等人(2022)提出从手持拍摄的RGB-D录像进行多视角人脸重建。对于给定的RGB-D录像,该团队提出了一种对视频帧进行筛选的方法,并将筛选结果作为一组多视角人脸图像进行后续的恢复工作。在几何恢复方面,为了解决基于PCA的模型表示的限制,他们提出了一种方法,将人脸的眼部、鼻部、嘴部3个区域进行替换、扰动、镜像化等增广处理,最终合成了100 000个面部模型,大大增强了PCA线性空间的拟合能力。在漫反射反照率贴图方面,该方法将贴图的区域分为8个子区域,并且每个区域都用PCA分析进行参数化表示,如图12所示。细节法向量贴图则是通过预训练的图像空间转换网络得到。其次,Bao等人(2022)另外训练了一个精细化网络,通过图像转化的方法向漫反射反照率贴图和法向量贴图添加毛孔等更多的细节。



图12 皮肤纹理的区域示意图(Bao等,2022)

Fig. 12 Overview of the different face texture region  
(Bao et al., 2022)

Wang等人(2023d)将太阳当作Light Stage的照明设备,提出了SunStage,对人脸几何与材质信息进行恢复。Zheng等人(2023)则提出了NeuFace,将人脸几何与反射材质信息的恢复扩展到了神经渲染领域,并采用了当前最受欢迎的神经渲染方法:神经辐射场(Mildenhall等,2022)。其利用隐式有向距离场表示几何,并结合隐式BRDF网络等基于多层感知机的轻量级神经网络,实现了对场景参数有较好解耦的渲染流程。NeuFace训练过程只需要同一个人的多个非约束拍摄视角下的图像,即可得到神经渲染下的几何表示与材质信息表示,如漫反射、镜面反射。虽然当前以神经辐射场为代表的神经渲染架构还未能很好地在传统渲染管线和实际应用中落地,这一方面的研究将仍然是接下来一段时间的热门。

除上述工作外,基于视频驱动的动态人脸构建也取得了重要进展。Zhao等人(2024)同样利用神经辐射场,首次实现了高精度且高稳定性的动态人脸数字化身重建。他们的研究通过从输入视频中每帧提取显式3D几何信息作为条件输入,以实现目标人脸的表情驱动。但不同于其他神经辐射场工作,该文章认为采用正面及左右侧视图的视图结构相较于传统的三正交结构更适合人脸的隐式表示。此外,该方法还提出了一个可学习的嵌入向量以保证人脸在多视角下的几何一致性,进一步提高了动态人脸的稳定性。Wang等人(2023a)的工作针对先前研究中仅聚焦人脸面部细节而忽略背景和头发等非面部元素的问题作出改进,通过将人脸图像拆分为面部、非面部前景以及背景3个区域以分别进行自适应的调整。此外,他们的工作还提出了StyleU-Net,通过结合UNet、StyleGAN以及时间编码的优点进行视频学习,从而实现了高质量的人脸视频生成。但需要注意的是,此类基于视频驱动的动态人脸构建工作并不能恢复用于人脸真实感渲染的纹理、网格等资源,因此并不是本文的重点,故不再过多赘述。

### 3 未来展望

随着真实感人脸领域的不断发展,其相关技术已在众多领域得到了广泛的应用。尽管如此,这一技术领域仍面临诸多挑战和问题,如何解决这些问题也同样指引着该领域未来的发展方向。正向渲染

方面,现有的工作往往使用纹理贴图表示人脸的材质特性,忽略了皮肤作为生物材质独特的成色原理;此外,深度学习技术的迅猛发展使得将其与现有渲染技术结合的探索变得愈发关键。逆向恢复方面,基于数据的人脸恢复方法往往受限于缺乏高质量的开源数据集,同时在皮肤毛孔级别的细节建模和恢复方面仍有较大的提升空间;此外,将逆向恢复与基于文本的生成工作相结合,具有巨大的潜力与应用场景。

### 3.1 对皮肤生物物理学属性的追求

现实世界中的人脸皮肤内部结构是十分复杂的:皮肤从外到内包含角质层、表皮层、真皮层,真皮层内还分布着包括肌肉、软骨、脂肪等组织。另外,表皮层内的黑色素含量和分布、真皮层内血红蛋白的分布等也是构成人脸皮肤外观的重要直接因素。但现有的方法经常将皮肤假设为各向同性的均匀介质,基于扩散近似的方法还会进一步将皮肤表面假设为厚度无限的平面。本文认为对于如此复杂的皮肤结构,如何进行准确的建模以实现更接近真实的皮肤效果将成为一个颇具潜力的研究方向。

在正向的真实感渲染中,一些研究(Donner和Jensen, 2006, 2008; Wang等, 2023b)注意到了皮肤的复杂性质,并尝试建立皮肤内生物物质成分与皮肤外观间的联系,使用血红蛋白及黑色素含量等参数控制皮肤渲染效果。还有一些研究(Donner等, 2008; Ghosh等, 2008; Jimenez等, 2010; Wang等, 2023b)为了更贴合皮肤的实际材质,将皮肤视为异质化介质进行处理。

逆向渲染方面,同样有研究人员考虑到了皮肤的生物物理属性对实现皮肤效果的影响。Aliaga等人(2023)的工作展示了基于全连接神经网络对皮肤内生物物质成分参数的恢复方法。然而他们对于皮肤的属性进行了简化,例如各层的折射率一致且均为各向同性的均匀介质、忽略真皮层皮肤组织并认为真皮层无限深、忽略角质层等。此外,该方法在自遮挡和黑色素及含氧量引起的深色耦合问题上并没有给出合理的分离方案。该方向仍需进一步的推进。

基于上述研究,未来的工作有望通过还原人脸皮肤的生物物理性质,进一步实现更为真实的皮肤材质。

### 3.2 人脸的渲染与深度学习结合

随着深度学习的发展,一些研究正在探索如何将深度学习与人脸渲染相结合,进而避免复杂的光线追踪,以对渲染过程进行加速。

此方面研究中,基于神经网络的材质表达方法目前得到了广泛的关注,已分别在多层材质(Fan等, 2022)、双向纹理函数(bidirectional texture function, BTF)(Rainer等, 2019, 2020; Sztrajman等, 2021; Fan等, 2023)以及空间变化双向反射分布函数(spatially varying bidirectional reflectance distribution function, SVBRDF)(Zeltner等, 2023)等表面材质建模中得到了应用。但以上方法仅能表达入射光和出射光在相同位置的材质,目前还无法很好地处理皮肤中包含的次表面散射。在未来的研究中,将神经网络的表达引入次表面散射中,以避免渲染过程中的多次散射,对提高人脸渲染的速度具有一定意义。

### 3.3 材质数据集的开源与增广

基于深度学习的人脸材质恢复方法对真实数据集有很强的依赖。然而,目前可用的开源数据集中很少包含完整的材质信息。其中一方面原因在于材质信息的珍贵性,其高昂的获取成本使得现有的研究团队不愿将数据集免费共享。另一方面,考虑到道德与隐私问题,研究团队通常不会公开包含个人身份信息的数据集,而材质数据集往往包含大量的个人身份信息。因此,目前可用的包含完整材质信息的数据集非常有限,而现有的数据集通常无法提供完整的材质贴图。唯一的例外是3DRFE(3D relightable facial expression)(Stratou等, 2011)数据集,它提供了一些包含人脸材质信息的图像,但是这些图像都是从固定视角拍摄的。

为了能够克服以上问题,更好地实现对材质数据集的开源,RealFaceDB(Lattas等, 2020)等人采取了一种将材质贴图分割成片段的方法,取出其中包含强身份信息部位(如眼睛、嘴唇)的部分,然后随机打乱。此外,对于有限的材质数据集,进行高度可信的数据增强也是一种提升材质数据集容量的方法。FitMe(Lattas等, 2023)提出了一种使用直方图匹配法来增加皮肤颜色多样性的方法,同时确保了个人信息的一致性。

### 3.4 面部毛孔层面细节的建模和可控恢复

人脸皮肤的细节包括宏观结构如皱纹、痘疤、肌肉凹陷等,以及更微观的细节,如毛孔和毛孔周围的

纹路。对毛孔的重建效果直接影响了皮肤的恢复质感与真实性,然而,在现有的人脸恢复工作中,对于毛孔级别的细节建模并不足够精确。一方面,一些利用图像转换的方法(Huynh等,2018;Chen等,2019;Yang等,2020)尽管可以直接从皮肤纹理预测高频的置换贴图,但是这些方法通常较为粗糙、缺乏可控性,且成本过高;另一方面,在基于自监督恢复几何细节的工作中,DECA(Feng等,2021)只能够恢复出皱纹等宏观层次的细节。HiFace(Chai等,2023)在皱纹细节层面具有更高的可控性,但仍然无法有效恢复毛孔级别的细节。

为了实现可控的毛孔几何细节恢复,应该需要对毛孔的分布情况、毛孔的疏密程度、毛孔的大小等均进行建模,并且应该考虑到人脸表情变化时毛孔分布随着褶皱的变化。Nagano等人(2015)曾研究了毛孔细节随着人物表情变化的模拟,通过对几何的拉伸和收缩进行建模,然后使用滤波器处理置换贴图以模拟皮肤拉伸和收缩引起的毛孔变化。最新的研究中,Weiss等人(2023)采用了一种基于图结构的方法来合成皮肤毛孔和微纹理。该方法将毛孔表示为点,微纹理表示为边,结合传统的几何捕捉技术,实现了非常逼真的恢复效果。本文认为这种程序化生成介观尺度细节的方法将进一步推动毛孔几何细节的逆向恢复研究。

### 3.5 基于文本描述的恢复与生成

以往的人脸恢复工作通常依赖于给定的人脸图像或视频序列,这可以视为是通过图像描述来预测或还原人脸信息的方法。然而,随着计算机视觉和自然语言处理领域的不断发展,一些建立二维图像与语言文本之间联系的模型相继提出,如CLIP(Radford等,2021),Stable-Diffusion(Rombach等,2022)等。这些模型的出现使得比较二维图像与语言文本之间的差距成为可能,也让基于文本描述的人脸恢复与生成工作成为该领域新的发展方向。

例如,RODIN(roll-out diffusion network)(Wang等,2023c)基于合成的多视角人脸数据集,率先将扩散模型引入3D人脸生成,通过三维感知扩散,实现了文本驱动的较高质量数字人脸生成。ClipFace(Aneja等,2023)尝试使用CLIP模型来评估皮肤纹理恢复结果与文本描述之间的损失,从而实现文本驱动的风格化皮肤纹理生成。Wu等人(2023)构建了一个文本标注的高质量3D人脸数据集Describe3D,

提出了匹配具体描述和抽象描述的两阶段生成框架。TG-3Dface(Yu等,2023)引入文本—人脸跨模态对齐模块,促进了生成人脸和输入文本之间的语义一致性。与上述工作不同,DreamFace(Zhang等,2023a)更进一步建立了一整套从几何模型恢复、细节雕刻到多通道材质生成的文本驱动生成工作流程。目前,基于文本描述的恢复和生成工作是一个崭新的领域,充满了潜力与挑战。

## 4 结 语

数字人脸作为当今互联网领域备受关注的方向,经过多年的发展与应用,取得了重要的进展与突破。本文针对数字人脸的相关技术,分别从正向的人脸真实感渲染以及逆向恢复外观两个方面对其发展历程进行了全面总结。

1)正向渲染。实现人脸皮肤的真实感渲染一直是渲染领域的一项重大挑战。为了实现皮肤这一复杂半透明介质的质感,从早年的标准蒙特卡洛方法发展到基于扩散方法及近年的蒙特卡洛加速收敛研究,皮肤的实现质感及渲染速度都有了很大提高,人们对皮肤的结构也有了更深的认识。本文详细介绍了基于扩散近似的方法和基于蒙特卡洛方法的方法。对基于扩散近似的方法根据扩散剖面的改进、次表面散射的实时实现以及结合蒙特卡洛方法的混合方法3个方面分别进行了详细的介绍以及复盘,分析了各方法对应的优点以及面临的问题。此外,本文还具体介绍了零方差随机游走、下一次事件估计以及参与介质路径指导等方法是如何在人脸渲染过程中加速蒙特卡洛收敛的。

2)逆向恢复。人脸逆向恢复是数字人脸生成领域的关键领域之一,经过多年的不懈努力,该领域的研究已经从早期的主动捕获方法与被动捕获方法逐步发展到如今与深度学习相结合的低精度方法。这些方法在恢复速度、恢复质量和成本控制方面都取得了显著进展。在逆向恢复部分,本文对主动捕获方法、被动捕获方法以及基于深度学习的低精度方法进行了系统性的论述,并对低精度方法在几何细节恢复、真实皮肤纹理恢复以及材质信息恢复等3个子领域进行了详细介绍。最后,本文对真实感人脸领域的未来研究方向进行了深入分析和展望。

希望本文能帮助研究人员了解人脸真实感渲染

以及逆向恢复领域的发展状况,并能对研究人员的后续工作有所启发。

## 参考文献 (References)

- Abdal R, Zhu P H, Mitra N J and Wonka P. 2021. StyleFlow: attribute-conditioned exploration of styleGAN-generated images using conditional continuous normalizing flows. *ACM Transactions on Graphics*, 40(3): #21 [DOI: 10.1145/3447648]
- Adrian O and Smith W A P. 2013. Inverse rendering of faces with a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(5): 1080-1093 [DOI: 10.1109/TPAMI.2012.206]
- Aliaga C, Xia M Q, Xie X, Jarabo A, Braun G and Hery C. 2023. A hyperspectral space of skin tones for inverse rendering of biophysical skin properties. *Computer Graphics Forum*, 42(4): #14887 [DOI: 10.1111/cgf.14887]
- Aneja S, Thies J, Dai A and Niessner M. 2023. ClipFace: text-guided editing of textured 3D morphable models//Proceedings of ACM SIGGRAPH 2023 Conference. Los Angeles, USA: Association for Computing Machinery: #70 [DOI: 10.1145/3588432.3591566]
- Azinović D, Maury O, Hery C, Nießner M and Thies J. 2023. High-res facial appearance capture from polarized smartphone images//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 16836-16846 [DOI: 10.1109/CVPR52729.2023.01615]
- Bai H R, Kang D, Zhang H X, Pan J S and Bao L C. 2023. FFHQ-UV: normalized facial UV-texture dataset for 3D face reconstruction//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 362-371 [DOI: 10.1109/CVPR52729.2023.00043]
- Bao L C, Lin X K, Chen Y J, Zhang H X, Wang S, Zhe X, Kang D, Huang H Z, Jiang X W, Wang J, Yu D and Zhang Z Y. 2022. High-fidelity 3D digital human head creation from RGB-D selfies. *ACM Transactions on Graphics*, 41(1): #3 [DOI: 10.1145/3472954]
- Beeler T, Bickel B, Beardsley P, Sumner B and Gross M. 2010. High-quality single-shot capture of facial geometry//Proceedings of ACM SIGGRAPH 2010 Papers. Los Angeles, USA: Association for Computing Machinery: #40 [DOI: 10.1145/1833349.1778777]
- Beeler T, Hahn F, Bradley D, Bickel B, Beardsley P, Gotsman C, Sumner R W and Gross M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics*, 30(4): #75 [DOI: 10.1145/2010324.1964970]
- Blanz V and Vetter T. 1999. A morphable model for the synthesis of 3D faces//Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA: ACM Press/Addison-Wesley Publishing Co.: 187-194 [DOI: 10.1145/311535.311556]
- Bolkart T and Wuhler S. 2015. A groupwise multilinear correspondence optimization for 3D faces//Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 3604-3612 [DOI: 10.1109/ICCV.2015.411]
- Booth J, Roussos A, Zafeiriou S, Ponniah A and Dunaway D. 2016. A 3D morphable model learnt from 10,000 faces//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 5543-5552 [DOI: 10.1109/CVPR.2016.598]
- Borshukov G and Lewis J P. 2005. Realistic human face rendering for "The Matrix Reloaded"//Proceedings of ACM SIGGRAPH 2005 Courses. Los Angeles, USA: Association for Computing Machinery: #13 [DOI: 10.1145/1198555.1198593]
- Bradley D, Heidrich W, Popa T and Sheffer A. 2010. High resolution passive facial performance capture//Proceedings of ACM SIGGRAPH 2010 Papers. Los Angeles, USA: Association for Computing Machinery: #41 [DOI: 10.1145/1833349.1778778]
- Brunton A, Bolkart T and Wuhler S. 2014a. Multilinear wavelets: a statistical shape space for human faces//Proceedings of the 13th European Conference on Computer Vision—ECCV. Zurich, Switzerland: Springer: 297-312 [DOI: 10.1007/978-3-319-10590-1\_20]
- Brunton A, Salazar A, Bolkart T and Wuhler S. 2014b. Review of statistical shape spaces for 3D data with comparative analysis for human faces. *Computer Vision and Image Understanding*, 128: 1-17 [DOI: 10.1016/j.cviu.2014.05.005]
- Cao C, Bradley D, Zhou K and Beeler T. 2015. Real-time high-fidelity facial performance capture. *ACM Transactions on Graphics*, 34(4): #46 [DOI: 10.1145/2766943]
- Cao C, Weng Y L, Zhou S, Tong Y Y and Zhou K. 2014. FaceWarehouse: a 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20(3): 413-425 [DOI: 10.1109/TVCG.2013.249]
- Chai Z H, Zhang T K, He T Y, Tan X, Baltrusaitis T, Wu H T, Li R N, Zhao S, Yuan C and Bian J. 2023. HiFace: high-fidelity 3D face reconstruction by learning static and dynamic details//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: #834 [DOI: 10.1109/ICCV51070.2023.00834]
- Chen A P, Chen Z, Zhang G L, Mitchell K and Yu J Y. 2019. Photo-Realistic facial details synthesis from single image//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9428-9438 [DOI: 10.1109/ICCV.2019.00952]
- Chiang M J Y, Kutz P and Burley B. 2016. Practical and controllable subsurface scattering for production path tracing//Proceedings of ACM SIGGRAPH 2016 Talks. Anaheim, USA: Association for Computing Machinery: #49 [DOI: 10.1145/2897839.2927433]
- Christensen P H and Burley B. 2015. Approximate reflectance profiles for efficient subsurface scattering [EB/OL]. [2023-09-25].

- <https://graphics.pixar.com/library/ApproxBSSRDF/paper.pdf>
- Dai H, Pears N, Smith W and Duncan C. 2017. A 3D morphable model of craniofacial shape and texture variation//Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy; IEEE: 3104-3112 [DOI: 10.1109/ICCV.2017.335]
- Debevec P, Hawkins T, Tchou C, Duiker H P, Sarokin W and Sagar M. 2000. Acquiring the reflectance field of a human face//Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. New Orleans, USA: ACM Press/Addison-Wesley Publishing Co.: 145-156 [DOI: 10.1145/344779.344855]
- Deng H, Wang B B, Wang R and Holzschuch N. 2020. A practical path guiding method for participating media. *Computational Visual Media*, 6(1): 37-51 [DOI: 10.1007/s41095-020-0160-1]
- Deng J K, Cheng S Y, Xue N N, Zhou Y X and Zafeiriou S. 2018. UV-GAN: adversarial facial UV map completion for pose-invariant face recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 7093-7102 [DOI: 10.1109/CVPR.2018.00741]
- Deng Y, Yang J L, Xu S C, Chen D, Jia Y D and Tong X. 2019. Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA; IEEE: 285-295 [DOI: 10.1109/CVPRW.2019.00038]
- d'Eon E. 2012. A better dipole [EB/OL]. [2023-09-25]. <http://www.eugenedeon.com/pdfs/betterdipole.pdf>
- d'Eon E and Irving G. 2011. A quantized-diffusion model for rendering translucent materials. *ACM Transactions on Graphics*, 30(4): 56 [DOI: 10.1145/2010324.1964951]
- d'Eon E, Luebke D and Enderton E. 2007. Efficient rendering of human skin//Proceedings of the 18th Eurographics Conference on Rendering Techniques. Grenoble, France: Eurographics Association: 147-157 [DOI: 10.2312/EGWR/EGSR07/147-157]
- Dib A, Ahn J, Thébault C, Gosselin P H and Chevallier L. 2023. S2F2: self-supervised high fidelity face reconstruction from monocular image//Proceedings of the 17th IEEE International Conference on Automatic Face and Gesture Recognition. Waikoloa Beach, USA; IEEE Press: 1-8 [DOI: 10.1109/FG57933.2023.10042713]
- Dib A, Bharaj G, Ahn J, Thébault C, Gosselin P, Romeo M and Chevallier L. 2021a. Practical face reconstruction via differentiable ray tracing. *Computer Graphics Forum*, 40(2): 153-164 [DOI: 10.1111/cgf.142622]
- Dib A, Thébault C, Ahn J, Gosselin P H, Theobalt C and Chevallier L. 2021b. Towards high fidelity monocular face reconstruction with rich reflectance using self-supervised learning and ray tracing//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada; IEEE: 12799-12809 [DOI: 10.1109/ICCV48922.2021.01258]
- Donner C and Jensen H W. 2005. Light diffusion in multi-layered translucent materials. *ACM Transactions on Graphics*, 24(3): 1032-1039 [DOI: 10.1145/1073204.1073308]
- Donner C and Jensen H W. 2006. A spectral BSSRDF for shading human skin//Proceedings of the 17th Eurographics Conference on Rendering Techniques. Nicosia, Cyprus: Eurographics Association: #2383946
- Donner C and Jensen H W. 2008. Rendering translucent materials using photon diffusion//Proceedings of ACM SIGGRAPH 2008 Classes. Los Angeles, USA: Association for Computing Machinery: #4 [DOI: 10.1145/1401132.1401138]
- Donner C, Lawrence J, Ramamoorthi R, Hachisuka T, Jensen H W and Nayar S. 2009. An empirical BSSRDF model//Proceedings of ACM SIGGRAPH 2009 Papers. New Orleans, USA: Association for Computing Machinery: #30 [DOI: 10.1145/1576246.1531336]
- Donner C, Weyrich T, d'Eon E, Ramamoorthi R and Rusinkiewicz S. 2008. A layered, heterogeneous reflectance model for acquiring and rendering human skin. *ACM Transactions on Graphics*, 27(5): #140 [DOI: 10.1145/1409060.1409093]
- egger B, Schönborn S, Schneider A, Kortylewski A, Morel-Forster A, Blumer C and Vetter T. 2018. Occlusion-aware 3D morphable models and an illumination prior for face image analysis. *International Journal of Computer Vision*, 126(12): 1269-1287 [DOI: 10.1007/s11263-018-1064-8]
- Fan J H, Wang B B, Hasan M, Yang J and Yan L Q. 2022. Neural layered BRDFs//Proceedings of ACM SIGGRAPH 2022 Conference. Vancouver, USA: Association for Computing Machinery: #4 [DOI: 10.1145/3528233.3530732]
- Fan J H, Wang B B, Hasan M, Yang J and Yan L Q. 2023. Neural biplane representation for BTF rendering and acquisition//Proceedings of ACM SIGGRAPH 2023 Conference. Los Angeles, USA: Association for Computing Machinery: #81 [DOI: 10.1145/3588432.3591505]
- Feng H W, Bolkart T, Tesch J, Black M J and Abrevaya V. 2022. Towards racially unbiased skin tone estimation via scene disambiguation//Proceedings of the 17th European Conference on Computer Vision (ECCV 2022). Tel Aviv, Israel: Springer: 72-90 [DOI: 10.1007/978-3-031-19778-9\_5]
- Feng Y, Feng H W, Black M J and Bolkart T. 2021. Learning an animatable detailed 3D face model from in-the-wild images. *ACM Transactions on Graphics*, 40(4): #88 [DOI: 10.1145/3450626.3459936]
- Frisvad J R, Hachisuka T and Kjeldsen T K. 2014. Directional dipole model for subsurface scattering. *ACM Transactions on Graphics*, 34(1): #5 [DOI: 10.1145/2682629]
- Fyffe G. 2010. Single-shot photometric stereo by spectral multiplexing//Proceedings of ACM SIGGRAPH ASIA 2010 Sketches. Seoul, Republic of Korea: Association for Computing Machinery: #20 [DOI: 10.1145/1899950.1899970]
- Fyffe G and Debevec P. 2015. Single-shot reflectance measurement from

- polarized color gradient illumination//Proceedings of 2015 IEEE International Conference on Computational Photography. Houston, USA: IEEE: 1-10 [DOI: 10.1109/ICCPHOT.2015.7168375]
- Fyffe G, Graham P, Tunwattanapong B, Ghosh A and Debevec P. 2016. Near-instant capture of high-resolution facial geometry and reflectance. *Computer Graphics Forum*, 35(2): 353-363 [DOI: 10.1111/cgf.12837]
- Garrido P, Valgaert L, Wu C L and Theobalt C. 2013. Reconstructing detailed dynamic face geometry from monocular video. *ACM Transactions on Graphics*, 32(6): #158 [DOI: 10.1145/2508363.2508380]
- Gecer B, Deng J K and Zafeiriou S. 2021. OSTeC: one-shot texture completion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7624-7634 [DOI: 10.1109/CVPR46437.2021.00754]
- Gecer B, Ploumpis S, Kotsia I and Zafeiriou S. 2019. GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1155-1164 [DOI: 10.1109/CVPR.2019.00125]
- Gecer B, Ploumpis S, Kotsia I and Zafeiriou S. 2022. Fast-GANFIT: generative adversarial network for high fidelity 3D face reconstruction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 4879-4893 [DOI: 10.1109/TPAMI.2021.3084524]
- Genova K, Cole F, Maschinot A, Sarna A, Vlastic D and Freeman W T. 2018. Unsupervised training for 3D morphable model regression//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8377-8386 [DOI: 10.1109/CVPR.2018.00874]
- Gerig T, Morel-Forster A, Blumer C, Egger B, Luthi M, Schönborn S and Vetter T. 2018. Morphable face models-an open framework. *Proceedings of the 13th IEEE International Conference on Automatic Face and Gesture Recognition*. Xi'an, China: IEEE: 75-82 [DOI: 10.1109/FG.2018.00021]
- Ghosh A, Fyffe G, Tunwattanapong B, Busch J, Yu X M and Debevec P. 2011. Multiview face capture using polarized spherical gradient illumination. *ACM Transactions on Graphics*, 30(6): 1-10 [DOI: 10.1145/2070781.2024163]
- Ghosh A, Hawkins T, Peers P, Frederiksen S and Debevec P. 2008. Practical modeling and acquisition of layered facial reflectance//Proceedings of ACM SIGGRAPH Asia 2008 Papers. Singapore: Association for Computing Machinery: #139 [DOI: 10.1145/1457515.1409092]
- Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y. 2014. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems—Volume 2. Montreal, Canada: MIT Press: 2672-2680
- Gosselin D. 2004. Real time skin rendering. *Game Developer Conference, D3D Tutorial (Vol. 9)*.
- Gotardo P, Riviere J, Bradley D, Ghosh A and Beeler T. 2018. Practical dynamic facial appearance modeling and acquisition. *ACM Transactions on Graphics*, 37(6): #232 [DOI: 10.1145/3272127.3275073]
- Graham P, Tunwattanapong B, Busch J, Yu X M, Jones A, Debevec P and Ghosh A. 2012. Measurement-based synthesis of facial microgeometry//Proceedings of ACM SIGGRAPH 2012 Talks. Los Angeles: Association for Computing Machinery: #9 [DOI: 10.1145/2343045.2343057]
- Guo Y D, Zhang J Y, Cai J F, Jiang B Y and Zheng J M. 2019. CNN-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(6): 1294-1307 [DOI: 10.1109/TPAMI.2018.2837742]
- Habel R, Christensen P H and Jarosz W. 2013. Photon beam diffusion: a hybrid monte carlo method for subsurface scattering. *Computer Graphics Forum*, 32(4): 27-37 [DOI: 10.1111/cgf.12148]
- Hanika J, Droske M and Fascione L. 2015. Manifold next event estimation. *Computer Graphics Forum*, 34(4): 87-97 [DOI: 10.1111/cgf.12681]
- Hanrahan P and Krueger W. 1993. Reflection from layered surfaces due to subsurface scattering//Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques. Anaheim, USA: Association for Computing Machinery: 165-174 [DOI: 10.1145/166117.166139]
- Herholz S, Zhao Y Y, Elek O, Nowrouzezahrai D, Lensch H P A and Krivánek J. 2019. Volume path guiding based on zero-variance random walk theory. *ACM Transactions on Graphics*, 38(3): #25 [DOI: 10.1145/3230635]
- Hernández C and Vogiatzis G. 2010. Self-calibrating a real-time monocular 3d facial capture system//Proceedings International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT). [s.l.]: [s.n.]
- Hertzmann A, Jacobs C E, Oliver N, Curless B and Salesin D H. 2001. Image analogies//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA: Association for Computing Machinery: 327-340 [DOI: 10.1145/383259.383295]
- Hu G S, Mortazavian P, Kittler J and Christmas W. 2013. A facial symmetry prior for improved illumination fitting of 3D morphable model//Proceedings of 2013 International Conference on Biometrics. Madrid, Spain: IEEE: 1-6 [DOI: 10.1109/ICB.2013.6613000]
- Hu L W, Saito S, Wei L Y, Nagano K, Seo J, Fursund J, Sadeghi I, Sun C, Chen Y C and Li H. 2017. Avatar digitization from a single image for real-time rendering. *ACM Transactions on Graphics*, 36(6): #195 [DOI: 10.1145/3130800.31310887]
- Huber P, Hu G S, Tena R, Mortazavian P, Koppen W P, Christmas W J, Ratsch M and Kittler J. 2016. A multiresolution 3D morphable

- face model and fitting framework//Proceedings of the 11th Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Application. Rome, Italy: SciTePress: 79-86 [DOI: 10.5220/0005669500790086]
- Huynh L, Chen W K, Saito S, Xing J, Nagano K, Jones A, Debevec P and Li H. 2018. Mesoscopic facial geometry inference using deep neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8407-8416 [DOI: 10.1109/CVPR.2018.00877]
- Ichim A E, Bouaziz S and Pauly M. 2015. Dynamic 3D avatar creation from hand-held video input. *ACM Transactions on Graphics*, 34(4): #45 [DOI: 10.1145/2766974]
- Jensen H W and Buhler J. 2002. A rapid hierarchical rendering technique for translucent materials. *ACM Transactions on Graphics*, 21(3): 576-581 [DOI: 10.1145/566654.566619]
- Jensen H W and Christensen P H. 1998. Efficient simulation of light transport in scenes with participating media using photon maps//Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques. Orlando, USA: Association for Computing Machinery: 311-320 [DOI: 10.1145/280814.280925]
- Jensen H W, Marschner S R, Levoy M and Hanrahan P. 2001. A practical model for subsurface light transport//Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques. Los Angeles, USA: Association for Computing Machinery: 511-518 [DOI: 10.1145/383259.383319]
- Jimenez J, Scully T, Barbosa N, Donner C, Alvarez X, Vieira T, Matts P, Orvalho V, Gutierrez D and Weyrich T. 2010. A practical appearance model for dynamic facial color. *ACM Transactions on Graphics*, 29(6): #141 [DOI: 10.1145/1882261.1866167]
- Jimenez J, Sundstedt V and Gutierrez D. 2009. Screen-space perceptual rendering of human skin. *ACM Transactions on Applied Perception*, 6(4): #23 [DOI: 10.1145/1609967.1609970]
- Jimenez J, Zsolnai K, Jarabo A, Freude C, Auzinger T, Wu X C, von der Pahlen J, Wimmer M and Gutierrez D. 2015. Separable subsurface scattering. *Computer Graphics Forum*, 34(6): 188-197 [DOI: 10.1111/cgf.12529]
- Johnson J, Ravi N, Reizenstein J, Novotny D, Tulsiani S, Lassner C and Branson S. 2020. Accelerating 3D deep learning with PyTorch3D//SIGGRAPH Asia 2020 Courses. [s.l.]: Association for Computing Machinery: #10 [DOI: 10.1145/3415263.3419160]
- Joseph J H, Wiscombe W J and Weinman J A. 1976. The Delta-Eddington approximation for radiative flux transfer. *Journal of the Atmospheric Sciences*, 33(12): 2452-2459 [DOI: 10.1175/1520-0469(1976)0332452:TDEAFR2.0.CO;2]
- Kampouris C, Zafeiriou S and Ghosh A. 2018. Diffuse-specular separation using binary spherical gradient illumination//Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas and Implementations. Karlsruhe, Germany: Eurographics Association: 1-10 [DOI: 10.2312/sre.20181167]
- Karras T, Laine S and Aila T. 2019. A style-based generator architecture for generative adversarial networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4401-4410 [DOI: 10.1109/CVPR.2019.00453]
- Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T. 2020. Analyzing and improving the image quality of styleGAN//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 8107-8116 [DOI: 10.1109/CVPR42600.2020.00813]
- Kim H, Zollhöfer M, Tewari A, Thies J, Richardt C and Theobalt C. 2018. InverseFaceNet: deep monocular inverse face rendering//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4625-4634 [DOI: 10.1109/CVPR.2018.00486]
- Kim J, Yang J L and Tong X. 2021. Learning high-fidelity face texture completion without complete face texture//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 13970-13979 [DOI: 10.1109/ICCV48922.2021.01373]
- Klaudiny M, Hilton A and Edge J. 2010. High-detail 3d capture of facial performance//Proceedings of International Symposium on 3D Data Processing, Visualization and Transmission. [s.l.]: [s.n.]
- Koerner D, Novák J, Kutz P, Habel R and Jarosz W. 2016. Subdivision next-event estimation for path-traced subsurface scattering//The 27th Eurographics Symposium on Rendering—Experimental Ideas & Implementations. Dublin, Ireland: Eurographics Association: 91-96 [DOI: 10.5555/3056507.3056525]
- Křivánek J and d'Eon E. 2014. A zero-variance-based sampling scheme for Monte Carlo subsurface scattering//ACM SIGGRAPH 2014 Talks (SIGGRAPH'14). Vancouver, Canada: Association for Computing Machinery: #66 [DOI: 10.1145/2614106.2614138]
- Kubelka P and Munk F. 1931. An article on optics of paint layers. *Z. Tech. Phys*, 12(593-601): 259-274.
- Lafortune E P and Willems Y D. 1996. Rendering participating media with bidirectional path tracing//Proceedings of the Eurographics Workshop on Rendering Techniques'96. Porto, Portugal: Springer-Verlag: 91-100 [DOI: 10.1007/978-3-7091-7484-5\_10]
- Lattas A, Lin Y M, Kannan J, Ozturk E, Filipi L, Guarnera G C, Chawla G and Ghosh A. 2022. Practical and scalable desktop-based high-quality facial capture//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 522-537 [DOI: 10.1007/978-3-031-20068-7\_30]
- Lattas A, Moschoglou S, Gecer B, Ploumpis S, Triantafyllou V, Ghosh A and Zafeiriou S. 2020. AvatarMe: realistically renderable 3D facial reconstruction “in-the-wild”//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 757-766 [DOI: 10.1109/CVPR42600.2020.00084]
- Lattas A, Moschoglou S, Ploumpis S, Gecer B, Deng J K and Zafeiriou

- S. 2023. FitMe: deep photorealistic 3D morphable model avatars// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 8629-8640 [DOI: 10.1109/CVPR52729.2023.00834]
- Lattas A, Moschoglou S, Ploumpis S, Gecer B, Ghosh A and Zafeiriou S. 2021. AvatarMe<sup>++</sup>: facial shape and brdf inference with photorealistic rendering-aware gans. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(12): 9269-9284 [DOI: 10.1109/TPAMI.2021.3125598]
- Lei B W, Ren J Q, Feng M Y, Cui M M and Xie X S. 2023. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 394-403 [DOI: 10.1109/CVPR52729.2023.00046]
- Lensch H P A, Kautz J, Goesele M, Heidrich W and Seidel H P. 2001. Image-based reconstruction of spatially varying materials//Proceedings of the Eurographics Workshop on Rendering Techniques 2001. London, United Kingdom: Springer: 103-114 [DOI: 10.1007/978-3-7091-6242-2\_10]
- Li H S, Pellacini F and Torrance K E. 2005. A hybrid monte carlo method for accurate and efficient subsurface scattering//Proceedings of the 16th Eurographics Conference on Rendering Techniques. Konstanz, Germany: Eurographics Association: 283-290 [DOI: 10.2312/EGWR/EGSR05/283-290]
- Li R L, Bladin K, Zhao Y J, Chinara C, Ingraham O, Xiang P D, Ren X L, Prasad P, Kishore B, Xing J and Li H. 2020. Learning formation of physically-based face attributes//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3407-3416 [DOI: 10.1109/CVPR42600.2020.00347]
- Li T Y, Bolkart T, Black M J, Li H and Romero J. 2017. Learning a model of facial shape and expression from 4D scans. ACM Transactions on Graphics, 36(6): #194 [DOI: 10.1145/3130800.3130813]
- Li T Y, Liu S C, Bolkart T, Liu J Y, Li H and Zhao Y J. 2021. Topologically consistent multi-view face inference using volumetric sampling//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3804-3814 [DOI: 10.1109/ICCV48922.2021.00380]
- Lin J K, Yuan Y, Shao T J and Zhou K. 2020. Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 5890-5899 [DOI: 10.1109/CVPR42600.2020.00593]
- Liu S C, Cai Y X, Chen H W, Zhou Y C and Zhao Y J. 2022. Rapid face asset acquisition with recurrent feature alignment. ACM Transactions on Graphics, 41(6): #214 [DOI: 10.1145/3550454.3555509]
- Ma W C, Hawkins T, Peers P, Chabert C F, Weiss M and Debevec P E. 2007. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination//Proceedings of the 18th Eurographics Conference on Rendering Techniques. Grenoble, France: Eurographics Association: 183-194 [DOI: 10.2312/EGWR/EGSR07/183-194]
- Meng J, Hanika J and Dachsbacher C. 2016. Improving the Dwivedi sampling scheme. Computer Graphics Forum, 35(4): 37-44 [DOI: 10.1111/cgf.12947]
- Mildenhall B, Srinivasan P P, Tancik M, Barron J T, Ramamoorthi R and Ng R. 2022. NeRF: representing scenes as neural radiance fields for view synthesis. Communications of the ACM, 65(1): 99-106 [DOI: 10.1145/3503250]
- Nagano K, Fyffe G, Alexander O, Barbič J, Li H, Ghosh A and Debevec P. 2015. Skin microstructure deformation with displacement map convolution. ACM Transactions on Graphics, 34(4): #109 [DOI: 10.1145/2766894]
- Nicodemus F E, Richmond J C, Hsia J J, Ginsberg I W and Limperis T. 1977. Geometrical Considerations and Nomenclature for Reflectance. U. S. Department of Commerce [DOI: 10.6028/nbs.mono.160]
- Papaoiannou A, Gecer B, Cheng S Y, Chrysos G, Deng J K, Fotiadou E, Kampouris C, Kollias D, Moschoglou S, Songsri-In K, Ploumpis S, Trigeorgis G, Tzirakis P, Ververas E, Zhou Y X, Ponniah A, Roussos A and Zafeiriou S. 2022. MimicME: a large scale diverse 4D database for facial expression analysis//Proceedings of the 17th European Conference on Computer Vision (ECCV 2022). Tel Aviv, Israel: Springer-Verlag: 467-484 [DOI: 10.1007/978-3-031-20074-8\_27]
- Pauly M, Kollig T and Keller A. 2000. Metropolis light transport for participating media//Proceedings of the Eurographics Workshop on Rendering Techniques 2000. Brno, Czech Republic: Springer-Verlag: 11-22 [DOI: 10.1007/978-3-7091-6303-0\_2]
- Paysan P, Lütthi M, Albrecht T, Lerch A, Amberg B, Santini F and Vetter T. 2009. Face reconstruction from skull shapes and physical attributes//Proceedings of the 31st DAGM Symposium on Pattern Recognition. Jena, Germany: Springer: 232-241 [DOI: 10.1007/978-3-642-03798-6\_24]
- Pérez P, Gangnet M and Blake A. 2003. Poisson image editing. ACM Transactions on Graphics, 22(3): 313-318 [DOI: 10.1145/882262.882269]
- Piotraschke M and Blanz V. 2016. Automated 3D face reconstruction from multiple images using quality measures//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 3418-3427 [DOI: 10.1109/CVPR.2016.372]
- Radford A, Kim J W, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G and Sutskever I. 2021. Learning transferable visual models from natural language supervision//Proceedings of the 38th International Conference on Machine Learning. [s.l.]: PMLR: #20
- Rainer G, Ghosh A, Jakob W and Weyrich T. 2020. Unified neural

- encoding of BTFs. *Computer Graphics Forum*, 39 (2) : 167-178 [DOI: 10.1111/cgf.13921]
- Rainer G, Jakob W, Ghosh A and Weyrich T. 2019. Neural BTF compression and interpolation. *Computer Graphics Forum*, 38 (2) : 235-244 [DOI: 10.1111/cgf.13633]
- Raman C, Hewitt C, Wood E and Baltrušaitis T. 2023. Mesh-tension driven expression-based wrinkles for synthetic faces//*Proceedings of 2023 IEEE/CVF Winter Conference on Applications of Computer Vision*. Waikoloa, USA: IEEE: 3504-3514 [DOI: 10.1109/WACV56688.2023.00351]
- Riviere J, Gotardo P, Bradley D, Ghosh A and Beeler T. 2020. Single-shot high-quality facial geometry and skin appearance capture. *ACM Transactions on Graphics*, 39 (4) : #81 [DOI: 10.1145/3386569.3392464]
- Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. 2022. High-resolution image synthesis with latent diffusion models//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 10674-10685 [DOI: 10.1109/CVPR52688.2022.01042]
- Saito S, Wei L Y, Hu L W, Nagano K and Li H. 2017. Photorealistic facial texture inference using deep neural networks//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 2326-2335 [DOI: 10.1109/CVPR.2017.250]
- Schlick C. 1994. An inexpensive BRDF model for physically-based rendering. *Computer Graphics Forum*, 13 (3) : 233-246 [DOI: 10.1111/1467-8659.1330233]
- Shen Y J, Gu J J, Tang X O and Zhou B L. 2020. Interpreting the latent space of GANs for semantic face editing//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 9240-9249 [DOI: 10.1109/CVPR42600.2020.00926]
- Shi F H, Wu H T, Tong X and Chai J X. 2014. Automatic acquisition of high-fidelity facial performances using monocular videos. *ACM Transactions on Graphics*, 33 (6) : #222 [DOI: 10.1145/2661229.2661290]
- Smith W A P, Seck A, Dee H, Tiddeman B, Tenenbaum J B and Egger B. 2020. A morphable face albedo model//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 5010-5019 [DOI: 10.1109/CVPR42600.2020.00506]
- Stam J. 1995. Multiple scattering as a diffusion process//*Proceedings of the Eurographics Workshop on Rendering Techniques '95*. Dublin, Ireland: Springer: 41-50 [DOI: 10.1007/978-3-7091-9430-0\_5]
- Stratou G, Ghosh A, Debevec P and Morency L P. 2011. Effect of illumination on automatic expression recognition: a novel 3D relightable facial database//*Proceedings of 2011 IEEE International Conference on Automatic Face and Gesture Recognition*. Santa Barbara, USA: IEEE: 611-618 [DOI: 10.1109/FG.2011.5771467]
- Sztrajman A, Rainer G, Ritschel T and Weyrich T. 2021. Neural BRDF representation and importance sampling. *Computer Graphics Forum*, 40(6): 332-346 [DOI: 10.1111/cgf.14335]
- Tewari A, Zollhöfer M, Garrido P, Bernard F, Kim H, Pérez P and Theobalt C. 2018. Self-supervised multi-level face model learning for monocular reconstruction at over 250 Hz//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 2549-2559 [DOI: 10.1109/CVPR.2018.00270]
- Tewari A, Zollhöfer M, Kim H, Garrido P, Bernard F, Pérez P and Theobalt C. 2017. MoFA: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 3735-3744 [DOI: 10.1109/ICCV.2017.401]
- Torrance K E and Sparrow E M. 1967. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9) : 1105-1114 [DOI: 10.1364/JOSA.57.001105]
- Tuan Tran A, Hassner T, Masi I and Medioni G. 2017. Regressing robust and discriminative 3D morphable models with a very deep neural network//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 1493-1502 [DOI: 10.1109/CVPR.2017.163]
- Valgaerts L, Wu C L, Bruhn A, Seidel H P and Theobalt C. 2012. Lightweight binocular facial performance capture under uncontrolled lighting. *ACM Transactions on Graphics*, 31 (6) : #187 [DOI: 10.1145/2366145.2366206]
- Wang L H, Jacques S L and Zheng L Q. 1995. MCML—Monte Carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, 47 (2) : 131-146 [DOI: 10.1016/0169-2607(95)01640-F]
- Wang L Z, Zhao X C, Sun J X, Zhang Y X, Zhang H W, Yu T and Liu Y B. 2023a. StyleAvatar: real-time photo-realistic portrait avatar from a single video//*ACM SIGGRAPH 2023 Conference Proceedings*. Los Angeles, USA: Association for Computing Machinery: #67 [DOI: 10.1145/3588432.3591517]
- Wang Q, Luan F J, Dai Y X, Huo Y C, Bao H J and Wang R. 2023b. A biophysically-based skin model for heterogeneous volume rendering [EB/OL]. [2023-09-25]. [https://luanfujun.com/files/cvm2023/skin\\_paper.pdf](https://luanfujun.com/files/cvm2023/skin_paper.pdf)
- Wang T C, Liu M Y, Zhu J Y, Tao A, Kautz J and Catanzaro B. 2018. High-resolution image synthesis and semantic manipulation with conditional GANs//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 8798-8807 [DOI: 10.1109/CVPR.2018.00917]
- Wang T F, Zhang B, Zhang T, Gu S Y, Bao J M, Baltrušaitis T, Shen J J, Chen D, Wen F, Chen Q F and Guo B N. 2023c. RODIN: a generative model for sculpting 3D digital avatars using diffusion//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 4563-4573

- [DOI: 10.1109/CVPR52729.2023.00443]
- Wang Y F, Holynski A, Zhang X M and Zhang X F. 2023d. SunStage: portrait reconstruction and relighting using the sun as a light stage// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 20792-20802 [DOI: 10.1109/CVPR52729.2023.01992]
- Weber P, Hanika J and Dachsbacher C. 2017. Multiple vertex normal estimation for lighting in dense, forward-scattering media. *Computer Graphics Forum*, 36(2): 21-30 [DOI: 10.1111/cgf.13103]
- Weiss S, Moulin J, Chandran P, Zoss G, Gotardo P and Bradley D. 2023. Graph-based synthesis for skin micro wrinkles. *Computer Graphics Forum*, 42(5): #14904 [DOI: 10.1111/cgf.14904]
- Weyrich T, Matusik W, Pfister H, Bickel B, Donner C, Tu C E, McAndless J, Lee J, Ngan A, Jensen H W and Gross M. 2006. Analysis of human faces using a measurement-based skin reflectance model. *ACM Transactions on Graphics*, 25(3): 1013-1024 [DOI: 10.1145/1141911.1141987]
- Wood E, Baltrušaitis T, Hewitt C, Dziadzio S, Cashman T J and Shotton J. 2021. Fake it till you make it: face analysis in the wild using synthetic data alone// Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 3661-3671 [DOI: 10.1109/ICCV48922.2021.00366]
- Woodham R J. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1): #191139 [DOI: 10.1117/12.7972479]
- Wrenninge M, Villemin R and Hery C. 2017. Path traced subsurface scattering using anisotropic phase functions and non-exponential free flights [EB/OL]. [2023-09-25].  
<https://graphics.pixar.com/library/PathTracedSubsurface/paper.pdf>
- Wu F Z, Bao L C, Chen Y J, Ling Y G, Song Y B, Li S M, Ngan K N and Liu W. 2019. MVF-Net: multi-view 3D face morphable model regression// Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 959-968 [DOI: 10.1109/CVPR.2019.00105]
- Wu M H, Zhu H, Huang L J, Zhuang Y Y, Lu Y X and Cao X. 2023. High-fidelity 3D face generation from natural language descriptions// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 4521-4530 [DOI: 10.1109/CVPR52729.2023.00439]
- Wu W S, Wang B B and Yan L Q. 2022. A survey on rendering homogeneous participating media. *Computational Visual Media*, 8(2): 177-198 [DOI: 10.1007/s41095-021-0249-1]
- Yamaguchi S, Saito S, Nagano K, Zhao Y J, Chen W K, Olszewski K, Morishima S and Li H. 2018. High-fidelity facial reflectance and geometry inference from an unconstrained image. *ACM Transactions on Graphics*, 37(4): #162 [DOI: 10.1145/3197517.3201364]
- Yang H T, Zhu H, Wang Y R, Huang M K, Shen Q, Yang R G and Cao X. 2020. Facescape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction// Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 598-607 [DOI: 10.1109/CVPR42600.2020.00068]
- Yu C C, Lu G S, Zeng Y H, Sun J, Liang X D, Li H B, Xu Z B, Xu S C, Zhang W and Xu H. 2023. Towards high-fidelity text-guided 3D face generation and manipulation using only images// Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 15280-15291 [DOI: 10.1109/ICCV51070.2023.01406]
- Zeltner T, Rousselle F, Weidlich A, Clarberg P, Novák J, Bitterli B, Evans A, Davidović T, Kallweit S and Lefohn A. 2023. Real-time neural appearance models [EB/OL]. [2023-09-25].  
<https://arxiv.org/pdf/2305.02678.pdf>
- Zhang L W, Qiu Q W, Lin H Y, Zhang Q X, Shi C, Yang W, Shi Y, Yang S B, Xu L and Yu J Y. 2023a. DreamFace: progressive generation of animatable 3D faces under text guidance. *ACM Transactions on Graphics*, 42(4): #138 [DOI: 10.1145/3592094]
- Zhang L W, Zhao Z J, Cong X Z, Zhang Q X, Gu S Q, Gao Y C, Zheng R, Yang W, Xu L and Yu J Y. 2023b. HACK: Learning a parametric head and neck model for high-fidelity animation. *ACM Transactions on Graphics*, 42(4): #41 [DOI: 10.1145/3592093]
- Zhao X C, Wang L Z, Sun J X, Zhang H W, Suo J L and Liu Y B. 2024. HAvatar: high-fidelity head avatar via facial model conditioned neural radiance field. *ACM Transactions on Graphics*, 43(1): #6 [DOI: 10.1145/3626316]
- Zheng M W, Zhang H Y, Yang H Y and Huang D. 2023. NeuFace: realistic 3D neural face rendering from multi-view images// Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 16868-16877 [DOI: 10.1109/CVPR52729.2023.01618]
- Zickler T, Mallick S P, Kriegman D J and Belhumeur P N. 2008. Color subspaces as photometric invariants// Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE: 2000-2010 [DOI: 10.1109/CVPR.2006.77]

## 作者简介

郝琮晖,男,硕士研究生,主要研究方向为真实感渲染,神经渲染。E-mail:2120230623@mail.nankai.edu.cn

王贝贝,通信作者,女,教授,主要研究方向为计算机图形学真实感渲染、三维视觉。E-mail:beibei.wang@nju.edu.cn

杜悠扬,男,硕士研究生,主要研究方向为真实感渲染,逆向渲染。E-mail:duyouyang957@gmail.com

王璐,女,教授,主要研究方向为真实感渲染、实时渲染、表现建模。E-mail:luwang\_hcivr@sdu.edu.cn