



(12)发明专利

(10)授权公告号 CN 107358648 B

(45)授权公告日 2019.08.27

(21)申请号 201710581769.0

(22)申请日 2017.07.17

(65)同一申请的已公布的文献号
申请公布号 CN 107358648 A

(43)申请公布日 2017.11.17

(73)专利权人 中国科学技术大学
地址 230026 安徽省合肥市包河区金寨路
96号

(72)发明人 张举勇 郭玉东

(74)专利代理机构 北京凯特来知识产权代理有
限公司 11260

代理人 郑立明 郑哲

(51)Int.Cl.

G06T 17/00(2006.01)

G06N 3/08(2006.01)

(56)对比文件

CN 106780662 A,2017.05.31,
CN 104978549 A,2015.10.14,
CN 106778682 A,2017.05.31,
CN 106599878 A,2017.04.26,
US 8582807 B2,2013.11.12,
US 2017083752 A1,2017.03.23,
Elad Richardson等.《Learning Detailed
Face Reconstruction from a Single Image》.
《ResearchGate》.2016,
郑梅兰.《基于学习的人脸图像超分辨率重
建方法》.《计算机工程与应用》.2009,第45卷(第
1期),

审查员 严凯丽

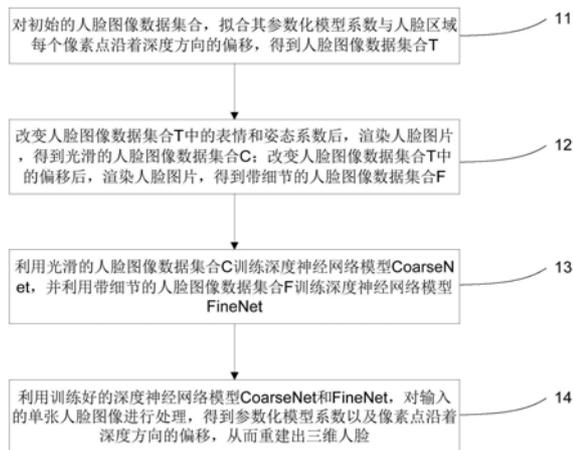
权利要求书2页 说明书6页 附图2页

(54)发明名称

基于单张人脸图像的实时全自动高质量三
维人脸重建方法

(57)摘要

本发明公开了一种基于单张人脸图像的实
时全自动高质量三维人脸重建方法,包括:对初
始的人脸图像数据集合,拟合其参数化模型系数
与人脸区域每个像素点沿着深度方向的偏移,得
到人脸图像数据集合T;改变集合T中的表情和姿
态系数后,渲染人脸图片,得到光滑的人脸图像
数据集合C;改变集合T中的偏移后,渲染人脸图
片,得到带细节的人脸图像数据集合F;利用集合
C训练深度神经网络模型CoarseNet,并利用集合
F训练深度神经网络模型FineNet;利用训练好的
深度神经网络模型CoarseNet和FineNet,对输入
的单张人脸图像进行处理,得到参数化模型系数
以及像素点沿着深度方向的偏移,从而重建出三
维人脸。该方法避免了复杂耗时的优化过程,能
够快速、高质量的重建三维人脸图像。



1. 一种基于单张人脸图像的实时全自动高质量三维人脸重建方法,其特征在于,包括:
对初始的人脸图像数据集合,拟合其参数化模型系数与人脸区域每个像素点沿着深度方向的偏移,得到人脸图像数据集合T;

改变人脸图像数据集合T中的表情和姿态系数后,渲染人脸图片,得到光滑的人脸图像数据集合C;改变人脸图像数据集合T中的偏移后,渲染人脸图片,得到带细节的人脸图像数据集合F;

利用光滑的人脸图像数据集合C训练深度神经网络模型CoarseNet,并利用带细节的人脸图像数据集合F训练深度神经网络模型FineNet;

利用训练好的深度神经网络模型CoarseNet和FineNet,对输入的单张人脸图像进行处理,得到参数化模型系数以及像素点沿着深度方向的偏移,从而重建出三维人脸;

所述对初始的人脸图像数据集合,拟合其参数化模型系数与人脸区域每个像素点沿着深度方向的偏移包括:

利用人脸颜色信息对参数化模型系数 x 进行拟合:

$$E(x) = E_{con} + w_{lan}E_{lan} + w_{reg}E_{reg}$$

上式中, E_{con} 为颜色项衡量由渲染的图片和实际图片的差异; E_{lan} 为关键点约束,使三维点的投影和图像上检测到的关键点尽可能接近; E_{reg} 是正则项,以保证参数化模型系数 x 中的身份 α_{id} 、表情 α_{exp} 及纹理 α_{alb} 系数在合理的范围内; w_{lan} 和 w_{reg} 分别是 E_{lan} 和 E_{reg} 相对于 E_{con} 的权重;

具体的:

$$E_{con}(x) = \frac{1}{|P|} \sum_{p \in P} \|C_S(p) - C_I(p)\|^2$$

其中, C_S 是由 x 渲染的图片, C_I 是输入的图片, P 是所有像素的集合;

$$E_{lan}(x) = \frac{1}{|\mathcal{F}|} \sum_{f_i \in \mathcal{F}} \|f_i - (\Pi R V_i + t)\|^2$$

其中, \mathcal{F} 是关键点的集合, i 为关键点的序号, f_i 为图像平面的二维关键点,

$\Pi = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ 为投影矩阵,其中的 s 为比例系数, V_i 为网格上的三维关键点, $t = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ 为平移

向量, R 为旋转矩阵;

$$E_{reg}(x) = \sum_{j=1}^J \left[\left(\frac{\alpha_{id,j}}{\sigma_{id,j}} \right)^2 + \left(\frac{\alpha_{alb,j}}{\sigma_{alb,j}} \right)^2 \right] + \sum_{k=1}^K \left[\left(\frac{\alpha_{exp,k}}{\sigma_{exp,k}} \right)^2 \right]$$

其中, σ 为对应的主成分方向的标准差, J 为身份与纹理的PCA维数, K 为表情的PCA维数; $\alpha_{id,j}$ 、 $\alpha_{alb,j}$ 分别为第 j 个分量的身份、纹理系数; $\sigma_{id,j}$ 、 $\sigma_{alb,j}$ 分别为第 j 个分量的身份、纹理系数对应主成分方向的标准差; $\alpha_{exp,k}$ 为第 k 个分量的表情系数, $\sigma_{exp,k}$ 为第 k 个分量的表情系数对应主成分方向的标准差;

利用人脸颜色信息对人脸区域每个像素点沿着深度方向的偏移进行拟合,假设所有像素偏移构成向量 d ,则拟合方式如下:

$$E(\mathbf{d}) = E_{\text{con}} + \mu_1 \|\mathbf{d}\|_2^2 + \mu_2 \|\mathbf{Ld}\|_1$$

其中, $\mu_1 \|\mathbf{d}\|_2^2$ 为正则项, 使 \mathbf{d} 接近 0; $\mu_2 \|\mathbf{Ld}\|_1$ 为光滑约束项, \mathbf{Ld} 表示偏移的拉普拉斯, μ_1 和 μ_2 分别是 $\|\mathbf{d}\|_2^2$ 和 $\|\mathbf{Ld}\|_1$ 相对于 E_{con} 的权重。

2. 根据权利要求1所述的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法, 其特征在于, 所述人脸图像数据集合T中包括: 若干彩色人脸图像以及对应的标签;

每一彩色人脸图像均为RGB三通道图像; 标签分为两部分, 一部分为参数化模型系数, 包括身份、表情、纹理、光照以及姿态系数; 另一部分为人脸区域每个像素点沿着深度方向的偏移。

3. 根据权利要求1所述的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法, 其特征在于, 所述光滑的人脸图像数据集合C中包括: 若干彩色人脸图像以及对应的标签;

每一彩色人脸图像均为RGB三通道图像; 标签是参数化模型系数, 包括身份、表情以及姿态系数。

4. 根据权利要求1所述的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法, 其特征在于, 所述带细节的人脸图像数据集合F中包括: 若干灰度人脸图像、粗糙深度图像以及对应的标签;

每一灰度人脸图像均为单通道图像; 粗糙深度图像是由参数化模型系统中的身份、表情与姿态系数得到; 标签为人脸区域每个像素点沿着深度方向的偏移。

5. 根据权利要求1所述的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法, 其特征在于, 所述深度神经网络模型CoarseNet为残差卷积神经网络模型, 依次包括: 输入层、卷积层、非线性层、池化层、全连接层与损失层。

6. 根据权利要求1所述的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法, 其特征在于, 所述深度神经网络模型FineNet为全卷积深度神经网络模型, 依次包括: 输入层、卷积层、非线性层、池化层、去卷积层、融合层与损失层。

基于单张人脸图像的实时全自动高质量三维人脸重建方法

技术领域

[0001] 本发明涉及人脸图像处理技术领域,尤其涉及一种基于单张人脸图像的实时全自动高质量三维人脸重建方法。

背景技术

[0002] 近年来,随着计算机技术和通信技术的快速发展,图像作为一种便捷的信息传递方式,已成为人们获取信息的重要手段。由于人脸图像在身份验证、人机交互、监控、取证、医疗诊断和治疗等领域有着广泛的应用,针对人脸图片的处理与识别已成为机器识别和计算机视觉的主要课题。更进一步,人脸作为身体中最有代表性的部分,人们能够仅凭面部信息就能够获得极为丰富的含义,如人的性别,种族、年龄、健康、情绪,甚至职业等等。

[0003] 在过去,三维人脸重建主要有三种方法:手动建模,仪器采集和基于图像的少量交互或全自动建模。手工建模作为最早的三维建模手段,现在仍然被广泛地使用。手工建模一般需要有经验的专业人士借助Autodesk Maya,3DMax等软件来完成。由于手动建模需要耗费大量的人力与时间,三维人脸建模仪器作为更为方便的方法得到了长期的研究和发展。其中的典型代表有基于结构光和激光扫描技术的精密三维采集仪器和深度相机等。基于精密仪器采集的三维模型,精度可达毫米级,是物体的真实三维数据,可以用来为基于图像的建模方法提供评价数据库,但这些设备一般价格高昂,且使用需经过专业培训,不适合于消费级的市场。最近,市场上出现了Microsoft Kinect,Intel RealSense,PrimeSense等深度相机,研究者可以利用深度相机采集到的深度信息来重建三维模型。和精密采集仪器相比,深度相机价格较低廉且更易于使用,但相比于RGB相机而言,这些设备仍较为少见。基于图像的建模技术是指通过多张或单张人脸图像来重建三维人脸模型。和人脸建模仪器相比,基于图像的建模方法只需要传统RGB相机采集的人脸图片,因此应用场景更为广泛。

[0004] 由于人脸具有较多共性,如具有特定数目的眼睛,嘴巴,鼻子,耳朵且相对位置不变,因此可以建立一个人脸的参数化模型,将复杂的三维人脸参数化到一个低维的空间。传统的基于图像的三维人脸建模技术一般把参数化模型作为先验,利用人脸关键点信息和颜色信息来优化参数化模型的相关系数。但这些方法存在一些问题:基于关键点信息的优化仅利用了稀疏的关键点信息,三维重建精度较低;基于颜色的优化要经过比较耗时的计算,且对光照比较敏感。

发明内容

[0005] 本发明的目的是提供一种基于单张人脸图像的实时全自动高质量三维人脸重建方法,能够快速、高质量的重建三维人脸图像。

[0006] 本发明的目的是通过以下技术方案实现的:

[0007] 一种基于单张人脸图像的实时全自动高质量三维人脸重建方法,包括:

[0008] 对初始的人脸图像数据集合,拟合其参数化模型系数与人脸区域每个像素点沿着深度方向的偏移,得到人脸图像数据集合T;

[0009] 改变人脸图像数据集T中的表情和姿态系数后,渲染人脸图片,得到光滑的人脸图像数据集C;改变人脸图像数据集T中的偏移后,渲染人脸图片,得到带细节的人脸图像数据集F;

[0010] 利用光滑的人脸图像数据集C训练深度神经网络模型CoarseNet,并利用带细节的人脸图像数据集F训练深度神经网络模型FineNet;

[0011] 利用训练好的深度神经网络模型CoarseNet和FineNet,对输入的单张人脸图像进行处理,得到参数化模型系数以及像素点沿着深度方向的偏移,从而重建出三维人脸。

[0012] 利用人脸颜色信息对参数化模型系数x进行拟合:

$$[0013] \quad E(x) = E_{con} + w_{lan} E_{lan} + w_{reg} E_{reg}$$

[0014] 上式中, E_{con} 为颜色项衡量由渲染的图片和实际图片的差异; E_{lan} 为关键点约束,使三维点的投影和图像上检测到的关键点尽可能接近; E_{reg} 是正则项,以保证参数化模型系数x中的身份 α_{id} 、表情 α_{exp} 及纹理 α_{alb} 系数在合理的范围内; w_{lan} 和 w_{reg} 分别是 E_{lan} 和 E_{reg} 相对于 E_{con} 的权重;

[0015] 具体的:

$$[0016] \quad E_{con}(x) = \frac{1}{|P|} \sum_{p \in P} \|C_S(p) - C_I(p)\|^2$$

[0017] 其中, C_S 是由x渲染的图片, C_I 是输入的图片,P是所有像素的集合。

$$[0018] \quad E_{lan}(x) = \frac{1}{|\mathcal{F}|} \sum_{f_i \in \mathcal{F}} \|f_i - (\Pi R V_i + t)\|^2$$

[0019] 其中, \mathcal{F} 是关键点的集合,i为关键点的序号, f_i 为图像平面的二维关键点, $\Pi = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ 为投影矩阵,其中的s为比例系数, V_i 为网格上的三维关键点, $t = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ 为平移向量,R为旋转矩阵;

$$[0020] \quad E_{reg}(x) = \sum_{j=1}^J \left[\left(\frac{\alpha_{id,j}}{\sigma_{id,j}} \right)^2 + \left(\frac{\alpha_{alb,j}}{\sigma_{alb,j}} \right)^2 \right] + \sum_{k=1}^K \left[\left(\frac{\alpha_{exp,k}}{\sigma_{exp,k}} \right)^2 \right]$$

[0021] 其中, σ 为对应的主成分方向的标准差, $\alpha_{id,j}$ 、 $\alpha_{alb,j}$ 分别为第j个分量的身份、纹理系数; $\sigma_{id,j}$ 、 $\sigma_{alb,j}$ 分别为第j个分量的身份、纹理系数对应主成分方向的标准差; $\alpha_{exp,k}$ 为第k个分量的表情系数, $\sigma_{exp,k}$ 为第k个分量的表情系数对应主成分方向的标准差;

[0022] 利用人脸颜色信息对人脸区域每个像素点沿着深度方向的偏移进行拟合,假设所有像素偏移构成向量d,则拟合方式如下:

$$[0023] \quad E(d) = E_{con} + \mu_1 \|d\|_2^2 + \mu_2 \|Ld\|_1$$

[0024] 其中, $\mu_1 \|d\|_2^2$ 为正则项,使d接近0; $\mu_2 \|Ld\|_1$ 为光滑约束项,Ld表示偏移的拉普拉斯, μ_1 和 μ_2 分别是 $\|d\|_2^2$ 和 $\|Ld\|_1$ 相对于 E_{con} 的权重。

[0025] 所述人脸图像数据集T中包括:若干彩色人脸图像以及对应的标签;

[0026] 每一彩色人脸图像均为RGB三通道图像;标签分为两部分,一部分为参数化模型系

数,包括身份、表情、纹理、光照以及姿态系数;另一部分为人脸区域每个像素点沿着深度方向的偏移。

[0027] 所述光滑的人脸图像数据集合C中包括:若干彩色人脸图像以及对应的标签;

[0028] 每一彩色人脸图像均为RGB三通道图像;标签是参数化模型系数,包括身份、表情以及姿态系数。

[0029] 所述带细节的人脸图像数据集合F中包括:若干灰度人脸图像、粗糙深度图像以及对应的标签;

[0030] 每一灰度人脸图像均为单通道图像;粗糙深度图像是由参数化模型系统中的身份、表情与姿态系数得到;标签为人脸区域每个像素点沿着深度方向的偏移。

[0031] 所述深度神经网络模型CoarseNet为残差卷积神经网络模型,依次包括:输入层、卷积层、非线性层、池化层、全连接层与损失层。

[0032] 所述深度神经网络模型FineNet为全卷积深度神经网络模型,依次包括:输入层、卷积层、非线性层、池化层、去卷积层、融合层与损失层。

[0033] 由上述本发明提供的技术方案可以看出,1)由深度神经网络训练得到参数化模型系数及偏移,既利用了完整的人脸图像信息,又避免了复杂耗时的优化过程,加快三维人脸重建的速度。2)通过改变表情、姿态以及皱纹等细节信息来生成训练数据,不仅提供大量带标签的数据用于训练深度神经网络模型,也增加了三维人脸重建方法对表情、姿态以及皱纹的鲁棒性。3)由输入单张图片直接得到参数化模型系数,不用预先检测输入人脸图片的关键点信息。

附图说明

[0034] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域的普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他附图。

[0035] 图1为本发明实施例提供的一种基于单张人脸图像的实时全自动高质量三维人脸重建方法的流程图;

[0036] 图2为本发明实施例提供的改变人脸图像数据集合T中的表情和姿态系数后,得到光滑的人脸图像数据集合C的示意图;

[0037] 图3为本发明实施例提供的改变人脸图像数据集合T中的偏移后,到带细节的人脸图像数据集合F的示意图。

具体实施方式

[0038] 下面结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明的保护范围。

[0039] 在三维人脸重建领域,传统的基于关键点的优化方法速度快,但因只利用了稀疏的关键点信息,重建结果不够精确;基于颜色的优化方法利用了整张图片的颜色信息,重建

结果较为精确,但需要经过复杂的优化,所需时间较长;基于神经网络的方法缺少带标签的训练数据。为此,本发明实施例提供一种基于单张人脸图像的实时全自动高质量三维人脸重建方法,如图1所示,其主要包括如下步骤:

[0040] 步骤11、对初始的人脸图像数据集合,拟合其参数化模型系数与人脸区域每个像素点沿着深度方向的偏移,得到人脸图像数据集合T。

[0041] 本发明实施例中,初始的人脸图像数据集合可以采用现有的人脸数据集300W。三维人脸用参数化模型作为先验,参数化模型系数主要包括:身份、表情、纹理、光照以及姿态系数。

[0042] 其中,几何和纹理信息用PCA表示:

$$[0043] \quad S = \bar{S} + A_{id}\alpha_{id} + A_{exp}\alpha_{exp}$$

$$[0044] \quad T = \bar{T} + A_{alb}\alpha_{alb}$$

[0045] 上式中,S为三维几何信息, \bar{S} 为平均几何信息, A_{id} 和 A_{exp} 分别为身份和表情的PCA基, α_{id} 和 α_{exp} 分别为身份和表情系数;T为纹理信息, \bar{T} 为平均纹理信息, A_{alb} 为纹理的PCA基, α_{alb} 为纹理系数。

[0046] 本领域技术人员可以理解,上式中的 A_{id} 、 A_{exp} 、 A_{alb} 均可通过常规技术获得。

[0047] 投影模型用弱透视投影:

$$[0048] \quad \begin{pmatrix} v_x \\ v_y \end{pmatrix} = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} R \begin{pmatrix} V_x \\ V_y \\ V_z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

[0049] 其中, (v_x, v_y) 和 (V_x, V_y, V_z) 分别是点在图像平面和世界坐标系下的坐标,s为比例系数,R是由欧拉角(pitch,yaw,roll)构造的旋转矩阵, (t_x, t_y) 是平移向量。

[0050] 对于光照,假设是兰伯特反射并用球面调和函数来逼近。这样,图像的渲染由参数化模型系数 $x = \{\alpha_{id}, \alpha_{exp}, \alpha_{alb}, s, pitch, yaw, roll, t_x, t_y, L\}$ 来决定。其中L为球面调和函数的系数。用人脸图像颜色信息对x进行拟合:

$$[0051] \quad E(x) = E_{con} + w_{lan}E_{lan} + w_{reg}E_{reg}$$

[0052] 上式中, E_{con} 是颜色项衡量由x渲染的图片和实际图片的差异; E_{lan} 是关键点约束使三维点的投影和图像上检测到的关键点尽可能接近; E_{reg} 是正则项,保证 α_{id} 、 α_{exp} 、 α_{alb} 在合理的范围内。 w_{lan} 和 w_{reg} 分别是 E_{lan} 和 E_{reg} 相对于 E_{con} 的权重。本发明实例中, w_{lan} 可取10, w_{reg} 可取 $5 \cdot 10^{-5}$ 。

[0053] 具体的:

$$[0054] \quad E_{con}(x) = \frac{1}{|P|} \sum_{p \in P} \|C_S(p) - C_I(p)\|^2$$

[0055] 其中, C_S 是由x渲染的图片, C_I 是输入的图片,P是所有像素的集合。

$$[0056] \quad E_{lan}(x) = \frac{1}{|\mathcal{F}|} \sum_{f_i \in \mathcal{F}} \|f_i - (\Pi R V_i + t)\|^2$$

[0057] 其中, \mathcal{F} 是眼睛轮廓、嘴巴轮廓、人脸轮廓等关键点的集合,i为关键点的序号, f_i

为图像平面的二维关键点, $\Pi = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ 为投影矩阵, V_i 为网格上的三维关键点, $t = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$ 为平移向量。

$$[0058] \quad E_{reg}(\chi) = \sum_{j=1}^J \left[\left(\frac{\alpha_{id,j}}{\sigma_{id,j}} \right)^2 + \left(\frac{\alpha_{alb,j}}{\sigma_{alb,j}} \right)^2 \right] + \sum_{k=1}^K \left[\left(\frac{\alpha_{exp,k}}{\sigma_{exp,k}} \right)^2 \right];$$

[0059] 其中, σ 为对应的主成分方向的标准差。下标 j, k 表示向量的第 j, k 个分量, 本发明实例中, 身份与纹理可用 100 维 PCA 表示 (即 $J=100$), 表情可用 79 维 PCA 表示 (即 $K=79$); 上式中各参数含义如下: $\alpha_{id,j}, \alpha_{alb,j}$ 分别为第 j 个分量的身份、纹理系数; $\sigma_{id,j}, \sigma_{alb,j}$ 分别为第 j 个分量的身份、纹理系数对应主成分方向的标准差; $\alpha_{exp,k}$ 为第 k 个分量的表情系数, $\sigma_{exp,k}$ 为第 k 个分量的表情系数对应主成分方向的标准差。

[0060] 由于参数化模型表达的几何信息比较光滑, 为了表示一些细节比如皱纹等信息, 需要对每个像素点沿着深度方向做个偏移, 假设所有像素偏移构成向量 d , 同样利用人脸颜色信息进行拟合:

$$[0061] \quad E(d) = E_{con} + \mu_1 \|d\|_2^2 + \mu_2 \|Ld\|_1$$

[0062] 其中, $\mu_1 \|d\|_2^2$ 为正则项, 使 d 接近 0; $\mu_2 \|Ld\|_1$ 为光滑约束, Ld 表示偏移的拉普拉斯, μ_1 和 μ_2 分别是 $\|d\|_2^2$ 和 $\|Ld\|_1$ 相对于 E_{con} 的权重。本发明实例中, μ_1 可取 0.001, μ_2 可取 0.3。

[0063] 本发明实施例中, 基于上述方式拟合后的参数化模型系数与偏移来获得人脸图像数据集合 T 。所述人脸图像数据集合 T 中包括: 若干彩色人脸图像以及对应的标签; 每一彩色人脸图像均为 RGB 三通道图像; 标签分为两部分, 一部分为参数化模型系数, 包括身份、表情、纹理、光照以及姿态系数; 另一部分为人脸区域每个像素点沿着深度方向的偏移。

[0064] 步骤 12、改变人脸图像数据集合 T 中的表情和姿态系数后, 渲染人脸图片, 得到光滑的人脸图像数据集合 C ; 改变人脸图像数据集合 T 中的偏移后, 渲染人脸图片, 得到带细节的人脸图像数据集合 F 。

[0065] 如图 2 所示, 改变人脸图像数据集合 T 中的表情和姿态系数后, 渲染人脸图片, 得到光滑的人脸图像数据集合 C 。

[0066] 所述光滑的人脸图像数据集合 C 中包括: 若干彩色人脸图像以及对应的标签; 每一彩色人脸图像均为 RGB 三通道图像; 标签是参数化模型系数, 包括身份、表情以及姿态系数, 即 $\{\alpha_{id}, \alpha_{exp}, s, pitch, yaw, roll, t_x, t_y\}$ 。

[0067] 本领域技术人员可以理解, 人脸图像数据集合 T 中的纹理和光照系数用于渲染人脸图片; 获得的光滑的人脸图像数据集合 C 在后续步骤用来训练深度神经网络模型 CoarseNet, 由于 CoarseNet 只重建了表情和姿态系数, 并没有重建纹理和光照系数, 因此在光滑的人脸图像数据集合 C 中无需纹理和光照系数。

[0068] 如图 3 所示, 改变人脸图像数据集合 T 中的偏移后, 渲染人脸图片, 得到带细节的人脸图像数据集合 F 。偏移的改变是通过细节迁移实现的, 给定带有丰富皱纹细节的原始图片 $source$ 以及较光滑人脸图片 $target$, 用步骤 11 可以算出各自的偏移, 通过迁移 $source$ 偏移的梯度到 $target$ 上来改变 $target$ 的偏移。

[0069] 所述带细节的人脸图像数据集合F中包括:若干灰度人脸图像、粗糙深度图像以及对应的标签;每一灰度人脸图像均为单通道图像;粗糙深度图像是由参数化模型系统中的身份、表情与姿态系数得到;标签为人脸区域每个像素点沿着深度方向的偏移。

[0070] 步骤13、利用光滑的人脸图像数据集合C训练深度神经网络模型CoarseNet,并利用带细节的人脸图像数据集合F训练深度神经网络模型FineNet。

[0071] 本发明实施例中,所述深度神经网络模型CoarseNet为残差卷积神经网络模型,依次包括:输入层、卷积层、非线性层、池化层、全连接层与损失层。

[0072] 本发明实施例中,所述深度神经网络模型FineNet为全卷积深度神经网络模型,依次包括:输入层、卷积层、非线性层、池化层、去卷积层、融合层与损失层。

[0073] 步骤14、利用训练好的深度神经网络模型CoarseNet和FineNet,对输入的单张人脸图像进行处理,得到参数化模型系数以及像素点沿着深度方向的偏移,从而重建出三维人脸。

[0074] 本发明实施例上述方案,相比传统人脸分割方法,主要具有以下优点:

[0075] 1) 由深度神经网络训练得到参数化模型系数及偏移,既利用了完整的人脸图像信息,又避免了复杂耗时的优化过程,加快了三维人脸重建的速度。

[0076] 2) 通过改变表情、姿态以及皱纹等细节信息来生成训练数据,不仅提供大量带标签的数据用于训练深度神经网络模型,也增加了三维人脸重建方法对表情、姿态以及皱纹的鲁棒性。

[0077] 3) 由输入图片直接得到参数化模型的系数,不用预先检测图片的关键点信息。

[0078] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例可以通过软件实现,也可以借助软件加必要的通用硬件平台的方式来实现。基于这样的理解,上述实施例的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例所述的方法。

[0079] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明披露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应该以权利要求书的保护范围为准。

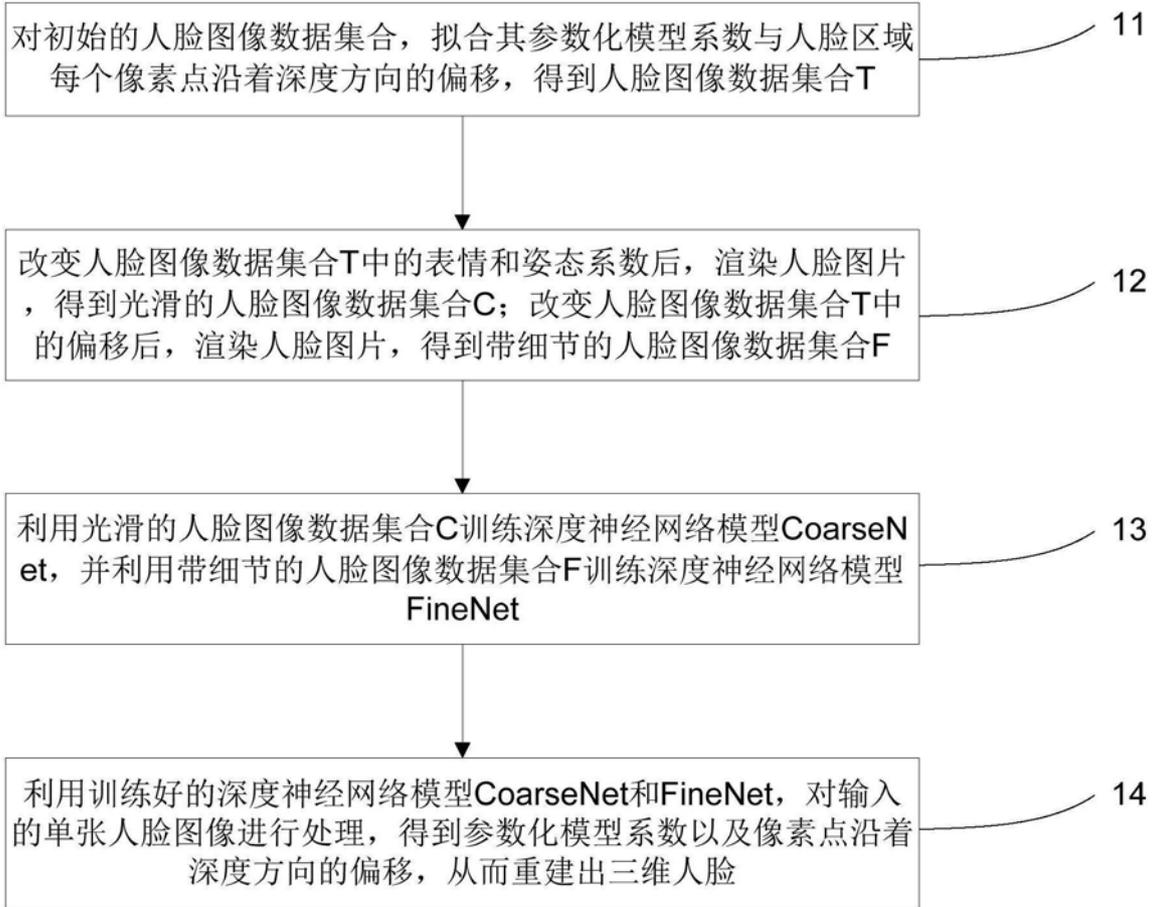


图1

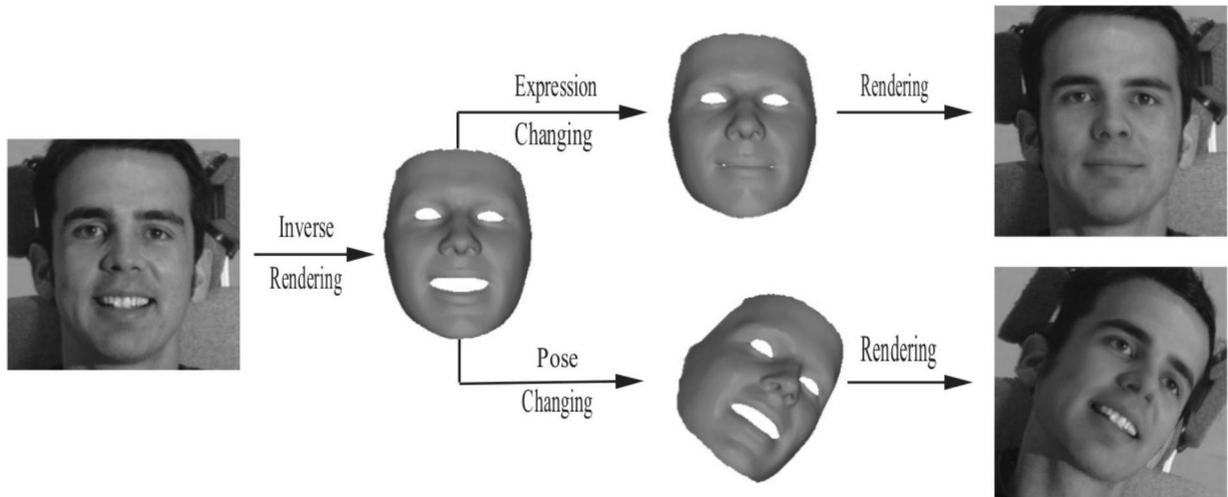


图2

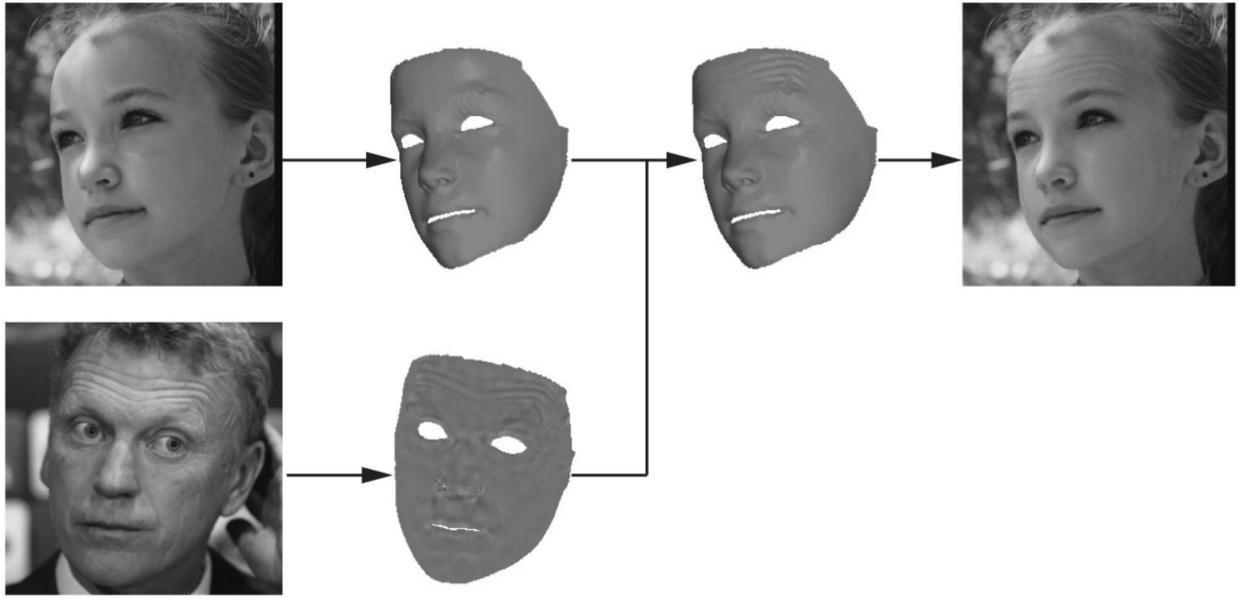


图3