

基于机器学习的三维骨架模型动作合成

黄元忠 宋 卫 卢庆华

(深圳市方直科技股份有限公司 广东 深圳 518000)

摘 要 骨架驱动是绘制三维动画的一种新的流行方式。通过动点采集、实时捕捉等方法获得的高精度三维数据可以凭借带有骨架的三维模型制作出逼真度高、细节精致、动作生动的三维动画。传统的三维模型素材在建模结构、建模样式等诸多方面存在着差异,这些差异使得骨架驱动模式在不同模型上使用时会产生不同程度的误差。文章提出一种机器学习的算法,用于解决骨架驱动模式在不同模型上的适应性,提高其针对不同三维模型驱动的兼容性,从而降低三维模型驱动的动作误差,提升动作质量,优化三维动画制作的效果。

关键词 机器学习;神经网络;骨骼动画

中图分类号:TP391.41;TP183 文献标识码:A

DOI:10.19414/j.cnki.1005-1228.2020.04.005

Skeleton-based Animation Synthesis using Machine Learning Framework

HUANG Yuan-zhong, SONG Wei, LU Qing-hua

(Shenzhen FangZhi Science & Technology Co., Ltd., Shenzhen 518000, China)

Abstract Nowadays, skeleton animation is a new approach to create 3D animation in main domains. The technic of skeleton animation is able to produce the animation with better authenticity, texture details and vivid movement based on dynamic joints tracking and real-time sensors capture. However, the effect of skeleton animation have obviously differences due to the traditional methods working on the different appearances, structures and meshes of the 3D models. Thus, we provide a novel algorithm based on machine learning theory to improve the practicality, adaptability and compatibility of the skeleton animation approaches, increasing the 3D objects movement quality and producing better 3D animations.

Key words: machine learning, neural network, skeleton-based animation

三维动画制作是计算机图形学应用的一个重要领域,在动画制作、电影工业和游戏领域中是一个十分重要的组成元素^[1]。骨架驱动模式是在静态三维模型的基础上,根据模型的结构和运动的区域,构造出基于模型的骨骼框架的一种技术方案。相比于传统的BlendShape模型动画制作技术,骨架驱动技术的优点在于制作简单,动作精度高以及在同类模型中的可迁移属性^[2]。然而,骨架驱动技术在实际使用时受到模型细节差异的影响,并不能完整的发挥其可迁移属性的优势。在很多动画制作过程中,动画师仍然需要根据新的模型定制出骨骼框架,并基于新的骨架重新设计和制作动画,这大大增加了三维动画制作的时间成本^[3]。

传统的骨架驱动动画是设计师通过模型骨架设计出该模型的运动轨迹,从而实现三维动画的绘制。在同类骨架模型中,骨骼架构在节点数量、连接点个数以及

连接结构上是一致的。然而,由于三维模型在外型的细节部分存在外形上的差异,导致一套模型动作在迁移到另一个模型上时。同样的动作参数不能在完成度和动作效果上与原有模型保持一致,使骨架动画在实际迁移的过程中产生了误差。传统的自适应算法是通过线性(Linear)策略去消除模型之间的个性化差异,然而实际的模型动画迁移的问题中,存在大量非线性(Nonlinear)的问题,传统的线性自适应算法并不能完全解决这个问题。机器学习是人工智能中的一个重要分支,并已被大范围应用并成功解决了很多实际案例中的回归、分类和预测等问题。其中,神经网络作为热门的一种技术被大量使用在解决非线性问题上。在图像分类问题中,神经网络也曾被用于基于人体骨骼的行为识别^[4]。因此,我们提出一种基于BP神经网络的模型,用于解决三维骨架动画的迁移问题。

2 构造基于骨架动作迁移的学习模型

骨架是三维骨架动画技术的核心部分,是解决模型运动的主要结构^[5]。骨架模型包括:关节点、关节连接线、骨架依附关系、IK 控制器等部分。在骨架动画制作中,人物面部骨架是结构最复杂、精度控制要求最高、动作绘制最难的部分^[6]。如图 1 所示,该骨架示意图是一个常用标准的面部骨架结构,图中包含双眼球中心、下巴中心点、嘴唇关键点、面部肌肉关键点等骨架关节点组成。通过蒙皮算法(Linear Blending Skinning)、方向性约束和权重分配,三维模型的 mesh 网格会根据骨骼的运动而产生联动效果,从而形成整体三维模型的运动^[7]。

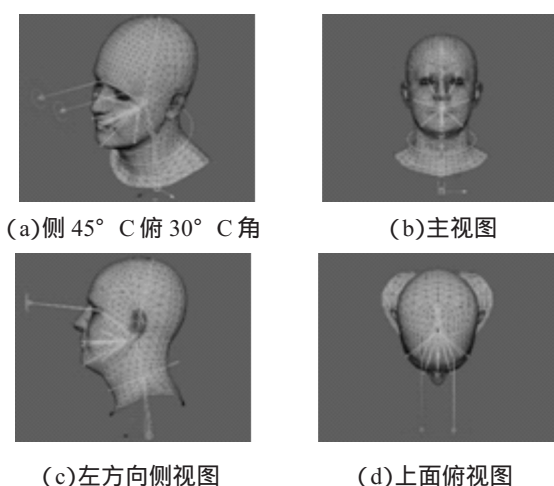


图 1 人体头部模型与面部骨骼结构示意图

三维模型骨架是根据模型本身的外形所构建的,同样一类三维模型在骨架的结构和设计上也保持着相同的规格^[8]。然而,由于同一类三维模型在建模轮廓、纹理细节上的差异,依据模型所构建的骨架之间也存在着些许不同。因此,在同一类三维模型的骨架动作迁移上,存在这因为骨骼结构差异导致的运动参数传递的误差^[9]。为了解决这一问题,我们构造了一个基于人脸头部三维模型的骨骼,如图 2 所示。该骨架包括 28 个骨骼关节,覆盖了三维人脸模型的大部分区域,我们将设计基于该骨骼的面部动作迁移系统。用来解决通过已知模型结构和运动向量来驱动目标模型的问题^{[10][11][12]}。

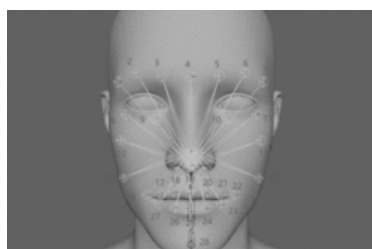


图 2 基于人脸三维模型构架的可驱动骨骼

我们通过构造 BP 神经网络用于解决源模型到目标模型的运动向量的迁移问题。BP 神经网络由多层节点组成,包含一个输入层、一个输出层和一个或多个隐层,如图 3 所示。其中,输入向量由 X_{dN}^T 和 V_N 两部分组成,其中 X_{dN}^T 表示在单位时间 T 内的所有骨骼的运动向量, V_N 表示目标模型的骨骼结构。对于任意时间点 t 的骨骼的运动向量定义为:

$$X_{dN}^t = \{X_{d1}^t, X_{d2}^t, \dots, X_{dN}^t\} \quad (1)$$

其中,任意时间点 $t \in T$, N 表示骨骼关节点的数量,在这里为 28,对应图 2 的骨骼模型关节点的数量。对于任意一个骨骼关节点 n ,在第 t 时间点上相对于初始模型的运动向量定义为:

$$X_{dN}^t = \{dx_n^t, dy_n^t, dz_n^t\} \quad (2)$$

其中 dx_n^t, dy_n^t, dz_n^t 分别表示骨骼关节点相对于初始状态下在 x, y, z 三个坐标轴方向上的偏移量。

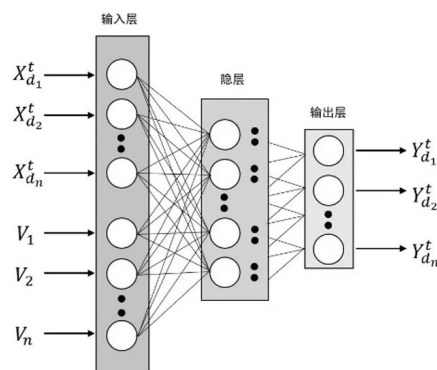


图 3 用于三维模型运动向量迁移的 BP 神经网络

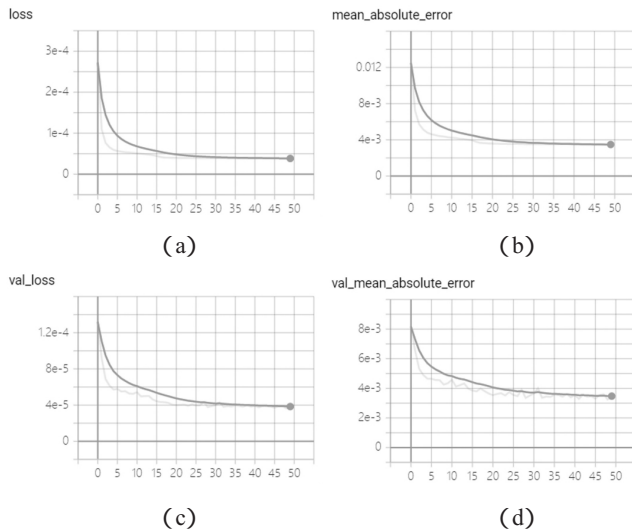
输入向量 V_N 表示目标模型,其中 N 同样表示骨骼关节点的数量,在单位时间 T 内,输入向量 V_N 不变,只有源模型运动向量 X_{dN}^t, X_{dN}^t 会发生改变。

3 试验仿真

我们通过 Maya 软件中内嵌 python 脚本,模拟产生出接近真实的三维人脸模型和相关的三维动作数据,并通过 blandshape 对齐面部纹理和模型骨架。我们基于非线性仿真的数据,构建出深度学习模型,分析深度学习模型在解决三维动作数据迁移问题上的表现。我们模拟出了 20 多个人脸骨架模型和超过 2 个小时的三维动作数据。通过预处理和随机选取,我们将骨架模型参数和动作数据转化为一个长度为 34000 的数据向量,并通过归一化将数值转化到 $[0,1]$ 区间,提高神经网络的学习效率。

我们采用了基于 Keras 的 Tensorflow 框架来实现

本实验中需要构建的深度学习模型。其中 relu 作为主要的激励函数用于输入层和所有隐层,训练过程采用 MSE 作为主要的梯度下降算法,模型优化器采用 Adam 动态优化学习率 learning rate 以及相关模型参数。模型初始学习率为 0.001,且所有神经网络权值使用均匀分布初始化器初始化。



(a)训练 MSE 值变化图 (b)训练 MAE 值变化图;
(c)预测 MSE 值变化图 (d)预测 MAE 值变化图;

图 4 深度神经网络训练和预测

通过深度学习模型的学习和相关参数的变化,如图 4 所示,其中 MSE 和 MAE 的变化呈现出稳定且平滑的下降趋势。并且训练和测试数据的下降效果均呈表现稳定,模型在学习过程中收敛效果明显,深度神经网络对于非线性数据的学习效果理想。

4 结束语

我们基于三维动画数据迁移问题构建了深度学习模型,用于解决数据迁移问题中的大量非线性问题。从模型的学习效果上来看,深度学习模型对存在于三维空间的非线性预测取得了良好的效果,证明了其可以用于解决三维动画数据迁移问题的能力。

参考文献:

[1] 翟浩澎. MAYA 三维动画制作技术[J]. 艺术研究. 2017(03):

42-43.

- [2] Shin Yoshizawa, Alexander Belyaev and Hans Peter Seidel. Skeleton based Variational Mesh Deformations [J]. Eurographics 2007, Volume 26 (2007), No. 3.
- [3] J. P. Lewis, K. Anjyo, T. Rhee, M. Zhang, F. Pighin, and Z. Deng. Practice and Theory of Blendshape Facial Models [J]. Eurographics 2014, 2014.
- [4] Yong Du, Wei Wang, Liang Wang. Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition [C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1110-1118.
- [5] M. Raptis, D. Kirovski, and H. Hoppe, Real-time classification of dance gestures from skeleton animation [C]. in Proc. ACM SIGGRAPH/Eurographics Symp. Comput. Animation, 2011, pp. 147-156.
- [6] G. Aujay, G. Hetro', F. Lazarus and C. Depraz. Harmonic skeleton for realistic character animation [C]. Symposium on Computer Animation, pp. 151-160, 2007.
- [7] Ladislav Kavan, Steven M Collins, Jiri Zara, et al. Geometric skinning with approximate dual quaternion blending [J]. ACM Transactions on Graphics, 27(4),2008.
- [8] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly. Realtime performance-based facial animation [C]. ACM SIGGRAPH 2011, pp 77:1-77:10, New York, NY, USA, 2011. ACM.
- [9] D. Bradley, W. Heidrich, T. Popa and A. Sheffer. High resolution passive facial performance capture [J]. ACM Trans. Graph. 29, 41:1-41:10, 2010.
- [10] H. Li, J. Yu, Y. Ye, and C. Bregler. Realtime facial animation with on-the-fly correctives [J]. ACM TOG, 32(4):42, 2013.
- [11] K. Li, F. Xu, J. Wang, Q. Dai, and Y. Liu. A data-driven approach for facial expression synthesis in video [C]. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 57-64, 2012.
- [12] F. Shi, H.-T. Wu, X. Tong, and J. Chai. Automatic acquisition of high-fidelity facial performances using monocular videos [J]. ACM TOG, 33(6):222, 2014.