

中图分类号: TP391.9 文献标志码: A 文章编号: 1006-8961(2011)08-1511-05

论文索引信息: 吕培, 张明敏, 徐明亮, 李灵, 宋青见, 潘志庚. 基于高斯过程动态模型的人体节奏运动合成[J]. 中国图象图形学报, 2011, 16(8): 1511-1515

基于高斯过程动态模型的人体节奏运动合成

吕培¹⁾, 张明敏¹⁾, 徐明亮¹⁾, 李灵¹⁾, 宋青见²⁾, 潘志庚¹⁾

¹⁾ (浙江大学计算机辅助设计与图形学国家重点实验室, 杭州 310058)

²⁾ (英特尔亚太研发有限公司, 上海 200241)

摘要: 提出一种新的基于高斯过程动态模型的节奏转移方法。该方法能够准确、有效地将现有运动中的节奏信息转移到新的运动中去, 适用于各种不同类型的运动。首先, 使用短时(short term)PCA 计算源运动的节奏点, 组合3种重要的运动特征求解目标运动的特征点; 然后, 使用动态规划算法来找到两者之间的最佳匹配, 最大化减少计算时间及对目标运动的修改; 最后, 使用高斯过程动态模型对目标运动进行学习, 并在隐空间进行节奏化插值, 最终合成新的节奏化运动。

关键词: 机器学习; 运动节奏; 运动合成; 插值

Rhythmical motion synthesis based on Gaussian process dynamical model

Lü Pei¹⁾, Zhang Mingmin¹⁾, Xu Mingliang¹⁾, Li Ling¹⁾, Song Qingjian²⁾, Pan Zhigeng¹⁾

¹⁾ (State Key Lab of CAD&CG Zhejiang University, Hangzhou 310058 China)

²⁾ (Intel Asia-Pacific Research & Development Ltd, Shanghai 200241 China)

Abstract: In this paper, we provide a new rhythm translation method based on Gaussian process dynamical model which can translate existing rhythm information to various types of motion accurately and effectively. Firstly, Short-Term PCA is used to calculate rhythm points in source rhythmical motion, three important motion features are combined to extract the motion feature points in target motion; Secondly, dynamical programming is utilized to find the best match between them in order to reduce the estimation time and damage to the target motion; Lastly, the Gaussian process dynamical model is applied to learn the target motion and rhythmical interpolation is done in latent space to get the new final rhythmical motion.

Keywords: machine learning; motion rhythm; motion synthesis; interpolation

0 引言

运动捕获技术的出现及运动数据重用技术的不断发展, 使人们不再满足于只是生成逼真的虚拟人运动, 而是希望能生成更多样化、更复杂的运动。但是, 由于运动捕获设备非常昂贵, 同时受限于捕获的外部条件, 能够捕捉到的可加以利用的运动数据非

常有限。如何充分利用这些有限的的数据, 尽可能多地生成符合人们要求的运动, 就成了摆在研究者及相关专业人员面前的一个巨大挑战。

音乐驱动人体动画的目标是给出一段音乐, 生成一段与音乐节奏相匹配的运动。节奏和运动特征提取是音乐驱动人体动画研究的基础和难点。Kim等人^[1]通过从未经标记过的样例运动中捕获运动节奏, 以此来提取这些运动中的基本运动和它们之

收稿日期: 2010-07-14; 修回日期: 2010-11-29

基金项目: 国家高技术研究发展计划(863)项目(2009AA062704); Intel - 高校合作项目。

第一作者简介: 吕培(1986—), 男, 浙江大学计算机辅助设计及图形学专业博士研究生, 主要研究方向为计算机动画, 虚拟人, 物理模拟。E-mail: lvpei@cad.zju.edu.cn。

间的转移关系,进而构造出相应的运动转移图。然而,这种方法的局限性在于必须事先知道运动的类型并且这种类型运动的节奏间隔要与相应的音乐类似。Lee 等人^[2]从运动和音乐中提取特征,然后使用动态规划算法在两者之间找到最佳匹配,通过对音乐和运动在时间上的修改来实现同步。Shiratori 等人^[3]基于人体运动易受到情感因素影响的事实,提出从舞蹈和音乐的情感角度寻找两者之间匹配的方法,也取得了很好的效果。Kim 等人^[4]从视觉感知的角度完成类似的工作。Xu 等人^[5]提出的短时主成分分析方法能够更加快速、高效地获取舞蹈运动的节奏信息。

Lawrence 等人^[6]提出的高斯过程隐变量模型表明,即使是一个非常小的样本空间仍然可以捕捉到人体运动的非线性特征。Grochow 等人^[7]提出的缩放高斯过程隐变量模型,在逆向运动学系统中学习人体姿势,并通过在不同的输入数据上训练模型以产生不同风格的逆向运动学(Inverse Kinematics, IK)。高斯过程动态模型通过学习低维隐空间和高维人体姿势空间的映射,能够准确地从单个实例中捕获人行走时的动力学特征^[8]。Wang 等人^[9]使用多因子高斯过程模型来区分人体运动的风格和内容。Lawrence 等人^[10]将高斯过程扩展到更为复杂的层次结构,进而应用于人体运动数据的可视化。Ikemoto 等人^[11]通过对运动学和动力学进行高斯过程建模,实现了对每一帧运动的预测。

国内的研究者也在虚拟人运动合成方面做了大量卓有成效的工作。王兆其等人^[12]提出的基本动作单元提取方法,充分考虑了运动序列间的时序性和不同关节之间的运动相关性,可用于合成新的逼真的运动。杨跃东等人^[13]针对已有运动分割方法的不足,提出一种可用于行为分割的运动串数据表示方法,使用该方法能够获得鲁棒的分割效果。王鑫等人^[14]提出一种为特定运动路径配上自然多样的人体运动数据的方法,能够有效处理待匹配运动路径运动范围较大、对多条运动路径同时计算匹配运动、多条运动路径之间交叉频繁的情景。

节奏化运动的生成,大都是由动画师依据音乐逐帧对运动进行调整得来的,整个过程非常繁琐且耗费时间。受到音乐驱动人体动画和风格化运动研究的启发,我们希望能自动完成这项工作。这项工作面临的主要难点:1)如何高效、准确地从源运动中提取节奏信息;2)如何在添加节奏信息的同时尽

量保留目标运动的原有内容。针对上述问题,我们提出基于高斯过程动态模型的节奏转移算法,本算法具有如下几个特点:

1)短时 PCA^[5](ST-PCA)应用于候选节奏点的提取。这种方法充分考虑运动数据在层次结构、空间和时间上的关系。使用短时 PCA,并结合文献[1]中的方法,能更加高效准确地处理非均匀间隔的节奏。

2)为了在计算代价和最终结果之间取得平衡,我们使用3个重要的运动特征来描述非节奏化运动。通过改变这3个运动特征在目标函数中所占的权重,可以获得不同运动类型之间细微的差别。

3)使用高斯过程动态模型对目标运动进行预测和插值。使用这样带有动力学约束的统计模型可以生成高质量的运动而无需进行后期处理,同时也可以根据误差估计来计算插值和预测的可靠性。

1 节奏和运动特征提取

人体的骨架模型可看作是由多个关节连接起来的骨骼构成的层次结构。第*i*帧运动表示为: $\mathbf{m}(i) = (\mathbf{p}_1(i), \mathbf{p}_2(i), \dots, \mathbf{p}_j(i))^T$, $\mathbf{p}(i) \in \mathbf{R}^3$ 表示每个关节的全局坐标, $j \in \mathbf{J}$ 表示关节索引。用于提取节奏信息的运动序列可以表示为: $\mathbf{S}_{\text{rhythm}} = (\mathbf{m}(1), \mathbf{m}(2), \dots, \mathbf{m}(M))$,用于提取运动特征并进行节奏转移的目标运动序列表示为: $\mathbf{T}_{\text{feature}} = (\mathbf{m}(1), \mathbf{m}(2), \dots, \mathbf{m}(N))$ 。数组*U*存放从运动序列 $\mathbf{S}_{\text{rhythm}}$ 中提取的节奏点对应帧的序号,数组*V*中存放从目标运动序列 $\mathbf{T}_{\text{feature}}$ 中提取特征点所对应的帧的序号。

1.1 节奏提取

样例运动 $\mathbf{S}_{\text{rhythm}}$ 按照固定大小窗口进行分割,每个窗口的大小为*W*(即窗口包含运动帧的数量),分割后的子窗口表示为 \mathbf{Win}_i , $1 \leq i \leq \lceil \text{length}(\mathbf{S}_{\text{rhythm}}) / W \rceil$ 。首先对第1个子窗口 \mathbf{Win}_1 进行PCA降维,得到它的第1个主成分分量对应的特征向量 \mathbf{B}_1 ,并将 \mathbf{Win}_1 在 \mathbf{B}_1 上进行投影,得到窗口 \mathbf{Win}_1 降维之后对应第1个主成分的系数 \mathbf{C}_1 ,它是一个1维向量。对第2个子窗口 \mathbf{Win}_2 执行同样的操作,得到 $\mathbf{B}_2, \mathbf{C}_2$ 。如果 $\mathbf{B}_1 \cdot \mathbf{B}_2 < 0$,说明前后两个窗口的第1个主成分分量方向相反,为了保持方向的一致性,需要对 $\mathbf{B}_2, \mathbf{C}_2$ 进行取反操作,否则, $\mathbf{B}_2, \mathbf{C}_2$ 保持不变。由于人为地将运动序列分成

了一个个窗口,破坏了运动本身的连续性,因此需要对降维得到的1维系数向量执行连接操作,以保持它们之间的平滑性。后续的其他窗口都跟它前一个窗口执行类似于 Win_2 和 Win_1 之间的操作,最终将所有的 C_i 存放到数组 W 中,选取 W 中的极值点作为候选节奏点。然后,使用文献[1]中的方法来计算参考节奏点和最终的节奏点。

1.2 运动特征提取

使用3个重要的运动特征来描述每一帧运动,并引入评价函数来选取合适的特征点。第1个运动特征是 weight effort,表示运动强度,用于区分简单和复杂运动。

$$we_i = \sum_{j=1}^J \arccos\left(\frac{p_j(i+1) \cdot p_j(i)}{\|p_j(i+1)\| \|p_j(i)\|}\right) \quad (1)$$

第2个运动特征是 motion span,表示一帧运动中所有关节偏离运动路径中心的距离。

$$c_j = \frac{1}{N} \left(\sum_{i=1}^N p_j(i) \right) \quad (2)$$

$$ms_i = \sum_{j=1}^J \|p_j(i) - c_j\| \quad (3)$$

c_j 表示关节 j 在整个运动序列中的中心位置。Motion density 是我们选择的第3个运动特征,用来表示关节的运动密度。

$$md_i = \frac{\sum_{j=1}^J \|p_j(i+1) - p_j(i)\|}{ms_i} \quad (4)$$

对每一帧的3个运动特征进行归一化操作,分别得到 we'_i, ms'_i, md'_i 。构造如下的评估函数来判断这一帧是否为我们所选取的特征帧。 a_1, a_2, a_3 ($0 \leq a_i \leq 1, \sum_{i=1}^3 a_i = 1$) 是它们的权重:

$$y(i) = a_1 we'_i + a_2 ms'_i + a_3 md'_i \quad (5)$$

运动特征点为函数 $y(i)$ 的局部最小值点。

$$Feature(i) = \begin{cases} 1 & \text{若 } y(i) \text{ 是局部最小值点} \\ 0 & \text{其他} \end{cases} \quad (6)$$

V 由 $Feature(i) = 1$ 所对应帧的序号组成。

2 节奏和运动特征匹配

假定源运动中提取的运动节奏点比目标运动中提取的运动特征点多,保证有足够数量的节

奏点供相应的运动特征点去匹配。 u_i, v_j 分别对应节奏点和运动特征点, $d(u_i, v_j) = (u_i - v_j)^2$ 表示两者之间的距离。动态规划计算大小为 $M \times N$ ($M = size(U), N = size(V)$) 的匹配矩阵来得到最终结果。

3 运动节奏转移

高斯过程动态模型(GPDM)定义了隐空间 X 到观察空间 Y 的一个非线性映射以及隐空间上的一个动态模型。 y 表示某一帧运动的位姿, $y_i = m(i)$, $i \in [1, 2, \dots, N]$ 定义了一段运动路径。 y' 表示一帧运动在观察空间中的位置, $y'_i = y_i - \frac{1}{N} \sum_{i=1}^N m(i)$, $i \in [1, 2, \dots, N]$ 。观察空间 $Y = [y'_1, y'_2, \dots, y'_N]$ 的高斯密度可以通过多个高斯分布相乘得到,每个高斯分布对应一个观察向量。 X 表示运动数据所对应的 d 维隐空间,这里我们设置隐空间的维度为3。高斯过程动态模型的具体学习过程可以参见文献[8]。

在节奏运动合成中,高斯过程动态模型将作为运动的先验模型,用来计算和评估观察空间中新的运动。给定一个学习模型 $\Gamma = \{Y, X, \alpha, \beta, W\}$,可以求得新的运动序列 Y^* 和它对应隐空间路径 X^* 的概率分布。

$$p(Y^*, X^* | \Gamma) = p(Y^* | X^*, \Gamma) p(X^* | \Gamma) \quad (7)$$

对目标运动中某个开始于第 i 帧,结束于第 j 帧的子序列,如果它与节奏运动中间隔为 T 帧的节奏点匹配,那么我们就可以从相应的隐空间中获得这个子序列对应的隐空间坐标 $S = \{x_i, x_{i+1}, \dots, x_j\}$ 。固定 x_i, x_j ,使用马尔科夫链蒙特卡罗采样器从序列 S 中等间距采样出 $T-2$ 个值,构造新的隐空间坐标序列 $S' = \{x_i, x'_i, \dots, x'_{j-1}, x_j\}$ 。有了这些新的隐空间坐标,通过式(8),可以得到相应的观察空间中的值。将所有位于运动特征点之间的子序列连接起来,最终得到合成后的节奏化运动。

$$\mu_Y(x') = Y^T K_Y^{-1} k_Y(x') \quad (8)$$

4 实验结果及分析

实验使用的运动数据经过向下采样之后保持相同的时间间隔,均为30帧/s。所有的实验均是在一

台 Intel Core 2 Quad 2.4 G HZ,2GB 内存的台式机上运行。

4.1 节奏提取

使用 5 段数据来提取运动节奏,其中 3 段来自 CMU 的运动捕获数据,剩余的 2 段是我们自己的捕获数据。首先使用窗口大小为 60 帧的 ST-PCA 对节奏运动进行分析,这里选取的运动均为舞蹈,得到相应的候选节奏点。图 1 是其中一个示例。然后,使用 Kim 等人^[1]提出的方法对候选节奏点进行细化处理进而得到最终的结果。

实验表明,实际的人体运动节奏并不完全是等间隔的,总是围绕着某一基准节奏在一定范围内变化,基于此,我们使用节奏平均帧来度量所提取的节奏,见表 1。

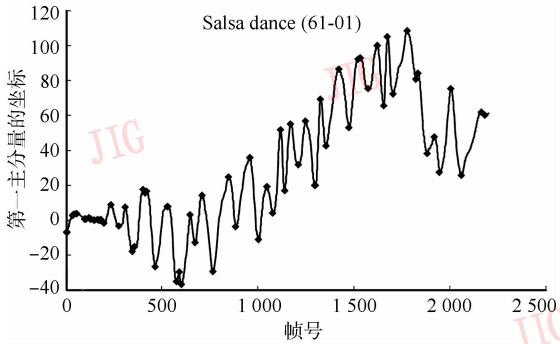


图 1 提取的候选节奏

Fig. 1 The extracted candidate rhythm points

表 1 节奏平均帧

Tab. 1 AFPR: average frames per rhythm

舞蹈数据	总帧数	真实值	节奏
Salsa(61_01,cmu)	2 241	7.5	7.6
Break(86_01,cmu)	998	12.4	12.1
Charleston(93_06,cmu)	402	7.4	7.9
Damage(our own)	501	15.0	15.3
Wanna(our own)	1 801	11.1	10.8

4.2 运动特征提取

用于节奏转移的运动来自 CMU 的运动捕获数据,它们分别是 walk (02_01), run/jog (35_18)。图 2 给出这两段运动所提取的特征点。实验证明选取的 3 个重要的运动特征可以涵盖绝大多数情形。在具体的应用中,例如,对简单的走路运动,weight effort 的权重大些,motion span 和 motion density 所占的权重相对较小,而要更精确的描述慢跑,则需要更

多的 motion density。在描述舞蹈运动时,相对于其他两个运动特征,weight effort 可以描述更多的情形,因此它在式(5)中的权重几乎总是大于 0.5,而 motion span 和 motion density 则可以更好地描述手部或腿部运动细节。

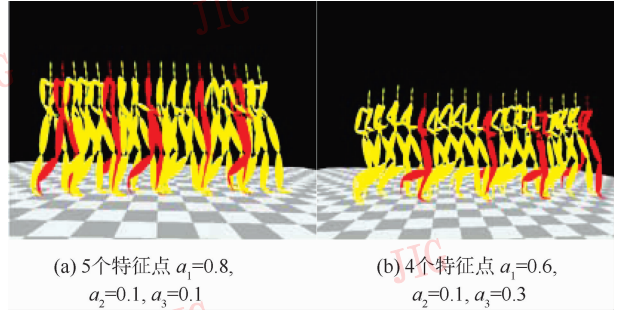


图 2 使用不同权重提取不同运动中的特征点
Fig. 2 Feature points extracted from different

4.3 GPDM 插值和节奏转移

将舞蹈中提取的运动节奏转移到走路运动中(图 2(a))。图 3(a)对应于走路运动一个周期的隐空间,数据来源为 02_01(CMU 运动数据库)15~150 帧。计算整个运动的隐空间坐标时,初始参数值设为 $\alpha = (0.9, 1.0, 0.1, e)$, $\beta = (1.0, 1.0, e)$, $W = (1, \dots, 1)$ 。MAP 算法的迭代次数设为 10,训练整个数据集所需时间约为两个小时。图 3(b)表示两个运动特征之间的隐空间坐标和它们对应观察空间中的位姿。

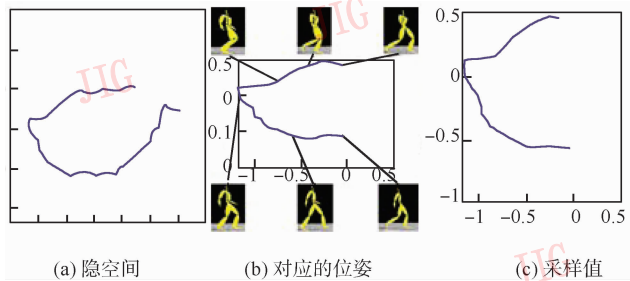


图 3 一段节奏化运动对应的采样值
Fig. 3 Sampled coordinates for one piece of rhythmical motion

对马尔科夫链初始化之后,使用马尔科夫链蒙特卡罗采样器对其进行采样,得到 T (两个相邻运动特征点对应的节奏间隔)个采样值。根据这些隐空间坐标,我们可以恢复观察空间中节奏化的运动数据,图 3(c)是我们选取的采样值。图 4(b)是最终的节奏化运动合成结果。

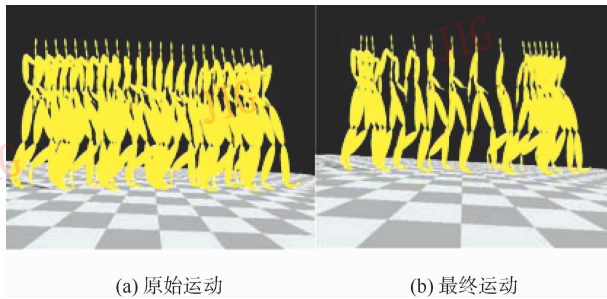


图4 两段运动对比

Fig.4 The comparison of two pieces of motion

5 结论

提出一种新的基于高斯过程动态模型的节奏转移算法。给定一段具有明显节奏信息的舞蹈运动,无须任何后期处理,该方法可以生成新的高质量的蕴涵源运动节奏信息的运动。使用充分考虑运动数据间时空关系的 ST-PCA 算法,有效提高运动节奏点提取的鲁棒性和准确性。不同的运动特征组合起来可以更好地描述目标运动所具备的特点。在第2节中我们假定源运动中的节奏点比目标运动中的特征点数量多,是为了提供更加丰富的节奏点供运动特征点去匹配,使得对目标运动的修改降到最小,减少后续插值操作所耗费的计算代价。高斯过程动态模型作为一种非线性的降维方法,能够有效地表示高维人体运动数据之间的非线性动态关系,对其隐空间进行插值操作,能生成更加自然、真实的运动。

参考文献 (References)

- [1] Kim T H, Park S I, Shin S Y. Rhythmic-motion synthesis based on motion-beat analysis [J]. *Acm Transactions on Graphics*, 2003, 22(3): 392-401.
- [2] Lee H C, Lee I K. Automatic synchronization of background music and motion in computer animation [J]. *Computer Graphics Forum*, 2005, 24(3): 353-362.
- [3] Shiratori T, Nakazawa A, Ikeuchi K. Dancing-to-music character animation [J]. *Computer Graphics Forum*, 2006, 25(3): 449-458.

- [4] Kim J W, Fouad H, Sibert J L, et al. Perceptually motivated automatic dance motion generation for music [J]. *Computer Animation and Virtual World*, 2009, 20(2-3): 375-384.
- [5] Xu J, Takagi K, Yoneyama A. Motion beat induction based on short-term principal component analysis [C]//*Proceeding of the 2009 ACM SIGGRAPH ASIA Sketches*. Yokohama. New York, USA:ACM, 2009: 1.
- [6] Lawrence N D. Gaussian process latent variable models for visualisation of high dimensional data [C]//*Proceedings of the 2009 Advances in Neural Information Processing Systems*. Cambridge, Canada: NIPS, 2004, 16: 329-336.
- [7] Grochow K, Martin S L, Hertzmann A, et al. Style-based inverse kinematics [J]. *Acm Transactions on Graphics*, 2004, 23(3): 522-531.
- [8] Wang J M, Fleet D J, Hertzmann A. Gaussian process dynamical models for human motion [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(2): 283-298.
- [9] Wang J M, Fleet D J, Hertzmann A. Multifactor Gaussian process models for style-content separation [C]// *Proceedings of the 24th International Conference on Machine Learning*. Corvallis, Oregon: ACM Press, 2007: 975-982.
- [10] Lawrence N D, Moore A J. Hierarchical Gaussian process latent variable models [C]//*Proceedings of the 24th International Conference on Machine Learning*. Corvallis, Oregon: ACM Press, 2007: 481-488.
- [11] Ikemoto L, Arikian O, Forsyth D. Generalizing motion edits with Gaussian processes [J]. *Acm Transactions on Graphics*, 2009, 28(1): 1-12.
- [12] Zhu Dengming, Wang Zhaoqi. Human animation synthesis based on primitive movement [J]. *Journal of Computer Research and Development*, 2009, 46(4): 610-617. [朱登明, 王兆其. 基于动作单元分析的人体动画合成方法研究 [J]. *计算机研究与发展*, 2009, 46(4): 610-617.]
- [13] Yang Yuedong, Wang Lili, Hao Aimin. Motion string: a motion capture data representation for behavior segmentation [J]. *Journal of Computer Research and Development*, 2008, 45(3): 527-534. [杨跃东, 王莉莉, 郝爱民. 运动串: 一种用于行为分割的运动捕获数据表示方法 [J]. *计算机研究与发展*, 2008, 45(3): 527-534]
- [14] Wang Xin, Sun Shouqian, Shao Ming. A path-driven character animation synthesis method [J]. *Journal of Computer-Aided Design & Computer Graphics*, 2009, 21(3): 319-324. [王鑫, 孙守迁, 邵明. 运动路径驱动的角色动画合成方法 [J]. *计算机辅助设计与图形学学报*, 2009, 21(3): 319-324.]