基于 3D 人脸跟踪的实时真人虚拟试发

唐博奕, 杨文武*, 赵叶清, 杨柏林, 金剑秋

(浙江工商大学计算机与信息工程学院 杭州 310018)
(wwwyang@zjgsu.edu.cn)

摘 要: 真人虚拟试发在虚拟社交、在线商城试戴和个人发型设计等领域具有明显的应用价值. 现有的基于图像的 2D 虚拟试发系统无法让用户从不同的角度和距离实时观看试发效果, 极大地限制了这类系统的可应用范围. 为此, 提出一个基于 3D 人脸跟踪的实时真人虚拟试发系统, 用户仅使用一个网络摄像头, 就能如同在镜子前试戴真实假发一样进行虚拟试发. 首先结合深度学习技术与传统优化算法, 从 2D 人脸视频帧中快速鲁棒地计算出对应的 3D 人脸形状和姿态, 实现 3D 人脸跟踪; 然后根据事先指定的若干对对应点, 自动把 3D 头发佩戴到视频帧所对应的 3D 人脸模型; 最后根据 3D 人脸模型对 3D 虚拟头发进行遮挡处理与绘制, 实现虚实混合的真人虚拟试发效果. 此外, 还提出一种简单实用的 3D 头发重上色算法, 以进一步增强虚拟试发效果. 在室内外环境下, 使用网络摄像头采集用户视频进行了大量的真人虚拟试发实验的结果表明, 该系统是有效的和实用的.

关键词: 虚拟试发; 3D 人脸跟踪; 虚拟戴发; 头发重上色

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18719

Real-Time Virtual Hair Try-on by 3D Face Tracking

Tang Boyi, Yang Wenwu*, Zhao Yeqing, Yang Bailin, and Jin Jianqiu

(School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018)

Abstract: Virtual hair try-on is becoming popular in many applications such as virtual social networking, try-on for online shopping, and hairstyle design. Most of existing systems are image-based where the user is not allowed to view the try-on effects from different perspectives and distances in real-time, which limits the usability of the system. A real-time virtual hair try-on system based on 3D face tracking is proposed, in this system, by using an ordinary web camera, the user is enabled to perform the virtual hair try-on in a way as s/he conducts the real hair try-on in front of a mirror. Firstly, the system realizes 3D face tracking, which combines the deep learning technique and the traditional optimization algorithm for determining the shape and pose of the 3D faces from the 2D video frames. Then, it automatically puts the 3D hair onto the 3D faces corresponding to the video frames with several pairs of user-specified points to wear hair. Finally, virtual-real rendering, which handles the occlusion of the 3D hair and renders it with the face in the 2D video frames, leads to the virtual-real effect of hair try-on. In addition, a simple but effective algorithm of recoloring the 3D hair is proposed so as to improve the effect of virtual hair try-on. In indoor and outdoor environments, extensive virtual try-on evaluations demonstrate the effectiveness and practicability of our system.

收稿日期: 2020-10-15; 修回日期: 2020-11-29. 基金项目: 国家重点研发计划(2018YFB1403200); 国家自然科学基金(U1609215); 浙江省自然科学基金(LY21F020010). 唐博奕(1994—), 男, 硕士研究生, 主要研究方向为计算机图形学; 杨文武(1981—), 男, 博士, 教授, 硕士生导师, CCF 会员, 论文通讯作者, 主要研究方向为计算机动画、计算机视觉、计算机图形学; 赵叶清(1997—), 女, 硕士研究生, 主要研究方向为计算机图形学; 杨柏林(1975—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为计算机图形学、计算机视觉、移动图形计算; 金剑秋(1977—), 男, 博士, 副教授, 主要研究方向为计算机视觉、图像处理.

Key words: virtual hair try-on; 3D face tracking; virtual hair-wearing; hair recoloring

头发是个人形象的一个明显特征. 在当今的个性化时代, 人们对个人形象越来越关注, 一款适合自己的发型或佩戴一顶合适的假发已经成为人们日常生活工作中的一个自然选择. 与实体店的假发试戴相比, 虚拟试发更便捷、成本更低且应用范围更广, 可广泛应用于虚拟社交、在线商城试戴和个人发型设计等领域.

目前, 主流的虚拟试发系统是基于 2D 图像的: 用户输入一张人脸相片, 并选择一款特定发型, 系统自动把对应发型的头发佩戴在相片中的人头顶; 但在该类系统中, 用户难以从不同的角度和距离实时观察试发效果. 此外, 因为该类系统只是在图像空间中把指定头发放置在用户人头顶, 而没有考虑用户人头的实际 3D 方位和尺寸, 所以容易产生失真的戴发效果. 为了解决上述问题, 一种可能的方案是重建一个与用户外观接近的 3D 人脸模型, 并把指定的 3D 虚拟头发佩戴到 3D 人脸模型的头顶, 用户可以通过旋转和移动 3D 人脸模型从多方位浏览试发效果. 但是, 实时重建外观相似的 3D 人脸模型仍是一个极具挑战性的课题^[1]; 此外, 由于用户只能通过旋转和移动 3D 模型观看试发效果, 该方案缺少用户沉浸感, 降低了用户的试发体验.

针对以上问题, 本文提出一种基于 3D 人脸跟踪的实时真人虚拟试发系统. 该系统基于增强现实技术, 通过一个普通的网络摄像头采集用户的试发视频, 对视频帧中用户的 3D 人脸形状与姿态进行实时跟踪结合戴发算法自动把 3D 虚拟头发佩戴到视频帧中的用户人头顶并进行增强现实绘制, 使用户可以实时观看虚实混合的试发效果.

为了把 3D 虚拟头发自然地佩戴到 2D 视频帧中的用户人头顶, 需要正确建立它们之间的位置和姿态关联. 针对这一目标, 本文训练了一个鲁棒且快速的基于深度神经网络的回归器, 可以有效地建立输入(一幅人脸图像)和输出(对应的 3D 人脸形状与姿态)之间的关系, 进而实现针对用户试发视频流中 3D 人脸形状及其姿态实时且有效的跟踪.

在进行真人虚拟试发时, 算法根据事先指定的若干对对应点, 自动把 3D 头发佩戴到视频帧对应的 3D 人脸模型上, 让其追随 3D 人脸的头部运动; 然后对 3D 虚拟头发进行遮挡处理并进行增强现实绘制, 最终实现虚实混合的真人虚拟试发效果.

1 相关工作

1.1 2D/3D 人脸关键点检测

人脸对齐即检测人脸图像上的关键特征点, 是诸多计算机视觉应用(如人脸跟踪)的一个重要步骤. 在传统的人脸对齐技术中, 较有代表性的方法可以分为 2 大类: 基于统计模型进行参数优化的方法^[1-5]和基于级联形状回归的方法^[6-9]. 虽然这些传统的方法已经取得了较好的人脸关键点检测结果, 但是近年来基于深度卷积神经网络的人脸对齐方法越来越流行^[10-13], 并且在人脸对齐的性能上取得了明显的提升. 上述方法检测的均是人脸图像中的可见 2D 关键点, 由于人脸大幅度旋转时发生的自遮挡, 对于不同角度的人脸图像检测到的人脸轮廓 2D 关键点通常对应着不同的人脸轮廓特征, 如图 1a 和图 1b 所示.

图 1 2D/3D 轮廓关键点

不同于 2D 人脸对齐技术, 基于深度学习的 3D 人脸对齐方法可为人脸图像检测到关键点的 3D 位置, 如图 1c 所示(蓝色表示由于遮挡不可见). 给定一幅人脸图像, 这些方法一般把它拟合到一个 3D 形变模型(3D morphable model, 3DMM)或一个 3D 人脸模板, 以获得对应关键点的 3D 位置. 例如, Zhu 等^[14]通过级联多个卷积神经网络估计 3DMM 参数; Bulat 等^[15]提出结合人脸特征点定位架构和残差模块的方法; Feng 等^[16]设计了 UV 位置图表示 UV 空间内完整人脸的 3D 形状特征. 为了解决缺少训练样本的问题, Tu 等^[17]通过自监督学习进一步提升 3D 人脸对齐的精度.

1.2 基于单摄像头的 3D 人脸跟踪

目前, 已有的一些基于特殊设备(如结构光扫描仪^[18-19]或多相机阵列^[20-21])的 3D 人脸跟踪技术, 可以实现高精度的 3D 人脸几何重建及其运动跟踪. 但这些技术计算耗时, 通常所用设备造价昂贵且使用复杂, 因而难以推广到一般的消费级应用.

目前, 市场上的移动设备大都配有普通的网络

摄像头, 因此利用单个网络摄像头进行实时的 3D 人脸跟踪是最为简单且可广泛应用的方案. Chai 等^[22]首先提取出人脸图像中的 2D 关键点等特征, 再通过拟合一个参数化的人脸表情模型, 实时地跟踪人脸的 3D 形状及其表情. 类似的, Weise 等^[23]使用现有的人脸表情样本数据进一步保证跟踪结果的有效性. 为了提高跟踪的准确性, Cao 等^[24]提出基于特定用户的 3D 人脸跟踪算法, 但需要事先为用户采集一批特定角度和表情的人脸图像, 增加了用户的使用难度. Cao 等^[25]提出刚性稳定性问题及其解决方案, 以提升 3D 人脸跟踪的稳定性. 在 3D 人脸的跟踪过程中, 上述方法通常使用 2D 人脸关键点作为跟踪特征. 如图 1 所示, 当人脸发生大幅度旋转时, 所检测到的轮廓处 2D 关键点会对应着不同的人脸轮廓特征, 这就需要跟踪算法及时更新这些 2D 关键点与 3D 人脸模型顶点的关系, 从而带来额外的计算量或产生不稳定的跟踪结果. 本文系统的 3D 人脸跟踪模块中采用 3D 关键点作为跟踪特征, 这些关键点对应着 3D 人脸模型上的固定顶点, 因而避免了上述问题.

1.3 基于增强现实的试穿试戴

近年来, 随着虚拟现实技术的发展以及互联网经济的兴起, 产生了一系列面向虚拟试穿试戴的相关应用, 如虚拟试衣、虚拟试发和虚拟试镜等, 这些应用同时吸引了学术界和工业界的诸多关注. 结合计算机视觉、计算机图形学、深度学习、虚拟现实和增强现实中的交互式控制技术^[26-27], 虚拟试衣无论在理论上还是产品应用上均取得了较大突破^[28-29].

相比于虚拟试衣, 虚拟试发和虚拟试镜(即眼镜试戴)方面的工作较少. 与虚拟试发类似, 已有的虚拟试镜方法也分为 2 类: 基于 2D 图像的方法^[30]和基于 3D 模型的方法. 在基于 2D 图像的虚拟试发和试镜系统中, 用户难以从不同的角度和距离实时观察试戴效果, 并且因为没有考虑用户人脸的实际 3D 方位和尺寸, 容易产生失真的试戴效果. 而在基于 3D 模型的系统中, 系统重建一个与用户外观接近的 3D 人脸模型, 并把 3D 头发或眼镜佩戴在该 3D 模型上, 从而用户可以通过旋转和移动 3D 模型从多方位观看试戴效果. 但是, 这类系统缺少用户沉浸感, 进而降低了试戴体验.

2 本文系统

本文提出基于 3D 人脸跟踪的实时真人虚拟试

发系统. 该系统基于增强现实技术, 把真实的人脸试戴视频与 3D 虚拟头发混合绘制在一起, 从而允许用户如同在镜子前试戴真实假发一样体验虚拟试发, 如图 2 所示.

图 2 用户试用本文系统

本文系统预先训练一个基于深度神经网络的回归器, 用于检测 3D 人脸关键点. 输入一幅人脸图像, 得到一组人脸关键特征点的 3D 位置. 根据检测到的 3D 人脸关键点, 再基于传统优化方法, 实现视频帧中用户 3D 人脸形状及其姿态的实时跟踪. 在虚拟试发过程中, 首先跟踪当前视频帧中用户的 3D 人脸; 然后根据事先指定的若干对对应点, 利用基于方位一致性的 3D 头发佩戴与虚实绘制的算法, 自动地把 3D 头发佩戴到当前帧的 3D 人脸模型, 并对 3D 头发进行遮挡处理和绘制; 最后结合原始视频帧与 3D 头发的绘制结果, 实现虚实混合的真人虚拟试发效果.

2.1 3D 人脸模型的表示

本文系统使用 Blanz 等^[31]提出的 3DMM 表示用户的 3D 人脸信息, 即

$$\mathbf{M} = \overline{\mathbf{M}} + \mathbf{B}_{\text{id}} \boldsymbol{\alpha}_{\text{id}} + \mathbf{B}_{\text{exp}} \boldsymbol{\alpha}_{\text{exp}} \quad (1)$$

其中, $\mathbf{M}$ 对应一个 3D 人脸模型; $\overline{\mathbf{M}}$ 为平均人脸形状; $\mathbf{B}_{\text{id}}$ 为一组个体基向量, 定义中性表情下不同个体的人脸形状空间; $\mathbf{B}_{\text{exp}}$ 为一组表情基向量, 定义人脸的表情空间; $\boldsymbol{\alpha}_{\text{id}}$ 和 $\boldsymbol{\alpha}_{\text{exp}}$ 分别为用户的个体和表情系数向量. 本文系统中, $\mathbf{B}_{\text{id}}$ 来自巴塞尔人脸模型(Basel face model, BFM)^[32]; $\mathbf{B}_{\text{exp}}$ 来自 FaceWarehouse^[33]. 通过组合个体和表情系数, 式(1)可以表示不同个体在各种表情下的 3D 人脸.

2.2 人脸坐标系表示

根据人脸的刚性运动(平移和旋转), 可以得到 3D 人脸模型在世界坐标系中的位置 $\mathbf{M}_{\text{w}} = \mathbf{R}\mathbf{M} + \mathbf{t}$. 其中, $\mathbf{R}$ 为旋转矩阵, 由围绕 x, y, z 轴的旋转角度

r_x, r_y, r_z 构成; t 为平移向量. 通过弱透视投影, 可以把 3D 人脸模型变换到屏幕(图像)坐标空间

$$M_s = f \times P_r R M + t_{2D} \quad (2)$$

其中, P_r 为正交投影矩阵 $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$; f 为透视缩放系数; t_{2D} 为 t 的 x, y 分量. 此外, 可以通过利用 M_w 中各顶点的 z 轴坐标生成 3D 人脸模型各顶点在屏幕空间中的深度坐标.

为便于下文的讨论, 令 Π_Q 为把 3D 人脸模型上的一个点 p 变换到图像坐标空间的映射函数, 即

$$\Pi_Q(p) = f \times P_r R p + t_{2D} \quad (3)$$

其中, $Q = \{\alpha_{id}, \alpha_{exp}, r_x, r_y, r_z, t, f\}$ 为相关参数的集合.

3 算法实现

3.1 面向虚拟试发的实时 3D 人脸跟踪

3.1.1 人脸 3D 关键点的检测

本文通过检测视频帧中的人脸 3D 关键点实现 3D 人脸的实时跟踪. Cao 等^[24]提出一个基于传统回归算法的检测器, 可以从人脸图像中检测到对应的 3D 关键点; 但是, 该方法需要为每个用户单独采集若干幅特定角度和表情的图像, 并且需要为每个用户单独训练一个检测器. 为了避免上述问题, 本文采用深度神经网络实现人脸图像到对应 3D 关键点的回归算法, 利用深度神经网络强大的泛化能力和大量的训练数据样本, 基于深度神经网络的回归器只需要训练一次, 就可用于检测任意用户的人脸 3D 关键点.

本文系统使用轻量级的深度神经网络 MobileNet^[34]作为回归算法的骨干神经网络. 与文献[14,17]类似, 该深度神经网络不是直接输出对应的 3D 关键点, 而是把人脸图像拟合到一个 3DMM 上, 即输入一幅人脸图像, 输出一组参数 Q . 为了训练该网络, 使用文献[14]中提出的参数距离和顶点距离定义损失函数, 同时使用文献[14]中的样本构建方法获得训练监督样本, 具体的训练参数设置见第 4.1 节, 更详细的方法见文献[14].

在 3DMM 上, 首先在眼角、鼻子和脸部轮廓处提取出 68 个特征点; 然后随机采样 132 个顶点, 它们和特征点一起组成了所需的人脸 3D 关键点, 用于稳定且鲁棒地对用户的 3D 人脸进行跟踪. 如图 3 所示, 给定一个人脸视频帧, 通过前述的深度神经网络得到一组对应参数 Q . 其中, 在图 3 最右侧的

人脸模型上, 红色对应着人脸特征点, 蓝色对应着额外的随机采样顶点(注意, 被遮挡而不可见的 3D 关键点也被画出). 利用这组参数 Q , 可以通过第 2 节中的式(1)(2)(3)算得到人脸 3D 关键点在 3DMM 上的位置坐标和在图像空间中的屏幕坐标, 分别记为 V_{3D} 和 V_{2D} .

图 3 人脸 3D 关键点检测

3.1.2 3D 人脸跟踪

与各帧独立进行人脸 3D 关键点检测不同, 本文在 3D 人脸跟踪时加入了相邻帧 3D 人脸运动的连贯性约束, 以保证用户在进行虚拟试发时的视觉流畅性. 同样地, 在 3D 人脸跟踪算法中, 也使用 3DMM 来表示用户的 3D 人脸信息. 因此, 对于每一时刻 m 的人脸视频帧, 跟踪其中的 3D 人脸几何与姿态信息, 即求解一组对应的参数 Q_t , 它包含着 3 部分信息: 3D 人脸几何与表情($\alpha_{id}, \alpha_{exp}$)、人脸姿态(r_x, r_y, r_z, t)和模拟 3D 人脸透视投影效果的缩放系数 f . 通过求解优化问题

$$\arg \min_{Q_m} \|u_m - V_{2D}^m\|^2 + w \|u_m - 2u_{m-1} + u_{m-2}\|^2$$

计算 Q_m . 其中, $u_m = \Pi_{Q_m}(V_{3D})$, V_{3D} 表示参数 Q_m 所对应的 3DMM 上的 3D 关键点; u_m 表示这些点投影后的图像坐标位置; V_{2D}^m 为当前帧检测到的人脸 3D 关键点的屏幕坐标; w 为能量权值, 在实验中取值为 0.5. 上述表达式对应着一个非线性最小二乘优化问题, 本文使用 Levenberg-Marquardt 算法^[35]进行求解. 在表达式中, $\|u_m - V_{2D}^m\|^2$ 表示跟踪到的 3D 人脸在屏幕空间中能够与检测到 3D 关键点对齐, 而 $w \|u_m - 2u_{m-1} + u_{m-2}\|^2$ 表示在屏幕空间中相邻帧之间 3D 关键点的运动速度能够保持一致, 即 $(u_m - u_{m-1}) - (u_{m-1} - u_{m-2}) \cong 0$. 之所以选择在屏幕空间中保持 3D 人脸运动的连贯性, 是因为本文的最终目标是希望在屏幕空间中绘制的虚拟试发结果能够保持视觉流畅性.

3.2 基于方位一致性的 3D 头发佩戴与虚实绘制

美工师在 3D 头发建模时, 通常需要一个 3D 人脸模型作为参考. 因此, 可以在该模型和 3DMM 的平均人脸模型上预先选择若干对对应顶

点. 本文系统中, 美工师使用 3DMM 的平均人脸模型作为头发建模时的参考, 因此只需要在 3DMM 的平均人脸模型上选择 4 个顶点(左右耳和前额处), 令这 4 个顶点的位置为 p .

对于任意人脸视频帧, 在跟踪得到的 3D 人脸模型上, 同样可以获得上述 4 个顶点的位置, 设为 q . 然后, 可以求得一个最优的相似变换矩阵 T_s , 包括旋转、平移和均匀缩放, 把 p 变换到 q , 即

$$\arg \min_{T_s} \|T_s p - q\|^2.$$

上述表达式对应一个最小二乘问题, 可通过奇异值分解进行求解^[36], 将相似变换 T_s 作用到 3D 头发, 即可实现把 3D 头发佩戴到当前帧的用户 3D 人脸, 如图 4 所示.

图 4 3D 头发配戴流程

完成 3D 头发与 3D 人脸的配准之后, 进一步根据 3D 人脸模型对 3D 虚拟头发进行遮挡处理并进行绘制, 最终实现虚实混合的真人虚拟试发效果. (1) 由于当前帧的用户 3D 人脸与视频帧中真实人脸形状精确对齐, 因此把 3D 头发与真实人脸的相互遮挡关系转变为 3D 头发与 3D 人脸的遮挡关系. (2) 只绘制未被 3D 人脸模型遮挡的 3D 头发, 可以通过开放式图形库(open graphics library, OpenGL)的深度测试完成, 并把绘制结果直接叠加在当前视频帧上, 进而得到虚实混合的真人虚拟试发效果.

3.3 保持相邻像素色差的 3D 头发重上色

为了进一步增强虚拟试发效果和用户体验, 本文系统允许用户指定所佩戴头发的颜色. 一种简单方法是直接把 3D 头发绘制成指定的颜色, 但是这种方法丢失了 3D 头发原有的明暗层次感. 如图 5 所示, 在进行 3D 头发建模时, 美工师通常是通过头发纹理设定头发的颜色和头发之间的明暗层次感, 而 3D 头发的明暗层次感主要体现在相邻像素之间的饱和度和亮度差别(本文称之为“色差”). 因此, 本文拟通过编辑原始的头发纹理来对 3D 头发进行重上色, 并且在重上色过程中保留原始头发纹理中蕴含的明暗层次感.

图 5 基于头发纹理的重上色

为了实现这个目标, 本文设计了一种简单实用的 3D 头发重上色方法, 可有效地保留相邻像素之间的色差. 用 HSB 颜色空间表示头发纹理中的像素颜色. 其中, H 为色相, 对应着通常意义上的颜色, 如红色、绿色或橙色; S 和 B 分别对应饱和度和亮度, 它们决定了颜色的纯度. 对头发进行重上色的公式为

$$\begin{cases} H_{\text{new}}(i, j) = H_{\text{reclr}} - \bar{H} + H(i, j) \\ S_{\text{new}}(i, j) = S(i, j) + \Delta T \\ B_{\text{new}}(i, j) = B(i, j) + \Delta T \end{cases} \quad (4)$$

其中, $H_{\text{new}}(i, j)$ 和 $H(i, j)$ 分别为各像素新的和原始的 H 值; $\bar{H}$ 为原始纹理中所有像素的平均色相; $S_{\text{new}}(i, j)$ 和 $S(i, j)$ 分别为各像素新的和原始的 S 值; $B_{\text{new}}(i, j)$ 和 $B(i, j)$ 分别为各像素新的和原始的 B 值; 用户通过 H_{reclr} 指定新的上色颜色; 而通过 ΔT 控制新上色颜色的纯度. 如图 5 所示, 算法通过式(4)修改了原始头发纹理中各像素的颜色, 在控制新上色颜色纯度的同时, 又保留了相邻像素之间饱和度和亮度的差别, 进而保持原始 3D 头发的明暗层次感.

4 实验及结果分析

为了验证本文系统的有效性, 分别对 3D 人脸跟踪、虚拟试发和头发重上色等算法进行实验. 本文给出相关算法的实现细节, 并围绕算法精度、时间性能以及用户体验等方面进行详细分析.

4.1 3D 人脸跟踪

如第 3.1.1 节所示, 在进行 3D 人脸跟踪时, 本文首先训练了一个轻量级的 MobileNet 检测人脸图像中的 3D 关键点, 通过 Pytorch 平台对深度网络进行训练, 采用随机梯度下降(stochastic gradient descent, SGD)优化器, 批处理大小设为 512. 在训练过程中, 使用文献[14]中提出的加权参数距离和顶点距离作为损失函数. 先使用加权参数距离损失函数训练 50 个 epoch, 学习率为 $10^{-5} \times 0.2^N$;

再使用顶点距离损失函数微调训练 50 个 epoch, 学习率为 $10^{-5} \times 0.2^N$. 其中, 当 epoch 个数小于 30 时, $N=0$; 当 $\text{epoch} < 40$ 时, 每增加一个 epoch, N 值加 1; 当 $\text{epoch} \geq 40$ 时, $N=10$.

根据检测到的人脸 3D 关键点, 再通过第 3.1.2 节中的优化表达式, 获得相邻帧连贯的 3D 人脸跟踪结果. 图 6 所示为不同用户和姿态的 3D 人脸跟踪结果(人脸模型为透明显示). 可以看出, 视觉上, 本文的 3D 人脸跟踪算法较好地逼近了视频帧中对应的 3D 人脸形状及其姿态, 能够满足后继的虚拟试发需要.

图 6 3D 人脸跟踪结果

4.2 真人虚拟 3D 试发

本文系统通过一个普通网络摄像头以 30 帧/s 采集大小为 640×480 像素的视频帧. 在 1 台配置为 Intel i7 2.81 GHz, 32 GB 内存的 PC 机上, 通过单线程方式进行虚拟试发实验. 其中, 在检测人脸 3D 关键点时, 通过 1 块 NVIDIA GeForce GTX 1070 GPU 显卡对深度神经网络进行加速. 实验结果显示, 本文系统的平均处理速度大约为 25 帧/s, 取得了接近实时的试发效果. 在当前实现中并没有对相关代码进行优化处理, 因此系统性能还有很大的提升空间.

图 7 所示为本文系统生成的虚拟试发效果实例, 包含不同的用户和发型以及不同的室内外环境. 本文系统产生了自然且较为真实的 3D 试发效果, 用户可以从不同角度和距离实时观看戴发效果, 就像在镜子前试戴真实假发一样. 因此, 本文系统不仅使用便捷, 并且普适性好, 适用于不同的用户、发型和环境.

4.2.1 试发结果比较

为了进一步验证本文系统的鲁棒性, 还测试在其他复杂环境下的试发效果, 如图 8 所示. 可以看出, 在这些复杂情况下, 本文系统能够鲁棒且稳定地处理外部光照和背景变化, 均生成了视觉上自然的戴发效果.

图 9 所示为同一用户试戴不同发型的效果, 进一步验证了本文系统针对不同发型的鲁棒性. 可以看出, 本文系统生成了非常自然的戴发效果. 根

图 7 本文系统生成的虚拟试发效果实例

a. 不同亮度条件下虚拟试发效果

b. 有他人进出情况下的试发效果

图 8 复杂环境下的试发效果

图 9 同一用户试戴不同的发型

据这些结果, 用户只需要通过虚拟试发, 就可以方便、快捷地发现适合自己的发型.

当前主流的虚拟试发系统是基于 2D 图像的, 因此本文选取国内 3 款具有一定知名度的基于图像的商业试发软件, 对比了它们的试发效果, 结果如图 10 所示. 由于每款软件的可选发型不尽相同, 为了保证对比的公平性, 本文尽可能地选取较接

近的造型和发色. 可以看出, 由于基于图像的试发系统没有考虑用户人脸的实际 3D 方位和尺寸, 当用户人脸非正面时极易产生失真的戴发效果. 此外, 相比于本文系统, 这些软件只允许用户通过单幅图像进行试发, 而不支持实时的虚拟试发.

a. 神奇发型屋 b. 试发型相机 c. 发型设计屋 d. 本文系统

图 10 基于图像的试发系统与本文系统的比较

4.2.2 用户调查研究

为了更加客观、全面地评估和对比本文系统的有效性, 本文进行一项用户调查研究, 邀请了 30 位与本文研究团队无关联的测试者, 均为在校本科生或研究生, 男女各 15 人. 首先选择 4 款本文系统中包含的 3D 头发, 同时选择一个国内具有一定知名度的基于图像的商业试发软件——神奇发型屋, 并从中挑选了 4 个与选择的 3D 头发较接近的发型作为对比; 然后要求每位测试者分别使用本文系统和神奇发型屋试戴上述 8 个发型, 并根据试戴结果, 从自然性、真实度、体验感 3 个方面为这 2 个系统打分(1~9 分). 在实际打分之前对测试者进行集中培训, 统一了相关标准: 自然性是指头发模型与真实人脸的相对位置是否准确、结合是否自然; 真实度是指头发模型的渲染效果是否接近真实头发; 体验感是指用户的试发过程是否接近真实世界的试发场景、是否能帮助用户挑选到合适的发型. 为了尽量避免个人的主观差异, 本文把评分划为 3 个等级: 差(1~3)、中等(4~6)和良好(7~9). 表 1 所示为测试者的评分结果.

从表 1 可以看出, 无论是自然性和真实度, 还是体验感方面, 大部分测试者都对本文系统感到满意, 并且明显超越了神奇发型屋的得分. 特别

地, 相比于自然性和真实度, 本文系统在体验感方面的平均得分更高. 这是因为, 本文系统允许用户如同在镜子前试戴真实假发一样进行实时的虚拟试发, 能够给予用户良好的体验感. 另外, 由于时间和成本的原因, 本文实验中的 3D 头发建模还比较粗糙, 在质感和形状方面与真实头发还有一定差距. 如果采用更高质量建模的 3D 虚拟头发, 本文系统在真实性和自然度以及体验感方面应该还可以取得更高的得分.

4.3 3D 头发重上色

图 11 所示为若干个 3D 头发重上色的效果实例. 本文系统中提供了 2 个滑动条让用户控制头发的重上色效果: 一个滑动条用于控制 H 的值, 即头发的颜色; 另一个滑动条用于控制 ΔT 的值, 即新颜色的纯度. 显然, 本文系统有效地保留了原始纹理中的头发明暗层次感, 同时允许用户通过一种简单直观的方式生成想要的头发颜色, 最终有效地增强了虚拟试发的效果及其体验感. 图 12 所示为神奇发型屋试发软件的重上色效果. 可以看

图 11 3D 头发重上色的效果实例

图 12 基于图像的商业试发软件的重上色效果实例

表 1 不同试发系统的用户评分结果

效果	神奇发型屋				本文系统			
	差	中	好	均值	差	中	好	均值
自然性	9	21	0	4.0	0	7	23	7.1
真实感	6	24	0	4.6	0	5	25	6.8
体验感	9	21	0	4.2	0	2	28	7.4

出, 对比基于图像的试发软件的头发重上色算法, 本文提出的重上色算法虽然针对的是更具有挑战性的 3D 头发重上色问题, 但是取得了不亚于上述软件的重上色效果。

4.4 算法的局限性

图 13 演示了本文系统的一个局限性: 跟踪到的 3D 人脸几何在脖子处与真实用户可能没有精确吻合, 可能造成后面的 3D 头发被不正确地挡住或露出。原因是本文系统在进行 3D 人脸跟踪时考虑的是人脸的面部特征, 如人脸轮廓、嘴巴和眼睛等, 而没有充分考虑用户在脖子处的特征轮廓。另外, 高精度的 3D 人脸重建并不是本文的目标, 大致逼近的 3D 人脸几何大部分情况下能够满足本文的虚拟试发需要, 因此, 本文的 3D 人脸跟踪算法更多考虑的是易用性和实时性。如何在 3D 人脸跟踪的实时性和精确性之间取得最佳平衡, 以实现更加完美的 3D 试发效果, 将是下一阶段的研究重点。

图 13 本文系统局限性的一个演示

5 结 语

本文提出一个基于 3D 人脸跟踪的实时真人虚拟试发系统, 包含以下 3 个方面:

- (1) 基于 3D 人脸跟踪, 设计并实现了一个实时的真人虚拟 3D 试发系统;
- (2) 结合轻量级深度神经网络以及传统优化技术, 实现了视频帧中用户 3D 人脸形状与姿态的实时跟踪;
- (3) 为了增强虚拟试发效果, 提出一种简单实用的 3D 头发重上色算法。

未来工作中拟结合具体应用, 如虚拟社交、在线商城试戴或个人发型设计等, 有针对性地完善本文系统。此外, 拟进一步改进本文提出的 3D 头发佩戴算法, 使头发能够全面贴合用户的头部轮廓, 并在试发时加入碰撞检测和物理模拟功能, 以产生更加逼真的头发飘动效果。

参考文献(References):

- [1] Wu C L, Shiratori T, Sheikh Y. Deep incremental learning for efficient high-fidelity face tracking[J]. *ACM Transactions on Graphics*, 2018, 37(6): Article No.234
- [2] Cootes T F, Taylor C J, Cooper D H, *et al.* Active shape models-their training and application[J]. *Computer Vision and Image Understanding*, 1995, 61(1): 38-59
- [3] Edwards G J, Cootes T F, Taylor C J. Face recognition using active appearance models[C] // *Proceedings of European Conference on Computer Vision*. Heidelberg: Springer, 1998: 581-595
- [4] Cootes T F, Edwards G J, Taylor C J. Active appearance models[C] // *Proceedings of European Conference on Computer Vision*. Heidelberg: Springer, 1998: 484-498
- [5] Cristinacce D, Cootes T. Automatic feature localisation with constrained local models[J]. *Pattern Recognition*, 2008, 41(10): 3054-3067
- [6] Dollár P, Welinder P, Perona P. Cascaded pose regression[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2010: 1078-1085
- [7] Cao X D, Wei Y C, Wen F, *et al.* Face alignment by explicit shape regression[J]. *International Journal of Computer Vision*, 2014, 107(2): 177-190
- [8] Xiong X H, de la Torre F. Supervised descent method and its applications to face alignment[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2013: 532-539
- [9] Ren S Q, Cao X D, Wei Y C, *et al.* Face alignment at 3000 fps via regressing local binary features[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2014: 1685-1692
- [10] Sun Y, Wang X G, Tang X O. Deep convolutional network cascade for facial point detection[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2013: 3476-3483
- [11] Zhang Z P, Luo P, Chen C L, *et al.* Facial landmark detection by deep multi-task learning[C] // *Proceedings of the European Conference on Computer Vision*. Heidelberg: Springer, 2014: 94-108
- [12] Wu Y, Hassner T, Kim K, *et al.* Facial landmark detection with tweaked convolutional neural networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 40(12): 3067-3074
- [13] Kowalski M, Naruniec J, Trzcinski T. Deep alignment network: a convolutional neural network for robust face alignment[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Los Alamitos: IEEE Computer Society Press, 2017: 88-97
- [14] Zhu X Y, Lei Z, Liu X M, *et al.* Face alignment across large poses: a 3D solution[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2016: 146-155
- [15] Bulat A, Tzimiropoulos G. How far are we from solving the 2D

- & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks)[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 1021-1030
- [16] Feng Y, Wu F, Shao X H, *et al.* Joint 3D face reconstruction and dense alignment with position map regression network[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2018: 534-551
- [17] Tu X G, Zhao J, Jiang Z H, *et al.* Joint 3D face reconstruction and dense face alignment from a single image with 2D-assisted self-supervised learning[OL]. [2020-10-15]. <https://arxiv.org/abs/1903.09359>
- [18] Zhang L, Snavely N, Curless B, *et al.* Spacetime faces: high-resolution capture for modeling and animation[M] //Data-Driven 3D Facial Animation. Heidelberg: Springer, 2008: 248-276
- [19] Weise T, Li H, van Gool L, *et al.* Face/off: live facial puppetry[C] //Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. New York: ACM Press, 2009: 7-16
- [20] Bradley D, Heidrich W, Popa T, *et al.* High resolution passive facial performance capture[J]. ACM Transactions on Graphics, 2010, 29(4): Article No.41
- [21] Valgaerts L, Wu C L, Bruhn A, *et al.* Lightweight binocular facial performance capture under uncontrolled lighting[J]. ACM Transactions on Graphics, 2012, 31(6): Article No.187
- [22] Chai J X, Xiao J, Hodgins J. Vision-based control of 3D facial animation[C] //Proceedings of the ACM SIGGRAPH/ Eurographics Symposium on Computer Animation. New York: ACM Press, 2003: 193-206
- [23] Weise T, Bouaziz S, Li H, *et al.* Realtime performance-based facial animation[J]. ACM Transactions on Graphics, 2011, 30(4): Article No.77
- [24] Cao C, Weng Y L, Lin S, *et al.* 3D shape regression for real-time facial animation[J]. ACM Transactions on Graphics, 2013, 32(4): Article No.41
- [25] Cao C, Chai M L, Woodford O, *et al.* Stabilized real-time face tracking via a learned dynamic rigidity prior[J]. ACM Transactions on Graphics, 2018, 37(6): Article No.233
- [26] Pan Y, Mitchell K. PoseMMR: a collaborative mixed reality authoring tool for character animation[C] //Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. Los Alamitos: IEEE Computer Society Press, 2020: 759-760
- [27] Vogel D, Lubos P, Steinicke F. AnimationVR-interactive controller-based animating in virtual reality[C] //Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces. Los Alamitos: IEEE Computer Society Press, 2018: 1-6
- [28] Han X T, Wu Z X, Wu Z, *et al.* Viton: an image-based virtual try-on network[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018:7543-7552
- [29] Minar M R, Tuan T T, Ahn H, *et al.* 3D reconstruction of clothes using a human body model and its application to image-based virtual try-on[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop. Los Alamitos: IEEE Computer Society Press, 2020: 1-4
- [30] Li J, Yang J. Eyeglasses try-on based on improved poisson equations[C] //Proceedings of the International Conference on Multimedia Technology. Los Alamitos: IEEE Computer Society Press, 2011: 3058-3061
- [31] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C] //Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques. New York: ACM Press, 1999: 187-194
- [32] Paysan P, Knothe R, Amberg B, *et al.* A 3D face model for pose and illumination invariant face recognition[C] //Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance. Los Alamitos: IEEE Computer Society Press, 2009: 296-301
- [33] Cao C, Weng Y L, Zhou S, *et al.* FaceWarehouse: a 3D facial expression database for visual computing[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 20(3): 413-425
- [34] Howard A G, Zhu M L, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications[OL]. [2020-10-15]. <https://arxiv.org/abs/1704.04861>
- [35] Moré J J. The Levenberg-Marquardt algorithm: implementation and theory[M] //Waston G A. Numerical analysis. Heidelberg: Springer, 1978: 105-116
- [36] Sorkine O, Rabinovich M. Least-squares rigid motion using SVD[J]. Computing, 2017, 1(1): 1-5