

《物联网与嵌入式技术》

大作业课题报告

基于图像的人脸三维重建综述

学 生 姓 名： 王仪凡

学科、 专业： 建工学院-智慧城市建造与管理专业

学 号： 22306123

完 成 日 期： 2024. 5. 5

大连理工大学

Dalian University of Technology

基于图像的人脸三维重建在人脸分析与娱乐领域里有巨大的应用场景，同时它也可以用于提升人脸关键点检测，人脸识别，人脸编辑等很多任务。1. 什么是人脸三维重建

人脸三维重建就是建立人脸的三维模型，它相对于二维人脸图像多了一个维度，在电影，游戏等领域应用广泛。目前获取人脸三维模型的方法主要包括三种，软件建模，仪器采集与基于图像的建模。

(1) 软件建模作为最早的三维建模手段，现在仍然是最广泛地在电影，动漫行业中应用。(2) 由于手工建模耗费大量的人力，三维成像仪器也得到了长期的研究和发展。基于结构光和激光仪器的三维成像是其中的典型代表，我们熟知的 iPhoneX 中的人脸识别就是基于结构光进行三维人脸重建，正因如此才有 iPhoneX 中的三维人脸表情包。这些基于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，也正好用来为基于图像的建模方法提供评价数据库。不过由于仪器的成本太高，一般的用户是用不上了。(3) 基于图像的建模技术 (image based modeling)，顾名思义，是指通过若干幅二维图像，来恢复图像或场景的三维结构，这些年得到了广泛的研究。我们这里说的人脸三维重建，就特指基于图像的人脸三维重建方法。人脸三维重建的研究已经有几十年的历史，但是基于图像的快速高精度三维人脸重建还没有工业落地，需要研究人员继续努力。

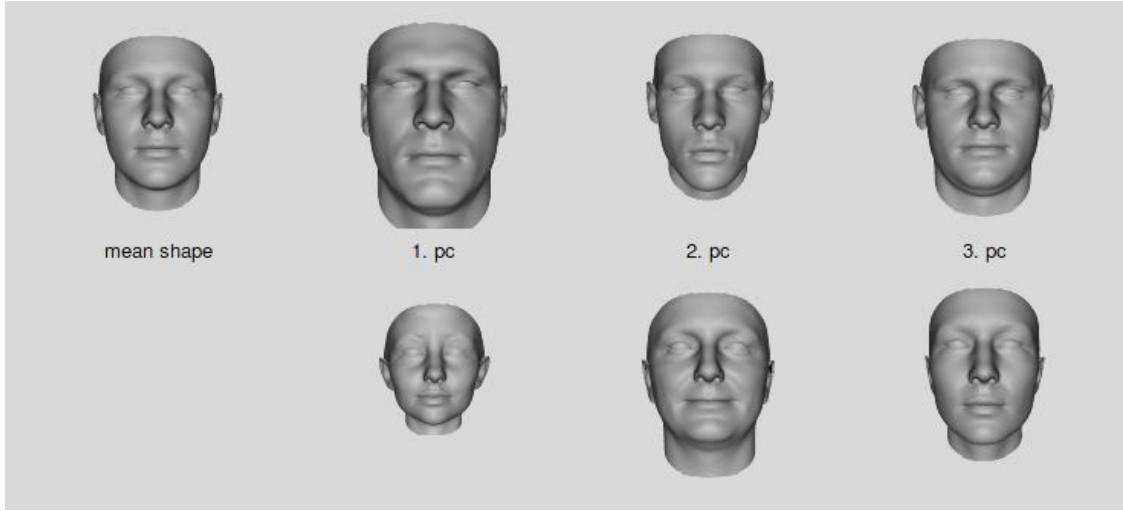
什么是 3DMM 模型，基于人脸图像的三维重建方法非常多，常见的包括立体匹配，Structure From Motion(简称 SfM)，Shape from Shading(简称 sfs)，三维可变形人脸模型 (3DMM)，本文就重点讲述 3D Morphable models(简称 3DMM)，其相关的传统方法和深度学习方法都有较多的研究。

基本思想 3DMM，即三维可变形人脸模型，是一个通用的三维人脸模型，用固定的点数来表示人脸。它的核心思想就是人脸可以在三维空间中进行一一匹配，并且可以由其他许多幅人脸正交基加权线性相加而来。我们所在的三维空间，每一点 (x,y,z) ，实际上都是由三维空间三个方向的基量， $(1,0,0)$ ， $(0,1,0)$ ， $(0,0,1)$ 加权相加所得，只是权重分别为 x,y,z 。

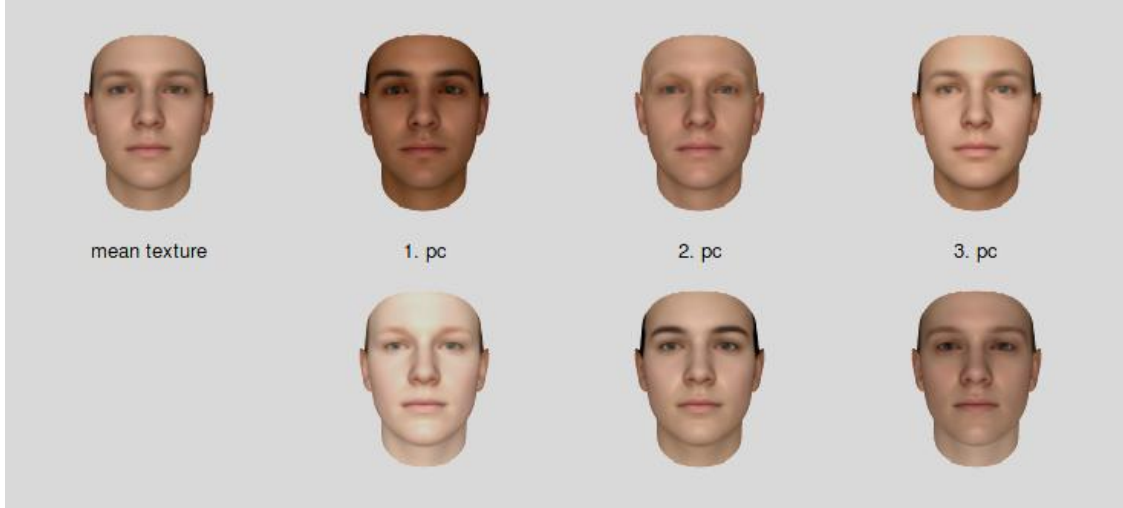
转换到三维空间，道理也一样。每一个三维的人脸，可以由一个数据库中的所有人脸组成的基向量空间中进行表示，而求解任意三维人脸的模型，实际上等价于求解各个基向量的系数的问题。

人脸的基本属性包括形状和纹理，每一张人脸可以表示为形状向量和纹理向量的线性叠加。

形状向量 Shape Vector: $S=(X_1,Y_1,Z_1,X_2,Y_2,Z_2,...,Y_n,Z_n)$ ，示意图如下：



纹理向量 Texture Vector: $T = (R1, G1, B1, R2, G2, B2, \dots, Rn, Bn)$ ，示意图如下：



任意的人脸模型可以由数据集中的 m 个人脸模型进行加权组合如下：

$$\mathbf{S}_{mod} = \sum_{i=1}^m a_i \mathbf{S}_i, \quad \mathbf{T}_{mod} = \sum_{i=1}^m b_i \mathbf{T}_i, \quad \sum_{i=1}^m a_i = \sum_{i=1}^m b_i = 1.$$

其中 $\mathbf{S}_i, \mathbf{T}_i$ 就是数据库中的第 i 张人脸的形状向量和纹理向量。但是我们实际在构建模型的时候不能使用这里的 $\mathbf{S}_i, \mathbf{T}_i$ 作为基向量，因为它们之间不是正交相关的，所以接下来需要使用 PCA 进行降维分解。

- (1) 首先计算形状和纹理向量的平均值。
- (2) 中心化人脸数据。
- (3) 分别计算协方差矩阵。

(4) 求得形状和纹理协方差矩阵的特征值 α , β 和特征向量 s_i , t_i 。

其中第一项是形状和纹理的平均值, 而 s_i , t_i 则都是 S_i , T_i 减去各自平均值后的协方差矩阵的特征向量, 它们对应的特征值按照大小进行降序排列。

等式右边仍然是 m 项, 但是累加项降了一维, 减少了一项。 s_i , t_i 都是线性无关的, 取其前几个分量可以对原始样本做很好的近似, 因此可以大大减少需要估计的参数数目, 并不失精度。

基于 3DMM 的方法, 都是在求解这几个系数, 随后的很多模型会在这个基础上添加表情, 光照等系数, 但是原理与之类似。

2.2 3DMM 模型求解方法

基于 3DMM 求解三维人脸需要解决的问题就是形状, 纹理等系数的估计, 具体就是如何将 2D 人脸拟合到 3D 模型上, 被称为 Model Fitting, 这是一个病态问题。经典的方法是 1999 年的文章“A Morphable Model For The Synthesis Of 3D Faces”, 其传统的求解思路被称为 analysis-by-Synthesis, 如下:

(a) 初始化一个 3 维的模型, 需要初始化内部参数 α , β , 以及外部渲染参数, 包括相机的位置, 图像平面的旋转角度, 直射光和环境光的各个分量, 图像对比度等共 20 多维, 有了这些参数之后就可以唯一确定一个 3D 模型到 2D 图像的投影。

(b) 在初始参数的控制下, 经过 3D 至 2D 的投影, 即可由一个 3D 模型得到 2 维图像, 然后计算与输入图像的误差。再以误差反向传播调整相关系数, 调整 3D 模型, 不断进行迭代。每次参与计算的是一个三角晶格, 如果人脸被遮挡, 则该部分不参与损失计算。

(c) 具体迭代时采用由粗到精的方式, 初始的时候使用低分辨率的图像, 只优化第一个主成分的系数, 后面再逐步增加主成分。在后续一些迭代步骤中固定外部参数, 对人脸的各个部位分别优化。

对于只需要获取人脸形状模型的应用来说, 很多方法都会使用 2D 人脸关键点来估计出形状系数, 具有更小的计算量, 迭代也更加简单, 另外还会增加一个正则项, 所以一个典型的优化目标是如下:

对于 Model fitting 问题来说, 除了模型本身的有效性, 还有很多难点。

(1) 该问题是一个病态问题, 本身并没有全局解, 容易陷入不好的局部解。

(2) 人脸的背景干扰以及遮挡会影响精度, 而且误差函数本身不连续。

(3) 对初始条件敏感, 比如基于关键点进行优化时, 如果关键点精度较差, 重建的模型精度也会受到很大影响。

2.3 3DMM 模型的发展

要使用 3DMM 模型来完成人脸重建，首先就需要一个数据库来建立人脸基向量空间，Blanz 等人在 1999 年的文章[1]中提出了采集方法，但是没有开源数据集，Pascal Paysan 等人在 2009 年使用激光扫描仪精确采集了 200 个人脸数据得到了 Basel Face Model 数据集[2] (简称 BFM 模型)，基本信息如下：

(1) 采用 ABW-3D 结构光系统进行采集，采集时间约 1s，相比于激光平均 15s 的采集方案更加具有优势。整个数据集包含 200 张三维的人脸，其中 100 张男性，100 张女性，大部分为高加索人脸。年龄分布 8~62 岁，平均年龄 24.97 岁，体重 40~123 千克，平均 66.48 千克。每一个人都被采集 3 次中性表情，并选择其中最自然的一次。

(2) 在对采集后的点进行处理的过程中，模型的每一个点的位置都进行了精确匹配，也就是说每一个点都有实际的物理意义，比如属于右嘴角等。经过处理后，每一个模型由 53490 个点描述。

Basel Face Model 数据集只有 200 个人，而近期研究者基于此模型采集了 9663 个人得到 LSFM 模型[3]，能够进一步提升表达能力。

2009 年发布的 Basel Face Model 版本中没有表情系数，而 2017 年发布的版本 BFM 2017[4]中提供了表情系数，同样还是一个线性模型。

当然了，在国内也有一个著名的数据集，就是 FaceWarehouse[5]，不过不开源，一般研究者拿不到数据。

当然也有一些商业号称会开源更好的模型，这个大家可以拭目以待。人脸的三维模型数据之所以不公开，是因为使用高精度的三维模型可以很容易仿真真实人脸，容易发生安全事故。

当前基于 3DMM 的表情模型主要有两个思路，分别是加性模型和乘性模型。加性模型就是线性模型了，将表情作为形状的一个偏移量， E_s ， E_e 分别表示形状和表情基， W_s ， W_e 分别表示对应的系数。

但是因为表情也会改变人脸的形状，因此它和形状并非完全正交的关系，所以有的研究者提出了乘性模型。其中 d_e 是一个表情迁移操作集合，第 j 个操作即为 T_j ， δ 都是校准向量。

另一方面，纹理模型也被称为表观模型，它相对于形状模型来说更加复杂，受到反射率和光照的影响，不过大部分的 3DMM 模型不区分两者，所以我们将其视为一个因素，即反射率。

光照模型通常采用的是球面模型，光照模型比较复杂，我们这里就不列出具体的表达式，大家可以自行阅读相关论文。

在 2009 年提出的 BFM 模型中，纹理模型是一个线性模型，即由多个纹理表情基进行线性组合。后续的研究者们在整个基础上增加了纹理细节，用于仿真脸部的皱纹等。