

引用格式:苏越阳,万静,易军凯. 基于单幅正面照和统计模型的三维人脸重建方法研究[J]. 北京化工大学学报(自然科学版), 2019, 46(1): 104–109.

SU YueYang, WAN Jing, YI JunKai. A 3D face reconstruction based on a single frontal image and a statistical model[J]. Journal of Beijing University of Chemical Technology (Natural Science), 2019, 46(1): 104–109.

基于单幅正面照和统计模型的三维人脸重建方法研究

苏越阳 万静* 易军凯

(北京化工大学 信息科学与技术学院, 北京 100029)

摘要:为提升单张图片三维重建的效率、提高重建的精准度,提出了一种基于单张图片和统计模型的三维人脸重建技术。首先基于智能行为理解组(intelligent behaviour understanding group, IBUG),采用监督下降算法(SDM)进行训练来建立二维人脸特征点参数模型;再应用该参数模型对给定图片进行特征点提取;然后基于北京工业大学3D数据库(BJUT-3D)进行三维人脸统计模型的构建;最后,以三维人脸统计模型为基础,使用学习因子自适应梯度下降法对能量函数进行迭代优化,得到统计模型参数化向量,使用该向量来调整统计模型,实现与给定图片相匹配的三维人脸模型构建。实验证明,与现有的方法相比,本文方法重建出的三维人脸模型具有更高的精确度和自适应度。

关键词:统计模型;三维人脸模型;图像配准;监督下降法(SDM);相机投影变换

中图分类号: TP391.4 **DOI:** 10.13543/j.bhxbzr.2019.01.016

引言

随着计算机图形学的快速发展,三维建模被广泛应用于动漫游戏、影视广告、安全认证以及电子医疗档案建立等领域。三维人脸重建技术作为计算机图形学的一个重要方向,已经成为研究的热点内容。由于人脸生理结构、几何复杂性以及光照特性等因素的综合影响,三维人脸建模成为一个极具挑战性的难题。

传统的三维人脸重建方法是以肌肉特征为主要研究对象,通过多边形网格模拟人脸,用网格形变模拟肌肉力的形变^[1]。Chu等^[2]构建的参数化人脸模型通过结构参数以及动作参数描述变化的面部特征,应用主成分分析降低数据复杂度,同时保留了足够的数据差异;Izadi等^[3]通过Kinect相

机的彩色传感器提取人脸面部的特征点,以此作为基准对一般人脸模型进行调整,实现与单幅深度图片的配准。Tran等^[4]将深度神经网络应用于三维人脸建模,提高了模型重建的效率,降低了实现的复杂度,该方法广泛应用于视频会议等实时性要求高的领域,但是使用单张图片进行人脸重建的效果不佳;Li等^[5]将局部数据集与公共数据集相结合,改进现有的三维形变模型(3D morphable model, 3DMM)拟合方法,然后使用卷积神经网络(convolutional neural network, CNN)来提高重建效果,具有很好的鲁棒性,一定程度上提高了单张图片的建模精度;Blanz等^[6]通过统计学习构建三维人脸形变模型,包括三维形状模型和对应的二维纹理模型,并采用牛顿迭代法进行模型处理,实现指定人脸特征的三维建模,保留了具体的纹理信息和精确的三维形状。

随着计算机图形学的不断发展,不断有新的人脸重建方法被提出,但是重建精度和自适应度都不太理想,而且数据处理速度慢。针对这些问题,本文提出了一种基于统计模型和单张正面照片的三维人

收稿日期: 2018-04-17

基金项目: 通用技术基础研究联合基金(U1636208)

第一作者: 男, 1992年生, 硕士生

* 通信联系人

E-mail: wanj@mail.buct.edu.cn

脸重建技术。该方法使用监督下降法 (supervised descent method, SDM) 训练得到二维人脸特征点参数模型, 再利用该模型提取人脸特征点, 最后通过对三维人脸统计模型的迭代优化, 使得到的模型最大程度地匹配该组特征点, 实现指定图像的三维人脸模型重建。

1 基于单幅正面照的三维人脸重建的基本原理

基于单张正面照片的三维人脸重建技术主要包括二维人脸特征点参数模型的建立以及三维人脸统计模型的生成两个方面。

目前实现人脸特征点标定的算法主要包括局部二元特征 (local binary feature, LBF) 算法、SDM 算法以及主动外观模型 (fast active appearance model, FAAM) 算法等。为了实现三维人脸重建的全自动化, 需要进行人脸特征点的自动标定, 本文使用基于 SDM 算法的特征点标定技术^[7-8], 该方法对于正面、侧面都具有较好的标定效果。求得人脸特征点后, 再按照一定规则不断调整三维统计模型的向量化参数, 使调整后的模型尽量接近这组特征点, 直到获得一个满意的效果^[8-11]。

2 基于统计模型的单幅正面照三维人脸重建

2.1 原始三维人脸模型预处理

为了得到精确的三维统计模型, 本文采用激光扫描得到准确度较高的人脸模型, 然后对原始三维人脸模型进行处理, 处理过程包括配准和序列化两个部分。图像配准^[12]是三维模型重建过程中的重要步骤, 配准的准确度对重建结果的精确度有很大影响。一般来说, 经过激光扫描后的三维模型 $I(h, \phi)$ 均未经过配准, 为了实现序列化, 本文采用光流法进行图像配准。光流法是一种图像运动的表达形式, 用光流场描述图像序列中的图像亮度模式变化来体现实际场景中的运动信息。

通过计算出流场 $(\delta h(h, \phi), \delta \phi(h, \phi))$ 中的最小范数 $\|I_1(h, \phi) - I_2(h + \delta h, \phi + \delta \phi)\|$, 可以确定两个三维模型中的对应点, 从而实现模型配准。经过配准后的人脸模型可以表示为分布在人脸曲面上的序列化点集

$$S_{\text{reg}} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_n, Y_n, Z_n)^T \quad (1)$$

式中 X_n, Y_n, Z_n 为第 n 个点的坐标。

三维统计模型由配准后的人脸模型的序列化点集的线性组合表示

$$S_{\text{mm}} = \sum_{i=1}^m a_i S_{\text{reg}_i} \quad (2)$$

式中 $\sum_{i=1}^m a_i = 1$ 。人脸模型集合 $\{S_{\text{reg}_i}\}$ 经过向量 $\mathbf{a} = (a_1, a_2, \dots, a_m)^T$ 参数化后, 就可以表示任意三维人脸模型, 参数化后的模型表示为 $S_{\text{mm}}(\mathbf{a})$, 所以通过调整向量 \mathbf{a} 就能够得到任意人脸模型。

经过配准和序列化后的三维人脸模型 S_{reg} 是一个 $3n$ 维的向量, 当模型的顶点非常多时会出现维度灾难, 而且参数向量 $\mathbf{a}' = (a'_1, a'_2, \dots, a'_n)^T$ 的计算也会非常复杂和耗时。本文引入主成分分析 (principal component analysis, PCA) 法进行降维, 主要内容是提取模型协方差矩阵的前 $m-1$ 个主成分特征向量, 由此得到的模型表示为

$$S_{\text{pea}} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, \dots, X_{m-1}, Y_{m-1}, Z_{m-1})^T \quad (3)$$

降维后的特征人脸的最优参数化向量 $\mathbf{a}'' = (a''_1, a''_2, \dots, a''_{m-1})^T$ 。所以降维处理之后的三维人脸模型可以表示为

$$S_{\text{mod}} = \bar{S} + \sum_{i=1}^{m-1} a''_i S_{\text{pea}_i} \quad (4)$$

式中 \bar{S} 表示平均人脸模型。

2.2 SDM 算法

本文应用 SDM 算法^[13]进行人脸特征检测, 标定给定图片上的特征点。SDM 算法相较于传统的梯度下降算法具有很大的优势, 它通过目标函数的优化求解进行特征检测, 迭代过程中的步长是通过样本训练而非计算所得, 所以有效避免了计算 Hessian 矩阵的逆等复杂过程, 提高了收敛效率。

SDM 算法的实现过程分为训练和检测两个部分。定义如下变量: X_* 为人脸特征点检测任务中的最优解坐标, X_0 为初始化时的特征点坐标矩阵; x 表示图像的特征点, $d(x)$ 为特征点坐标, p 为特征点数目, 则 $d(x) \in \mathbf{R}^{p \times 1}$ 为图像中的 p 个特征点坐标; 在特征点 x 周围提取的尺度不变 (scale invariant feature transform, SIFT) 特征^[14]可以记为 $\phi_* = h(d(x))$ 。本文选取的 SIFT 特征是一种稳定的局部特征, 在旋转

角度变化、亮度改变等情况下能保持不变,并且含有丰富的信息。给定一组训练图片集如图 1 所示,图 1(a)是手工标记的人脸特征点坐标,即训练过程中的最优解 X_* ,图 1(b)是初始化时的特征点坐标 X_0 。

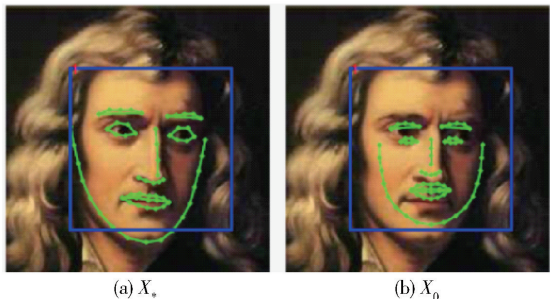


图 1 特征点最优解和初始值

Fig. 1 The optimal solution and initial value of the feature point

在 SDM 的训练过程中, ϕ_* 为已知量, X_0 则通过对训练样本中手动标注的人脸特征点坐标取平均值得到,训练的目标是从训练集中获取梯度下降方向 R_k 和 b_k 。训练过程中,求取人脸特征点的过程可以抽象为线性回归问题,该问题的输入特征为在 X_0 处的 SIFT 特征 ϕ_* ,其目的是使 X_0 收敛到 X_* 过程中的迭代步长趋于 0。由此得模型训练过程中的目标函数为

$$\operatorname{argmin}_{R_0, b_0} \sum_{d^i} \|\Delta x_*^i - R_0 \phi_0^i - b_0\|^2 \quad (5)$$

求得第一次的迭代系数 R_0 和 b_0 之后,一系列的系数 R_k 和 b_k 可以根据式(6)计算

$$\operatorname{argmin}_{R_k, b_k} \sum_{d^i} \|\Delta x_*^i - R_k \phi_k^i - b_k\|^2 \quad (6)$$

在检测过程中,每个特征点需要提取 128 维的 SIFT 特征,首先根据训练过程中的平均形状来确定初始形状,然后利用得到的 R_0 和 b_0 进行迭代,对初始形状进行优化更新,直至获取到准确的特征点位置。SDM 进行人脸特征点检测的过程可以看作是对目标函数 f 在 Δx 基础上的最小化

$$f(x_0 + \Delta x) = \|h(d(x_0 + \Delta x)) - \phi_*\|^2 \quad (7)$$

2.3 三维人脸模型重建

经过 2.2 节获取到人脸特征点之后,以特征点集为目标约束条件,不断调整人脸统计模型参数化向量的取值,使统计模型经过相机变换投影到二维平面上的点高度接近特征点集。

人脸生成过程可以分为三维统计模型的投影变

换和相似度评估两个部分。设三维人脸统计模型为 S_{mm} ,特征点集为 $P = \{P_i\}$,含有 r 个特征点,相机投影变换参数为 ρ 。在三维模型向二维平面投影的过程中,需要操作人员手动初始化相机参数 ρ 和模型参数化向量 a 。二维图像可以表示为

$$I_{\text{mod}} = P(S_{\text{mm}}, \rho, a) \quad (8)$$

式中, I_{mod} 表示渲染图像特征点集。

在相似度评估过程中,将重建的人脸模型和真实人脸模型的差异抽象为输入图像特征点集和渲染图像特征点集之间的欧氏距离平方和,即相似度偏差函数,定义为

$$E = \|I_{\text{input}} - I_{\text{mod}}\|^2 \quad (9)$$

式中, I_{input} 为输入图像特征点集。偏差函数的值越小,则相似度越高,重建效果越好。

在偏差函数最小化过程中,采用梯度下降法调整参数化向量

$$a'_j = a_j - \lambda_j \frac{\partial E}{\partial a_j} \quad (10)$$

式中, a_j 为原参数化向量, a'_j 为调整后的参数化向量, $\frac{\partial E}{\partial a_j}$ 为梯度矢量, λ 为逼近因子。

这里假设相机参数的估计值是正确的,那么 ρ 可以保持不变。 λ 的值由经验给出, λ 值越大,在优化过程中的逼近速度越快,但是 λ 值过大会影响模型的精确度,还可能使偏差函数不收敛;反之,设置过小则会导致优化速度慢、花费时间长甚至陷入局部极值。为了选取最合适的 λ 值,本文使用了动态调节策略:①初始化时根据经验选取一个较大的逼近因子 λ_{init} ;②迭代过程中如果梯度方向发生变化就以 $\lambda' = k\lambda$ 缩小 λ ,其中 k 是每次缩小的比例,一般取 0.7 ~ 0.8。

3 实验及结果分析

3.1 实验过程

本文采用北京工业大学 3D(BJUT-3D)人脸数据库^[15-16],该数据库共包含 1 200 个人脸样本,其中 500 个人脸样本数据是对外公布的,包含男女各一半,年龄范围 16 ~ 49 岁。本文以此数据库作为数据基础进行三维人脸统计模型的构建。BJUT-3D 人脸数据库采用 Cyber Ware 3030 RGB/PS 激光扫描仪获取人脸三维模型的原始数据,扫描过程中被扫描者要除去眼镜等可能影响识别度的

头、面部饰品,保持中性面部表情和端正的坐姿。图 2 是 Cyber Ware 扫描仪获取到的三维人脸图像。



图 2 Cyber Ware 扫描仪及其获得的三维人脸数据
Fig. 2 The Cyber Ware scanner and 3D face data acquired using this device

在实验的第二个阶段,采用 IBUG 提供的二维人脸数据库及特征点进行 SDM 模型的训练。该二维人脸数据库包含了 135 张人脸图片,每张图片被标注了 68 个特征点,存储在. pts 文件中,其标注规则如图 3 所示。

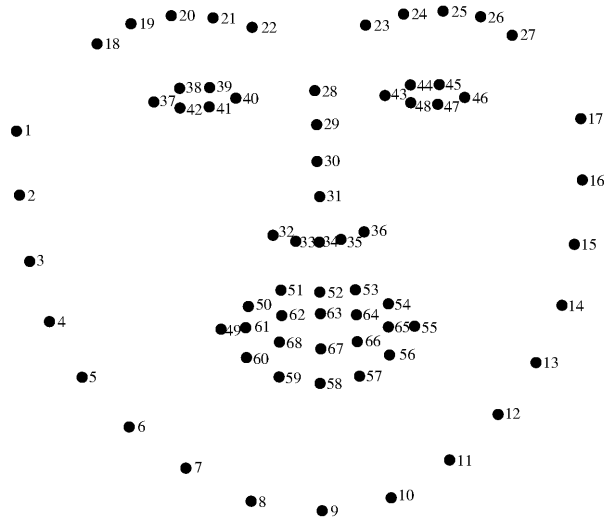


图 3 IBUG 标注规则
Fig. 3 IBUG labeling rules

训练完成后,使用 SDM 模型识别输入图片的特征点,并以该组特征点为基准不断调整优化统计模

型各分量的系数,使得最终得到的模型尽量接近该组特征点。为了得到更好的效果,需要重复迭代这一过程,直至生成具有真实感的三维人脸模型。具体流程如图 4 所示。

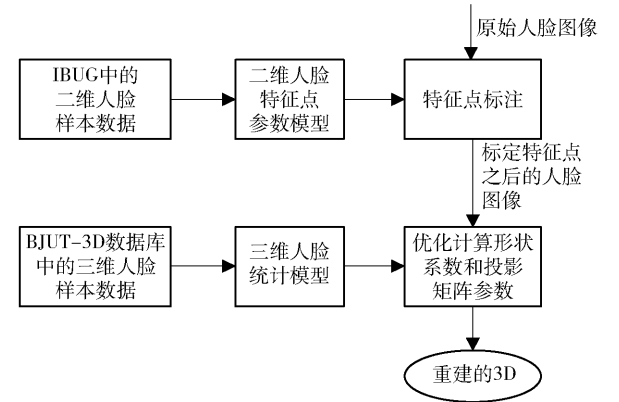


图 4 基于单张图像的 3D 人脸重建流程
Fig. 4 The 3D face reconstruction process based on single image

3. 2 结果分析

3. 2. 1 三维人脸重建过程

针对三维人脸重建过程中的输入需求、形状建模、模型构建和自动化程度几个方面,与 Blanz 方法进行了对比,结果如表 1 所示。

表 1 本文方法与 Blanz 方法的重建过程比较
Table 1 Comparison of the reconstruction process of this method and the Blanz method

方法	输入需求	形状建模	模型构建	自动化程度
Blanz	任意角度图片	特征点对应	稠密模型组合	手动初始化
本文方法	单张正面照片	特征点对应	统计模型变形	全自动

3. 2. 2 三维人脸重建效率

三维人脸重建过程耗费的时间是实际应用中需要考虑的关键指标。本文选取 50 个样本进行三维人脸重建,并与 Blanz 方法的重建效率比较,结果如表 2 所示。

表 2 本文方法与 Blanz 方法重建效率比较
Table 2 Comparison of the reconstruction efficiency of this method and the Blanz method

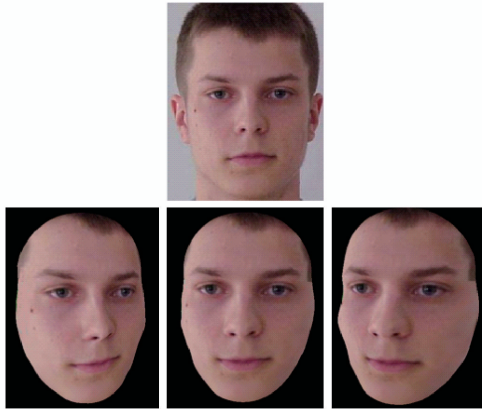
方法	实验环境	时间/min
Blanz	2G pentium4 Workstation	4. 5
本文方法	intel core i5 2. 50GHZ 处理器	0. 30

从实验结果可以看出,相比 Blanz 方法,本文的

方法耗时短,重建效率提高 14 倍,具有很高的实用价值。

3.2.3 三维人脸重建精确度

三维人脸重建的精确度也是重要的评价指标,高精度是所有重建算法的改进目标。本文方法与三维人脸形变方法的重建效果比较如图 5 所示。



(a) 本文方法重建效果



(b) 基函数取 $e^{-\frac{1}{2}}$ 的三维人脸形变方法重建效果

图 5 本文方法与三维人脸形变方法重建精度比较

Fig. 5 Comparison of the reconstruction accuracy of this method and the 3D face deformation method

从结果可以看出,该方法生成的人脸模型和真实人脸非常接近,轮廓饱满,刻画细腻,且具有很多细节特征,完全可以满足公安刑侦、影视动漫、安全认证、视频会议等方面的应用要求。

4 结束语

本文提出的基于单幅正面照和统计模型的三维人脸重建技术在重建效率和准确率方面都有明显的提高。实验结果得到的人脸模型与真实人脸具有较高的相似度,可满足相关专业领域的绝大部分需求。

本文结果还有待从以下方面进行进一步完善和

优化:①构建统计模型时使用的评估偏差函数包含信息较为单一,对纹理信息考虑不足,下一步可以将位置信息和纹理信息综合考虑,以提升重建模型的真实感;②以整个人脸为基本单位导致对局部特征的调整难度变大,下一步将考虑通过对人脸进行区域分割并对各部分分别构建统计模型,来实现整体重建效果的提高。

参考文献:

- [1] 李想. 基于单张正面图象的三维人脸建模研究[D]. 苏州: 苏州大学, 2016.
LI X. Research on 3D face modeling based on single front image [D]. Suzhou: Soochow University, 2016. (in Chinese)
- [2] CHU C H, WANG I J, WANG J B, et al. 3D parametric human face modeling for personalized product design[J]. Advanced Engineering Informatics, 2017, 32: 202–223.
- [3] IZADI S, KIM D, HILLIGES O, et al. Kinect fusion: real-time 3D reconstruction and interaction using a moving depth camera [C] // ACM Symposium on User Interface Software & Technology. Santa Barbara, 2011: 559–568.
- [4] TRAN A T, HASSNER T, MASI I, et al. Regressing robust and discriminative 3D morphable models with a very deep neural network[C] // IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1493–1502.
- [5] LI F M, CHEN K, LIU X H. 3D face reconstruction based on convolutional neural network[C] // International Conference on Intelligent Computation Technology and Automation. Changsha, 2017: 71–74.
- [6] 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. Los Angeles, 1999: 187–194.
- [7] XIONG X, TORRE F D L. Global supervised descent method[C] // IEEE Conference on Computer Vision and Pattern Recognition. Boston, 2015: 2664–2673.
- [8] YANG H, JIA X H, PATRAS I, et al. Random subspace supervised descent method for regression problems in computer vision [J]. IEEE Signal Processing Letters, 2015, 22(10): 1816–1820.
- [9] HUBER P, FENG Z H, CHRISTMAS W, et al. Fitting 3D morphable models using local features[C] // IEEE International Conference on Image Processing. Quebec, 2015.

- [10] HU G S, YAN F, KITTLER J, et al. Efficient 3D morphable face model fitting[J]. Pattern Recognition, 2017, 67: 366–379.
- [11] MOEINI A, MOEINI H, FAEZ K. Pose-invariant facial expression recognition based on 3D face reconstruction and synthesis from a single 2D image[C]//International Conference on Pattern Recognition. Stockholm, 2014: 1746–1751.
- [12] CHEN Z C. Sensorcam cameras and image registration [J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 104(27): 11483.
- [13] XIONG X, TORRE F D L. Supervised descent method and its applications to face alignment[C]//IEEE Conference on Computer Vision and Pattern Recognition. Portland, 2013: 532–539.
- [14] AMERINI I, BARNI M, CALDELLI R, et al. Counterforensics of SIFT-based copy-move detection by means of keypoint classification[J]. Eurasip Journal on Image & Video Processing, 2013(1): 1–17.
- [15] 许磊, 黎智辉, 王俊娟, 等. 高精度三维人脸图像数据库[J]. 刑事技术, 2015, 40(2): 94–97.
XU L, LI Z H, WANG J J, et al. High-precision 3D face image database[J]. Criminal Technology, 2015, 40(2): 94–97. (in Chinese)
- [16] SANG G L, LI J, ZHAO Q J. Pose-invariant face recognition via RGB-D images[J]. Computational Intelligence and Neuroscience, 2016(3): 1–9.

A 3D face reconstruction based on a single frontal image and a statistical model

SU YueYang WAN Jing* YI JunKai

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: 3D face model reconstruction based on a single frontal photo is widely used in practical applications. In order to improve the efficiency and accuracy of 3D reconstruction based on single frontal image, this work proposes an effective approach to reconstruct a 3D face model based on a single image and a statistical model. First, we train with a SDM algorithm to establish a two-dimensional human face feature point parameter model based on the IBUG two-dimensional face database. The parameter model is used to extract feature points from a given image. A 3D face statistics model is then constructed based on the BJUT-3D database. In order to determine the parametric vector of the statistical model, we optimize the energy function with a learning factor adaptive gradient descent method iteratively based on the 3D face statistics model. Finally, we adjust the statistical model to reconstruct the 3D face model of the given image. The experimental results show that our face reconstruction method has higher accuracy and adaptability than representative commonly used algorithms.

Key words: statistical model; 3D face model; image registration; supervised descent method(SDM); camera projection transformation

(责任编辑:汪 琴)