

基于样例的面部表情基生成

高娅莉, 谭光华, 郭松睿, 范晓伟

(湖南大学信息科学与工程学院, 长沙 410082)

摘 要: 针对现有 blendshape 表情基的生成多依赖手工建模和微调的问题, 提出一种基于样例的表情基自动生成方法。以一组通用的面部模型为先验模型, 实现每个 blendshape 的表情语义逐步求精。通过该方法生成的 blendshape 表情基能够将通用模型的表情语义和表情动态映射到目标模型上, 从而实现模型在应用上的可扩展性, 使用户能够按照自身需求逐步计算出更为丰富的面部表情模型。实验结果表明, 在梯度空间建立形变约束, 并将其转化为一个关于 blendshape 的优化问题, 能更加有效地控制 blendshape 表情基的变化趋势, 并且计算时间更短, 真实感更强。

关键词: blendshape 表情基; 基于样例; 迭代优化; 梯度空间; 真实感; 表情动画

中文引用格式: 高娅莉, 谭光华, 郭松睿, 等. 基于样例的面部表情基生成[J]. 计算机工程, 2015, 41(3): 258-261.

英文引用格式: Gao Yali, Tan Guanghua, Guo Songrui, et al. Facial Expression Radical Generation Based on Example[J]. Computer Engineering, 2015, 41(3): 258-261.

Facial Expression Radical Generation Based on Example

GAO Yali, TAN Guanghua, GUO Songrui, FAN Xiaowei

(College of Information Science and Engineering, Hunan University, Changsha 410082, China)

【Abstract】 To solve the problem of blendshapes generation mostly depending on manual modeling and fine-tuning, this paper proposes an automatic generation method about example-based expression blendshapes. Using a set of generic face models as the prior, it makes each generated blendshapes approach to its real semantics. Expression blendshapes generated by this method can map semantics and expression dynamics from the generic model to the target model. The method achieves model's scalability in application and enables users to gradually calculate more expression models according to their own needs. Experimental results show that the method, constructing a gradient space to optimize blendshapes, outperforms the state-of-the-art methods both in speed and reality.

【Key words】 blendshape expression radical; example-based; iterative optimization; gradient space; reality; facial animation

DOI: 10.3969/j.issn.1000-3428.2015.03.048

1 概述

人脸模型参数化是特效电影和计算机游戏中构建虚拟角色的一项重要工作。基于 blendshape 的动画技术能够逼真地创建出各种虚拟角色的表情动画。由于其直观的控制方式, 深受动画师们喜爱。然而, 为了追求高度的真实感, 动画师即使已经最大限度地控制了一个特定面部的各种细微变化, 也还需要调整数百个 blendshape 表情基。例如, 电影《指环王》中的 Gollum 模型总共用了 946 个 blendshapes^[1]。因此, 需要动画师既要高效地建立一组个

性化的面部表情基又能够使其自动调整以匹配表演者的面部表情。

文献[2-5]使用多线性的主成份分析(Principal Component Analysis, PCA)模型, 提出了面部表情基的自动创建方法, 然而最终的线性表情基对于表情动画的控制并没有直观上的意义, 导致动画师无法直接使用。文献[6]提出一种基于非线性联合学习的三维人脸表情合成方法, 该方法的表情重定向合成结果较好, 但通过系数插值合成表情时可能产生由于过度变形导致的扭曲表情。文献[7]提出了一种通用的面部表情模拟技术, 该方法能有效地对 3D

基金项目: 国家科技支撑计划基金资助项目“文化旅游资源挖掘与体验式平台研发与示范”(2014BAK08B00, 2014BAK08B01)。

作者简介: 高娅莉(1988-), 女, 硕士研究生, 主研方向: 计算机动画, 数字图像处理; 谭光华, 助教、博士; 郭松睿, 博士研究生; 范晓伟, 本科生。

收稿日期: 2014-04-02 **修回日期:** 2014-05-09 **E-mail:** gaoyali2007@163.com

模型或 2D 图像进行表情转移,但该方法计算效率不高且真实感不够好。

本文提出一种基于样例的 blendshape 表情基自动生成方法。该方法利用一组通用的面部模型作为先验模型,与新创建的目标模型作为输入,将模板模型的表情动态映射到目标网格上^[8],最终采用迭代优化算法求解最优 blendshape 表情基及对应的混合权重。该方法摒弃了每一个 blendshape 必须与一个精确表情对应的限制,而只需要一组数量较少的样例模型及满足粗略初始猜想的混合权重即可。

2 基于样例的 blendshape 表情基自动生成

2.1 问题描述

基于样例的表情基生成问题可描述为:给定已知样例人脸模型 S 和 S' ,如何根据目标人脸模型 T 生成 T' 。

假设给定一组通用的 blendshape 模型的集合 $S = \{S_0, S_1, \dots, S_n\}$,其中, S_0 表示自然表情,表情基 $S_i (i > 0)$ 表示自然表情在局部范围内的变形。则任意的表情可以表述为 S_i 的线性组合,即 $E_j = S_0 + \sum_{i=1}^n x_{ij} S_i$ ^[9]。其中, x_{ij} 表示对应表情基 S_i 的混合权重。

创建新的目标模型(其中包括自然表情和 m 个其他表情)作为输入。根据非刚性匹配算法^[10]将通用模型的自然表情 S_0 与目标模型对齐,产生一组网格集合 $T = \{T_1, T_2, \dots, T_m\}$ 作为训练表情。本文的目的是计算一组新的 blendshape 表情基 $B = \{B_0, B_1, \dots, B_n\}$,使之能够最大程度匹配表演者的面部外观特征和运动特征。因此,需要求解出目标 blendshape 表情基 B_i 以及对应的权重 x_{ij} ,使得表情基的线性组合与训练表情的误差最小,即:

$$E_{\text{fit}} = \min \sum_{j=1}^m \|T_j - (B_0 + \sum_{i=1}^n x_{ij} B_i)\|_F^2 \quad (1)$$

该问题是一个双线性问题,它的求解存在 2 个主要问题:(1)在训练表情很少($m < n$)的情况下,如何计算目标表情基 B_i ;(2)在相同权重设置下,如何确保模板模型和目标模型之间的表情语义距离最小。

为了解决上述问题,本文采用了如下迭代优化方法交替迭代进行求解,其步骤如下:(1)保持混合权重 x_{ij} 不变,优化 blendshape 表情基;(2)保持 blendshape 表情基不变,优化混合权重。为了提高计算速度,用户需选择适当的混合权重合成一个粗略的表情 E_j ,使其与训练表情 T_j 大致对应,从而产生逼近的权重 \tilde{x}_{ij} ,为步骤(1)提供初始值,为步骤(2)提供表情的语义约束。开始训练表情基时,只需设定几个足够表现力的表情的混合权重即可,通常不超过 4 个,即 $m \leq 4$ 。

2.2 目标 blendshape 模型的重构

在明确问题之后,首先需要将形式化的问题与实际网格模型建立起联系,即:为了实现模板模型的表情能映射到目标模型上,需要对各个模型的网格进行操作,假设源模型和目标模型都是由三角网格构成,则可以针对网格中任意一个由顶点 v_1, v_2, v_3 构成的三角形 r ,定义一个 3×3 的矩阵 $A_r = [v_3 - v_1, v_2 - v_1, n]$, $n = (v_3 - v_1) \times (v_2 - v_1)$ 为三角形的法向量。这里将源三角形 r 到目标三角形 d 的变形梯度定义为 $H_{r \rightarrow d} = A_d \cdot A_r^{-1}$ 。其次,在梯度空间建立约束形成优化问题,求解这个优化问题便能重构出最终的 blendshape 表情基。

由于后续的优化是针对网格模型中各个三角形进行,因此忽略三角形的索引,网格 B_i 中每个三角形记为矩阵 A_i^B 中的每个元素。由自然表情 B_0 和训练表情 T_j ,可获得矩阵 A_0^B 和 A_j^T 。为了重构训练表情,定义如下拟合公式:

$$E_{\text{fit}} = \min \sum_{j=1}^m \|A_j^T - (A_0^B + \sum_{i=1}^n x_{ij} A_i^B)\|_F^2 \quad (2)$$

假定目标表情基 B_i 和模板表情基 S_i 的形变梯度近似,即 $H_{B_0 \rightarrow B_0 + B_i} \approx H_{S_0 \rightarrow S_0 + S_i}$,其中, $H_{B_0 \rightarrow B_0 + B_i} = (A_0^B + A_i^B)(A_0^B)^{-1}$,由此可解决训练数据不足的问题。定义如下调整公式:

$$E_{\text{reg}} = \min \sum_{i=1}^n w_i \|\tilde{A}_i^S - A_i^B\|_F^2 \quad (3)$$

其中, $\tilde{A}_i^S = (A_0^S + A_i^S)(A_0^S)^{-1} \cdot A_0^B - A_0^B$ 。调整权重 w_i 确保了通用模板的基本表情语义,一旦模板网格发生形变,就要求重构的目标 blendshape 模型也产生同样的变化。实验结果表明,调整权重 $w_i = ((1 + \|A_i^S\|_2)/(\mu + \|A_i^S\|_2))^\sigma$, $\mu = 0.1$ 且 $\sigma \geq 1$ 时,便足以达到优化效果。实验中 $\sigma = 2$,采用其他值也可以得到相似的结果。使用调整权重约束顶点并非限制目标模型表情的范围,因为通过优化将改变其他 blendshape 表情基的权重,从而获得其他个性化的表情。

结合式(2)和式(3)得到:

$$E = E_{\text{fit}} + \alpha E_{\text{reg}} \quad (4)$$

其中, α 表示平衡因子。由于矩阵 A_i^B 的第 3 个列向量定义为向量叉积,使得误差 E 呈非线性。文献[11]证明,可以忽略法向量成分进行重构,因此只需求解线性部分,即矩阵 A_i^B 的第 1 列和第 2 列。得出 A_i^B 就能使用最小二乘优化方法^[12]重建每个 blendshape 模型的顶点位置。

2.3 优化混合权重

给定 blendshape 表情基集合 B ,就能求出最优的混合权重 x_{ij} ,使用最小二乘拟合算法重构训练表情 T_j 。定义如下公式求解 x_{ij} :

$$E' = \min \sum_{j=1}^m \sum_{i=1}^N \left\| v_l^{T_j} - (v_l^{B_0} + \sum_{i=1}^n x_{ij} v_l^{B_i}) \right\|^2 + \lambda \sum_{i=1}^n (x_{ij} - \tilde{x}_{ij})^2 \quad (5)$$

其中, $v_l^{T_j}$ 和 $v_l^{B_i}$ 分别是训练表情 T_j 和表情基 B_i 的顶点; N 表示顶点数; 参数 λ 为平衡因子。由于混合权重约束在 $[0, 1]$ 区间, 因此可使用有效集算法^[12]对其进行求解。此外, 在 Maya 等动画编辑软件中, 当操作或微调 blendshape 滑动条时, 动画师往往添加强制约束, 阻止 2 个 blendshape 同时形成同一个表情。例如, 微调嘴巴单元时, 其约束条件为不同时从左和右进行挤压。这可以表述为 2 个互斥的 blendshape 表情基 B_i 和 B_k 的一个非线性约束形式 $x_{ij}x_{kj}=0$ 。采用针对非线性最小二乘问题的二次优化方法^[12]更新该混合权重 x_{ij} 。

3 实验结果与分析

首先, 选用文献[8]中的面部网格作为模板模型 S 。其次, 创建目标模型作为训练表情集合 T 。模板模型中包括 10 种基本表情, 分别是自然表情、生气、哭泣、暴怒、露齿笑、嘲笑、愤怒、悲伤、微笑、惊讶。

每次设置中, 主要控制参数 α 和 λ , 便能调整输出最优的 blendshape 表情基。当 $\alpha \gg 1$ 时, 输出的 blendshape 模型接近于直接进行变形传输后产生的结果。在这种情况下, 即便 $\lambda = 0$, 实验中也观察不到明显的变化。当 α 接近于 0.1 时, 输出的 blendshape 有可能组合后逼近训练表情, 但其变化可能比 $\alpha \ll 0.1$ 时更敏感。所有的实验中, 仅使用了 10 次交替迭代, 其中第一次迭代时 $\alpha = 0.5, \lambda = 100$ 。最后一次迭代时权重逐渐减小到 $\alpha = 0.1, \lambda = 10$ 。

本文应用的优化方法对于初始混合权重 \tilde{x}_{ij} 的选择具有鲁棒性。随机增加一个 $[-x, x]$ 之间的值, 直到 $|x| = 0.5$ 时, 重构所有样例都没有产生任何明显差异。图 1 显示了拟合咧嘴表情的 blendshape 模型时, \tilde{x}_{ij} 增加随机变量 x 后对模型的影响。

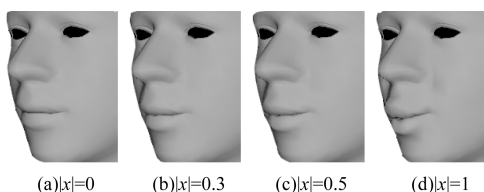


图 1 初始混合权重的变化对模型拟合效果的影响

图 2 显示了参数 α 对式(4)中 E 的影响。它有效地控制了通用模型和目标模型之间的映射。当其减小到 0.05 时出现过拟合现象, 在重构 blendshape 表情基时导致失真。实验发现 $\alpha = 0.1$ 时折中地满

足了精度和鲁棒性的要求。

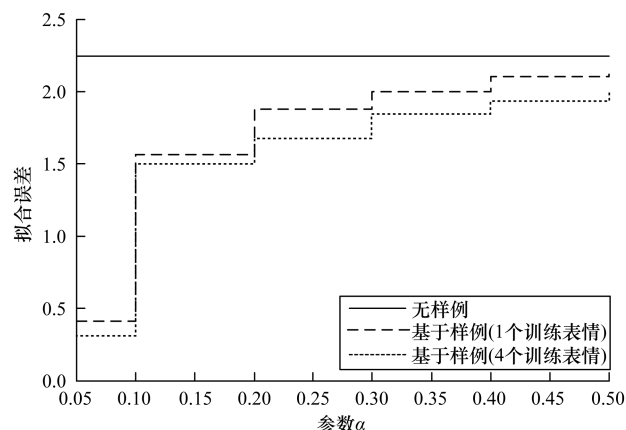


图 2 参数 α 对误差 E 的影响

本文应用的方法主要针对非标记面部表情动画。实验表明很少的训练表情(4 个)便足以精确地表达一个密集的表情空间。如果没有基于样例进行优化, blendshape 表情基模型将不具备足够的表现力。

本文与 3 种具有代表性的方法: 基于主成份分析(PCA)的方法^[5]、面部表情模拟(Facial Expression Analogy, FEA)方法^[6]和非线性协同学习(Non-linear Co-learning, NCL)方法^[7]进行比较。不同表情基生成方法对比如图 3 所示。从表 1 可以看出, 对于 4 个训练表情, 进行优化的实验大约需要 81 s, 约等于分别进行 blendshape 优化、重构和参数优化所用的时间总和。

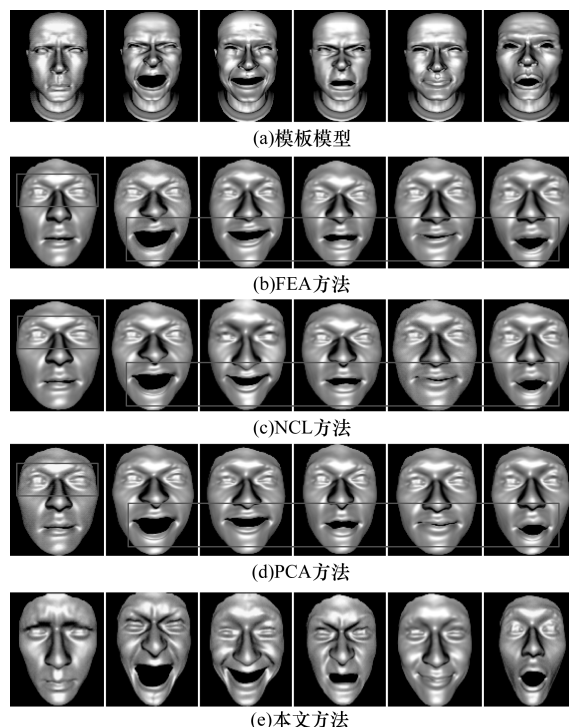


图 3 不同表情基生成方法对比

表 1 4 种方法消耗时间比较		s
方法	消耗时间	
FEA 方法	127.90	
NCL 方法	86.60	
在顶点空间基于 PCA 的方法	84.01	
在梯度空间基于样例的方法	81.37	

本文方法假定训练样例对应有效的表情 blendshape 组合。例如,使用先验模型无法表达鼓嘴吹气的表情。然而,通过验证 E 是否超过某一阈值就可以检测出模型中的表情是否丢失。因此,需增加语义不同的表情补充表情基空间。输入其他模型产生的效果如图 4 所示。目前,算法的交互速率还不够快。算法的复杂度为 $O(a(m+N))$,非线性求解部分的复杂度为 $O(n^3)$ 。

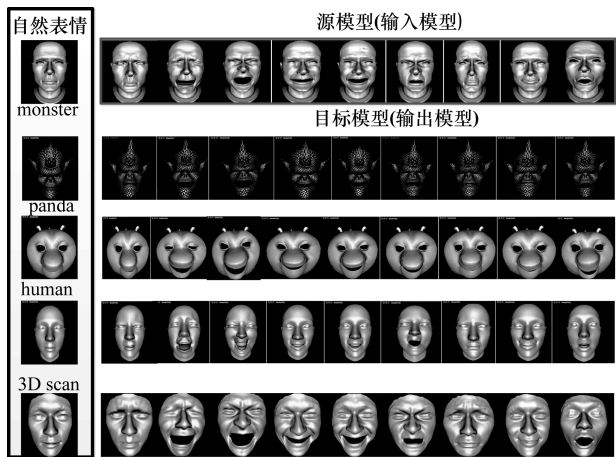


图 4 输入其他模型产生的效果

4 结束语

本文从自动生成个性化的 blendshape 表情基入手,根据样例表情迭代地优化表情基和混合权重,从而达到提高动画师工作效率,甚至让非专业用户也能很快地生成个性化的 blendshape 模型的目的。相较于单纯依赖手工建模并微调每个 blendshape 模型的方式,本文方法更方便快捷且具有可扩展性。从实验结果可以看出,本文方法是可行的,真实感高,能够生成高质量、个性化的 blendshape 模型。下一步工作将尝试如何使用较少的样例实现与更多的表

情基自动建立对应关系,同时,进一步提高表情映射的精度。

参考文献

[1] Lewis J P, Mooser J, Deng Zhigang, et al. Reducing Blendshape Interference by Selected Motion Attenuation[C]//Proceedings of Symposium on Interactive 3D Graphics and Games. Washington D. C. ,USA:[s. n.], 2005:25-29.

[2] 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, USA: ACM Press, 1999: 187-194.

[3] Blanz V, Basso C, Poggio T, et al. Reanimating Faces in Images and Video[J]. Computer Graphics Forum, 2003, 22(3):641-650.

[4] Vlasic D, Brand M, Pfister H, et al. Face Transfer with Multilinear Models[J]. ACM Transactions on Graphics, 2005, 24(3):426-433.

[5] Weise T, Gool L V, Venue M P, et al. Face/off: Live Facial Puppetry [C]//Proceedings of Symposium on Computer Animation. [S. l.]:ACM Press, 2009:7-16.

[6] Song Mingli, Zhao Dong, Theobalt C, et al. A Generic Framework for Efficient 2-D and 3-D Facial Expression Analogy[J]. IEEE Transactions on Multimedia, 2007, 9(7):1384-1395.

[7] 黄晓钦,林裕旭,宋明黎,等. 非线性联合学习的三维人脸表情合成方法[J]. 计算机辅助设计与图形学学报, 2011, 23(2):363-370.

[8] Sumner R W, Popović J. Deformation Transfer for Triangle Meshes [J]. ACM Transactions on Graphics, 2004, 23(3):399-405.

[9] Liu Xuecheng, Mao Tianlu, Xia Shihong, et al. Facial Animation by Optimized Blendshapes from Motion Capture Data [J]. Computer Animation and Virtual Worlds, 2008, 19(3-4):235-245.

[10] Coleman T F, Li Yuying. An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds[J]. SIAM Journal on Optimization, 1996, 6(2):418-445.

[11] Li Hao, Adams B, Guibas L J, et al. Robust Single-view Geometry and Motion Reconstruction [J]. ACM Transactions on Graphics, 2009, 28(5).

[12] 马昌凤. 最优化方法及其 Matlab 程序设计[M]. 北京: 科学出版社, 2010:166-178.

编辑 顾逸斐