

硕 士 研 究 生 读 书 报 告



题目 用于角色运动合成和编辑的深度学习框架

作者姓名 修玉同

作者学号 21651099

指导教师 李启雷

学科专业 移动互联网与游戏开发技术

所在学院 软件学院

提交日期 二○一七 年 四 月

A Deep Learning Framework for Character Motion Synthesis and Editing.

A Curriculum Report to

Zhejiang University

and for the Course

3D animation and interactive technology.

Major Subject: Software Engineering

Advisor: Qilei Li

By

Yutong Xiu

Zhejiang University, P.R. China

2017

摘要

我们提出了一个基于高级参数来合成角色运动的框架，使起所产生的运动分析人体运动的多个部分，并在大型运动图像数据集上训练。由卷积自动编码器的隐藏单元表示的学习运动流形表示可以组合以产生广泛复杂运动的稀疏分量中的运动数据。为了从高级参数映射到运动流形，我们在多次训练的自动编码器之上堆叠一个深的前馈神经网络。这个网络经过训练，可以从诸如角色应该遵循的地形上的曲线等参数或打孔和踢球的目标位置产生逼真的运动序列。前馈控制网络和运动模块被独立训练，允许用户根据需求轻松地在前馈网络之间切换，而无需重新训练运动模块。一旦生成运动，可以通过在运动歧管的空间中进行优化和编辑。这允许施加运动约束，或改变运动的风格，同时确保编辑的动作保持自然。因此，该系统可以产生平滑，高质量的运动序列，而无需对训练数据进行手动预处理。

**关键词**：深层学习，卷积网络，自动编码器，人体运动，角色动画，多元学习

Abstract

We present a framework to synthesize character movements based on high level parameters, such that the produced movements respect the manifold of human motion, trained on a large motion capture dataset. The learned motion manifold, which is represented by the hidden units of a convolutional autoencoder, represents motion data in sparse components which can be combined to produce a wide range of complex movements. To map from high level parameters to the motion manifold, we stack a deep feedforward neural network on top of the trained autoencoder. This network is trained to produce realistic motion sequences from parameters such as a curve over the terrain that the character should follow, or a target location for punching and kicking. The feedforward control network and the motion manifold are trained independently, allowing the user to easily switch between feedforward networks according to the desired interface, without retraining the motion manifold. Once motion is generated it can be edited by performing optimization in the space of the motion manifold. This allows for imposing kinematic constraints, or transforming the style of the motion, while ensuring the edited motion remains natural. As a result, the system can produce smooth, high quality motion sequences without any manual preprocessing of the training data.

**Keywords:** deeplearning, convolutional, neuralnetworks, autoencoder, human motion, character animation, manifold learning

1 介绍

数据驱动的运动合成允许动画师从高层中产生角色运动。这样的方法大大有助于动画制作，因为动画师只需要提供高级别的指令，而不是通过关键帧提供底层细节。在计算机动画中已经提出了利用大动作捕获数据集和机器学习参数化运动的各种技术。

目前可用的大多数数据驱动方法需要大量的手动数据预处理，包括运动分割，对齐和标签。任何阶段的错误都可能导致最终动画失败。因此，这种预处理通常通过大量人为干预来仔细执行，确保输出运动显得平滑自然。这使得全自动化变得困难，因此这些系统通常需要专门的技术开发人员进行维护。

在本文中，我们提出了一种基于深度学习框架的动画合成和编辑模型，可以使用大量的人体运动数据自动学习运动数据在非线性流形中的嵌入，无需手动数据预处理或人为干预。我们在大型运动数据库上训练卷积自动编码器，使其可以再现作为输入给出的运动数据，以及通过插值合成新的运动。这种无监督的非线性歧管学习过程不需要任何运动分割或对准，这使得该过程比以前的方法显着更容易。在这个自动编码器的顶部，我们堆叠另一个前馈神经网络，将高级别参数映射到低级别人体运动，由自动编码器的隐藏单元表示。通过这种方式，用户可以从直观的输入中轻松地产生逼真的人体运动序列，例如某个地形上的曲线，角色应该如何，或者用于打孔和踢球的最终效果器的轨迹。随着前馈控制网络和运动控制器的独立训练，用户可以根据需要的接口轻松地交换和重新训练前馈网络。我们的方法本质上是并行的，这使得计算速度非常快，并且适合主流动画包。

我们还提出了在运动歧管的空间中编辑运动数据的技术。卷积编码器的隐藏单元以稀疏和连续的方式表示运动，使得在该空间中调整数据保持运动的自然度和平滑度，同时仍然允许复制身体的移动。这种编辑的一个示例性例子是通过最小化合成运动的隐藏单元和参考样式运动的Gram矩阵的差异来将一个运动与另一运动相结合。

总之，我们的贡献如下：

1. 一种用于从高级参数合成角色动画的快速、平行的深度学习框架。
2. 满足用户约束和改变运动风格的方法。

**2 相关工作**

**2.1 基于内核的运动混合方法**

基于内核的运动合成方法是通过混合运动捕捉数据合成运动的主要技术。我们接下来将回顾交互式角色控制的方法，其中使用用户指令来使用运动数据库来合成新的动作。最后，我们回顾深入学习的方法，以及它们如何应用于角色动画。诸如RBF和GP之类的基于内核的方法面临巨大的成本，可以混合的运动数量有限。我们的方法没有这样的限制，可以扩展到大量的培训数据。

**2.2交互式角色控制**

在交互式角色控制期间应用运动共同运动混合的许多方法需要对运动进行分类，分段和对齐，以产生每个运动类的丰富模型。来自不同时期或不同阶层的姿势不能混合。虽然Kovar和Gleicher 试图使这个过程自动化，但运动序列的运动和分割标准之间的距离度量的选择可以显着地影响性能和精度。明确地执行这些步骤的要求可能是其性能的瓶颈。相反，我们的无监督框架自动混合附近的运动来合成实际运动，而不需要进行运动分割和分类。

**2.3运动数据的深度学习**

基于深度学习的技术目前是图像和语音识别领域最先进的技术，基于深度学习的框架的一个重要优点是它们自动学习数据集的特征。例如，当将卷积神经网络应用于图像识别时，类似于Gabor滤波器的滤波器出现在底层，对于不同对象的更复杂的滤波器出现在较高层中。我们的主要方法是利用这样的角色动画的力量。

**3 系统总览**

系统的轮廓如图1所示，它使用来自大型人体运动数据库的数据，对卷积编码器进行了训练，并因此发现了一个通用运动歧管。在这次训练之后，运动可以由网络的隐藏单元来表示，我们通过堆叠在卷积自动编码器顶部的前馈神经网络给出这个表示，并在高级别控制参数和隐藏单元之间产生映射。这项工作中显示的高级别控制参数是地形上的人物轨迹以及末端执行器的移动。由于这些参数设置可能包含歧义，使用另外一个小型网络用于计算和消除歧义，输入参数只有与该任务相关的运动捕捉数据的子集用于训练这些前馈网络。使用该框架，用户可以通过在地形上绘制曲线来生成角色行走和运行的动画，或者用户可以通过指定手脚的轨迹来让角色踢打运动。一旦生成了运动，就可以在隐藏单元的空间中进行编辑，从而使所得的运动满足诸如脚踏滑冰等的位置限制的约束。使用这种技术，我们描述了一种使用短样式化剪辑作为参考来转换角色运动的风格的方法。

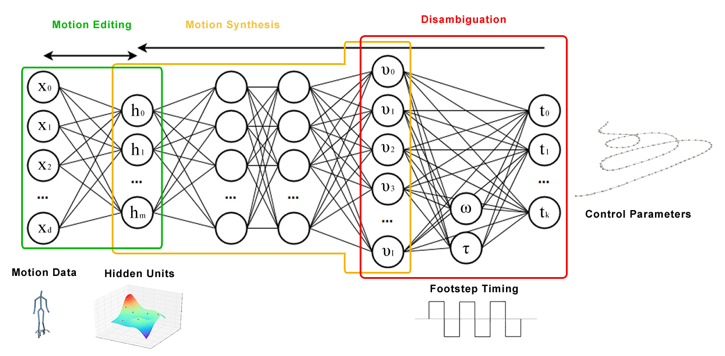


图1: 我们的方法的大纲。高级参数被消歧，用作前馈神经网络的输入。这些网络在卷积自动编码器的隐藏单元的空间中产生运动，可以进一步用于编辑生成的运动。

**4 数据采集**

**4.1 用于深度学习的的运动数据集合**

我们通过收集许多免费提供运动捕获数据的大型在线数据库，以及从内部捕获中添加数据，并将其重新定位到具有单个刻度和相同骨长度的统一骨架结构。 重定向是通过首先将源骨架结构中的任何对应关节角度复制到目标骨架结构，然后将源骨架与目标骨架尺寸相同，最后执行全身反向运动学方案移动目标骨架的关节以匹配源骨架中的任何对应的关节位置。这样的数据集一旦构建，最终的结果大约是CMU运动捕捉数据库大小的两倍，并且包含以每秒120帧采样的单个字符的大约600万帧的高质量运动捕捉数据。

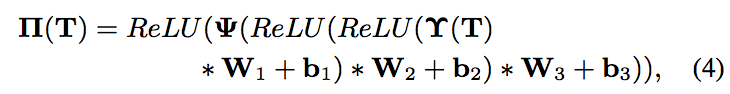
**4.2 训练数据的样式**

我们将数据集转换为适合培训的格式。我们将数据库中的所有运动都采样到每秒60帧，并将数据从原始数据集中的关节角度表示转换为3D关节位置格式。关节位置在身体的局部坐标系中定义，其原点位于根位置投影的地面上。使用左右肩和臀部的向量计算身体的前向（ZZ轴），对其进行平均并计算与垂直轴（Y轴）的交叉乘积。这使用高斯滤波器平滑以去除任何高频运动。 XZ平面中的全局速度以及每帧中垂直轴（Y轴）的身体的旋转速度被附加到输入表示。随着时间的推移，这些可以被整合以恢复运动的全局平移和旋转。通过检测脚趾或脚趾的脚趾何时低于一定的高度和速度来发现足部通信标签，并且也附加到输入表示。速度和足部通信标签也被他们自己的标准差消除。

**5 映射高级参数到人类运动监控**

**5.1 前馈网络的结构**

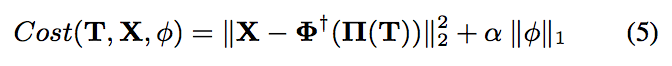
首先我们介绍前馈卷积网络，其将高级参数T映射到前面部分中构造的自适应编码网络的隐藏层，使得最终系统输出角色的运动。前馈卷积网络包含三个层次，另外还有一个额外的操作，这是解决模糊问题的任务特定操作。 前馈操作由以下给出：



其中 是馈送网络的两个隐藏层中的隐藏单元数，是三个卷积运算的的过滤器宽度， l是高级参数的自由度，分别为64, 128, 45, 25, 15和7。这些过滤器宽度确保每个运动帧都是使用大约1秒的轨迹信息生成的。 因此，用于回归的前馈网络的参数由= {W1，W2，W3，b1，b2，b3}给出。

**5.2 前馈网络的训练**

为了训练高水平参数和输出运动之间的回归，我们需要将随机下降的成本函数最小化，保持自动编码网络的参数固定。成本函数定义如下：



第一项计算回归的均方误差，第二项是稀疏项，以确保使用最小数量的隐藏单位来执行回归。 如前所述，设置为0.1。

培训这个网络时，我们只使用与该任务相关的数据。 例如，在运动任务期间，我们只使用特定于运动的数据。 因此，培训所需的时间远远少于自动编码器。 对于运动任务，对200个时期进行了训练，大约需要1个小时。

**5.3 运动歧义的消除**

在本节中，我们将描述我们的解决方案，以消除运动的歧义，给出在地形上绘制的曲线。单独的地形上的曲线没有给出足够的信息来充分描述由于上述模糊性问题而产生的运动。 我们检查了各种类型的投入，并发现，当脚与地面接触时提供时间可以大大地消除运动的歧视。 事实上，接触时间甚至可以区分步行和跑步，因为步行中总是有双重支持阶段，并且运行中有飞行阶段。 因此，我们训练一个可以用于从给定轨迹自动计算脚部触点的模型。 我们将其包含在前馈网络的输入中，以解决歧义。

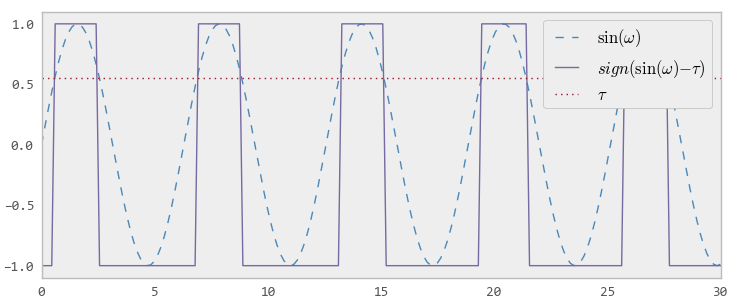


图2: 代表从正弦波计算的脚接触的方波，当值为1时，脚与地面接触。

该网络的输入是在XZ平面上的平移速度和围绕Y轴的旋转速度的轨迹，给定每个时间步长和相对于路径的向前方向，其中n 是轨迹中的帧数，k是轨迹输入的维度，其值为3。角色高度被认为是常数。输入被传递到方程(4)中的 ，它将足部通信信息加入到了轨迹输入：



其中是表示每个帧处的左脚跟，左脚趾，右脚跟和右脚趾的接触状态的矩阵，并且当与地面接触时其值为1，否则为-1。

**6 讨论与总结**

运动合成的许多其他方法是时间序列方法，但我们的运动合成方法是一种程序性的方法，因为它不需要逐步计算，任意时间的单个帧可以根据需要生成。这使它非常适合动画制作软件如Maya，允许动画师跳到时间轴中的任意点。动画师也不需要担心在本地约束窗口之外影响生成的动画的更改。这也使得系统高度并行化，因为可以独立地计算所有帧的运动。因此，可以在运行时间内以快速的速度在GPU上生成许多字符的轨迹。跨时间线并行生成运动需要帧之间的连续性。这由我们的框架以两种方式处理。通过生成脚部接触信息来提供高电平连续性（例如定时），而通过歧管确保低电平连续性（平滑度等）。

虽然程序方法在角色动画中并不新颖，但是以前的方法需要在将数据包含在模型中之前对运动进行分段，对齐和标记。相反，我们的模型会自动从大量运动数据中学习模型，而无需手动标记或分割。这使得系统非常实用，因为用户可以轻松地将新的运动数据添加到训练集中以丰富模型。

我们提出一个深入的学习框架，将高级参数映射到输出运动，首先使用大运动数据库学习运动流形，然后在用户输入到输出运动之间产生映射。 我们还提出了在同一框架下编辑和改变议案方式的方法。

目前，我们的自动编码器只有一层，因为深层堆叠自动编码器在脱模过程中会出现模糊。在我们的系统中，组合和抽象低级特征的作用由叠加在其上的前馈网络覆盖。然而，如果使用堆叠的深度自动编码器来学习运动歧管，则可以使用更简单的前馈网络，其更容易训练。

参考文献

[1] Daniel Holden, Jun Saito, Taku Komura, 2016. A Deep Learning Framework for Character Motion Synthesis and Editing.