

基于 Kinect 的高维人体动画实时合成研究

王万良 马庆 王鑫* 汪晓妍

(浙江工业大学计算机科学与技术学院 浙江 杭州 310000)

摘 要 基于动作的视频交互游戏一直是游戏市场上非常受消费者青睐的主流游戏之一。研究利用 Kinect 从用户动作中获取低维控制信号,然后通过双级结构来重建高维度运动控制信号,以实现高质量人体动画的实时合成。其中第一级先通过构造一个邻居图缩小搜索空间,再通过 K-D 树加速搜索算法得到 k 个相似数据,最后基于主成分分析法来构建一个线性运动实时合成模型;第二级则是利用平滑参数对线性模型进行优化。实验结果表明,即使在场景受到严重干扰的情况下,该方法仍然可以重建出高质量的人体动画。

关键词 Kinect 人体运动合成 K-D 树 邻居图

中图分类号 TP391

文献标识码 A

RESEARCH ON REAL-TIME SYNTHESIS OF HIGH-DIMENSIONAL HUMAN BODY ANIMATION BASED ON KINECT

Wang Wanliang Ma Qing Wang Xin* Wang Xiaoyan

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310000, Zhejiang, China)

Abstract The interactive video game based on human action is always one of the most popular games on the game market. This research gets low-dimensional control signals from users' actions, then rebuilds high-dimensional motion control signals through two-level structure to realize quality real-time synthesis of human body animation. The first level is to get k similar data by constructing an adjacency graph for a smaller searching space using a K-D tree accelerating searching algorithm, until finally construct a real-time linear motion synthesis mode. The second level is to optimize the mode by smoothing parameters. Experimental results show that even when the scene is interfered seriously, the method can still rebuild quality human body animations.

Keywords Kinect Human motion synthesis K-D tree Adjacency graph

0 引 言

人体动画合成作为计算机图形学研究的热门内容之一,已经被广泛应用于基于视频交互的动作游戏中,有关视频交互游戏的开发也非常受到业界关注。包括 Microsoft、Nintendo 等知名厂商已经投入大量的研究经费来进行人体动画合成的研究和相关设备的开发。准确而高效的人体动画合成可应用于很多领域,如角色控制、虚拟现实和增强人机交互等。Microsoft 在 2010 年推出的 Kinect 体感游戏设备利用即时动作识别让玩家摆脱传统游戏手柄的束缚,通过自己的肢体来控制虚拟人。在此之前,电脑和视频游戏为控制人体运动提供了大量的接口,例如:鼠标、操作杆和按钮等等。但是,这些接口只是提供了少量的自由度上的控制并且必须自动计算运动细节。索尼的 EyeToy^[1]是一个基于视觉的系统,该系统不需要任何的标记,就能够提取出一个拳击或者挥手的简单 2D 动作。文献[2]建立了一个视觉的接口来把从单个视频摄像机上获得的噪音轮廓数据转换成约有 2 秒延迟的全身运动。文献[3]提出了基于视频跟踪来捕获用户运动的方法,但是这个方法并不能够实时并且精确地重建复杂人体运动。同时,文献[3]最早提出了利用 PCA 来减少人体运动维数的方法。磁性、机械和基于外骨骼的运动捕获系统

虽然能完成一个人体的捕获,但是这些技术往往存在以下缺点:设备昂贵;难以满足普通用户的需求;难以捕获复杂的人体运动。Kinect 基本上克服了上述的缺点,但是还无法呈现高维人体动画^[4]。大量的研究者已经研究出了利用运动捕获数据合成人体动画的技术^[5-7]。目前最常使用的方法有三种:构建人体运动模型,利用运动图重排运动^[2]和插值^[8]。文献[9]提出了基于低维信号来生成表演动画的方法。他们从附着少量反光标记的用户的动作中获取低维控制信号,通过在线局部建模和在线运动两个过程,重建出一个完整的人体动画。本文提出了利用 Kinect 捕获运动数据来获取用户动作的低维控制信号 c_i ,然后采用双级结构来合成人体动画,以实现从低维控制信号到高维控制信号的实时合成。其中第一级首先通过构造一个邻居图缩小搜索空间,再利用 K-D 树搜索算法搜索数据库得到 k 个相似数据,最后利用 PCA^[10]学习一个简单的线性局部模型,每一个新的局部模型都会被用于合成一个动作,第二级是利用平滑参数对线性模型进行优化。

收稿日期:2011-09-01。2011 中国计算机大会论文。浙江省自然科学基金(Y1110882);浙江省教育厅科研项目(Y200907765, Y201018160)。王万良,教授,主研领域:智能调度与控制,数字媒体。

1 系统框架

基于 Kinect 获得运动捕获数据合成人体实时动画相关的工作,主要由两部分构成:1) 运动捕获;2) 利用低维信号合成人体实时动画。系统的框图如图 1 所示。

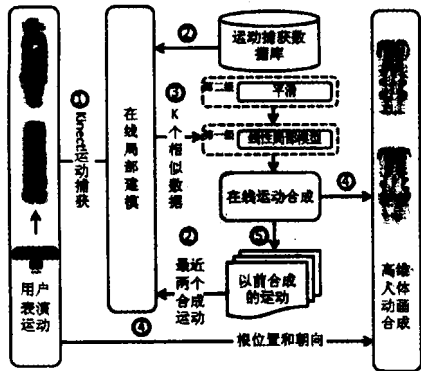


图 1 系统框图

1) 运动捕获

本文利用 Kinect 捕获运动数据来获得用户的低维控制信号。Kinect 的光学部分包括两个主要部件:红外线发射器和红外线/VGA 摄像头组。红外线发射器发出一道“激光”覆盖整个 Kinect 的可视范围,摄像头组接收反射光线来识别人。首先,通过红外摄像头获得图像的一个“深度场”。接着,Kinect 会对深度图像进行像素级评估,来辨别人体的不同部位,并采用分割策略来将人体从背景环境中区分出来,即从噪音中提取出有用信号。然后,一个人工智能系统(Exemplar 系统)会教会 Kinect 去辨识人体的各个部位。最后,Kinect 会根据追踪到的 20 个关节点来生成一幅骨架系统。Kinect 可以主动追踪最多两个玩家的全身骨架,或者被动追踪最多六名玩家的形体和位置。该过程如图 2^[11]所示。



图 2 Kinect 运动捕获过程

2) 利用低维信号合成人体实时动画

因为上述的运动捕获过程只是获得了用户的低维运动信号,所以合成高维的、准确的人体动画还需要经过以下几个步骤:① 根据标记位置和最近合成的两个运动在运动捕获数据库中进行局部运动选择;② 根据选择出的运动构建局部线性模型,并对模型进行平滑处理来实现在线运动合成;③ 根据根位置和朝向对在线合成的运动进行处理以合成新运动,使得能在动画中使用;④ 将合成的新运动连接至前面合成的运动上。

2 人体动画合成

2.1 数据处理

$[q_1, q_2, \dots, q_n]$ 表示存放在运动捕获数据库中的高维数据,为 87 维向量,这些数据包括走、跑、跳跃等等。 q_n 表示完整的人体运动数据,记录了一帧中人体动画中所有的关联角度。通过这些关联角度,计算出人体骨架点的三维空间位置 c_n 。 $[c_1, c_2$

$\dots, c_n]$ 表示数据库中的低维特征数据。 $[c_1, c_2, \dots, c_i]$ 表示通过 Kinect 获得的人体运动的低维特征数据。 c_n, c_i 均为 18 维向量,表示人体运动的特征。

为了提高计算效率,我们建立一个邻居图来描述不同姿势的关系。邻居图中的每个点表示运动捕获数据库中的一个高维姿势 q_n 。本文通过下式来描述两个姿势之间的相关性。如果第 i 个节点和第 j 个节点满足以下式子, a_{ij} 就定义为 1, 否则为 0。

$$\|q_i - q_j\|_{L_1} < \varepsilon \quad (1)$$

$\|\cdot\|_{L_1}$ 表示 L_1 的距离(曼哈顿距离)。 q_i 和 q_j 表示在运动捕获数据库的不同帧。 ε 表示在一个特定的搜索中,最近邻居的搜索范围半径,这用于判定两个姿势是否具有相关性。通过实验,我们取 ε 为 5。

在系统运行之前,就需要将邻居图加载到系统之中^[12]。当系统需要选择训练样本时,系统只需要搜索最近合成的两个人体运动,而不需要遍历整个数据库。这个方法不仅缩小了检索空间,提高了计算效率,而且使得运动合成更加有效。

2.2 搜索算法

c_i 和 c_n 位于同一个超平面内。当系统从 Kinect 中获得一帧 c_i , 它必须在数据库选择中选择接近 c_i 的 c_1, c_2, \dots, c_i 作为训练样本。本文中通过构造一个 K-D 树来提高搜索速度。

K-D 树,是一种由二叉搜索树推广而来的用于多维搜索的树形结构(K 即空间维数)。它的每一层通过检测不同的属性(关键字)值以决定选择分枝的方向。本文在 c_n 与 c_i 进行匹配时,我们采用 18-D 树来提高搜索速度。本文中构造的 18-D 树是一个二叉树,它的每一个节点有 18 维点。这个树状模型将父节点空间分成了两个邻接的空间,而且每个空间有一半的之前的节点^[13]。这两个空间的距离为欧氏距离,算法为:

$$|c_i - c_j| = \sqrt{\sum_{k=1}^{18} (c_{ik} - c_{jk})^2} \quad (2)$$

本文中的 18-D 树是利用迭代的方式构造的。在每一个内部节点,父节点的值是两个子节点的中间值,左节点小于根节点,根节点小于右节点。因此,当系统搜索一个点时,它的平均搜索时间复杂度为 $O(\log_2 n)$, 从而提高搜索效率。

2.3 在线局部模型

本文为当前姿势建立一个线性模型:

$$\hat{f}(q) = w_1 a_1(q) \vec{e}_1 + w_2 a_2(q) \vec{e}_2 + w_3 a_3(q) \vec{e}_3 + \dots + w_n a_n(q) \vec{e}_n \quad (3)$$

$a_i(q)$ 表示样本 q 的第 i 个属性。样本 q 表示通过搜索算法从运动捕获数据库中查询到与当前姿势最相似的高维样本数据。 \vec{e}_i 表示 i 维的单位向量。

然后,本文采用了一个梯度下降查找 w_1, w_2, \dots, w_n 的系数的方法,并利用这个方法来查找具有线性功能的训练样本。为了得到一个可以接受的权重向量,我们先选择一个随机权重,再将这个训练值反复应用到每个训练样本中。当这个训练值不适合训练样本时,改变这个值。这个过程是迭代的,直到这个训练值适合每一个训练样本。权重会随着每一步训练规则的变化而不断变化,并根据输入 q_i 修改权重 w_i 的值。

$$w_i \leftarrow w_i + \Delta w_i \quad (4)$$

$$\tilde{q}_i \leftarrow \hat{f}(q) \quad (5)$$

$$\Delta w_i = \eta \sum_{q \in k \text{ nearest nbors of } q_0} K(d(q_0, q)) (f(q) - \hat{f}(q)) a_i(q) \quad (6)$$

实例 q 的作用就是为了不断更新权重值。它乘以 $K(d(q_0, q))$ 后,运动合成的错误只会受 k 个相似数据的影响。 η 表示一

个称为学习率的常数,它决定梯度下降搜索的步长。

根据上面的模型,我们首先利用搜索算法选择训练样本 q_1, q_2, \dots, q_n 。每一个训练样本都是高维向量。这里需要注意的是 q 是在每个维度的适当区域内选择出来的随机数,而不是采用这个区域的平均值。这样做的目的是为了缩短计算时间。这种方法能够满足每秒合成几个实时动作的要求,并充分体现实时性。假设在训练样本中, q_i 与当前姿势最相似,系统会选择 $q_1, q_2, \dots, q_{i-1}, q_{i+1}, \dots, q_n$ 作为训练样本,并使用 q 来计算 w_i 。接着就可以使用线性模型来合成当前运动。

2.4 在线运动合成

本文使用局部线性模型作为一个强制性的约束,再利用平滑项优化处于低维空间 w_i 的当前动作 q_i 。

平滑项测量了如何将 q_i 排在 $[q_1, q_2, \dots, q_{i-1}]$ 之后,保证了合成的人体运动的连续性。我们假设在 t 时刻的动作只和 $t-1, t-2$ 时刻的动作有关,算法^[9]为:

$$E_{smoothness} = \|q_t - 2q_{t-1} + q_{t-2}\|^2 \quad (7)$$

其中 q_{t-1} 和 q_{t-2} 为前两帧生成的动作。

3 实验结果和分析

在线局部模型和以前的即时学习方法是不同的,因为前者是从最近实例的集合而不是从直接插值局部实例中构建一个低维空间来合成运动的。下图,我们将本文的方法和两个流行的局部学习方法进行了比较:近邻 NN(nearest neighbor)和局部加权回归 LWR(locally weighted regression)。近邻合成简单选择了最近的实例。局部加权回归通过它们到查询点的距离回归插入邻接点。图3展示本文的方法比近邻合成和局部加权回归更为准确。

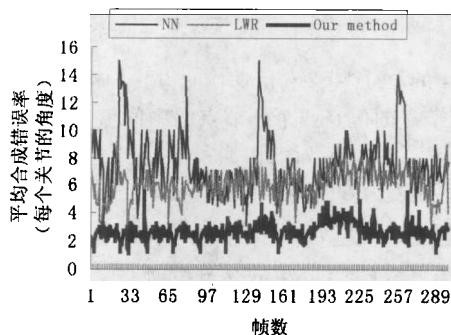


图3 行走的平均合成错误:对 NN 的每帧 7.56 度/关节,对 LWR 的每帧 6.0 度/关节,对本文的方法每帧 2.7 度/关节

为了验证算法的正确性和有效性,本文采用 Kinect 捕获设备采集行走、跑步和跳跃这三组视频。见图4、图5、图6。在实验中,我们的主要工作是让系统读入 Kinect 已经提取好的低维控制信号,然后将它还原成高维人体动画。

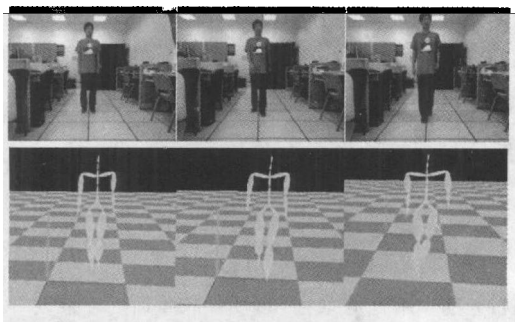


图4 Kinect 捕获行走姿势和相应的高维人体动画

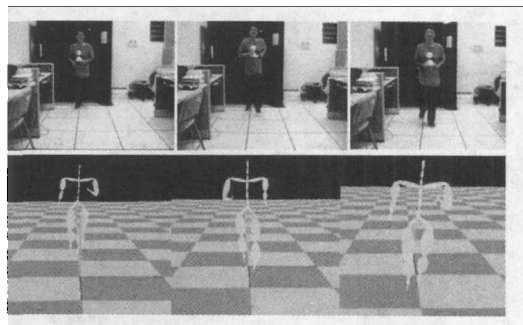


图5 Kinect 捕获跑步姿势和相应的高维人体动画

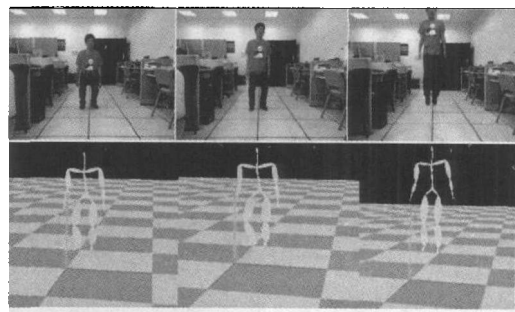


图6 Kinect 捕获跳跃姿势和相应的高维人体动画

通过实验表明,即使在场景受到严重干扰的情况下,本文方法仍然可以重建出高质量的人体动画,且系统性能良好,平均合成时间为每秒 30 帧,没有明显的延迟。

4 结 语

本文描述了如何利用 Kinect 获取人体运动的低维信号,并利用低维信号来重建高维人体动画的方法。本文在数据处理、搜索算法以及在线局部建模中使用了一些新的算法,使得人体动画合成更加准确。本文利用邻居图缩短数据处理的时间;利用 K-D 树缩短数据搜索的时间;利用局部模型来实现人体运动的实时合成。在运动合成的过程,巧妙地使用之前合成的人体运动来保证当前运动的流畅性。最后,利用实验充分验证了本文方法的有效性。

由于 Kinect 本身能跟踪两个人的运动,本文后续工作主要集中于如何利用 Kinect 来实现两个人之间的交互运动合成。另外,如何使用多个 Kinect,来完成一群人之间的交互式运动合成,也是下一步的工作重点之一。

参 考 文 献

- [1] Sony Eye Toy Systems. 2003. <http://www.eyetoy.com>.
- [2] Lee J, Chai J, Reitsma P, et al. Interactive control of avatars animated with human motion data[J]. ACM Transactions on Graphics, 2002, 21(3): 491-500.
- [3] Howe N, Leventon M, Freeman W. Bayesian reconstruction of 3d human motion from single-camera video[J]. Advances in Neural Information Processing Systems, 1999, 12: 820-826.
- [4] Blewett A, Eshar D, Kutliroff G, et al. Enhanced interactive gaming by blending full-body tracking and gesture animation[C]. International Conference on Computer Graphics and Interactive Techniques, 2010.
- [5] 王鑫. 数据驱动的人群动画仿真技术研究[D]. 浙江大学计算机科学与技术学院, 2009.
- [6] 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.
- [7] 刘更代, 潘志庚, 程熙, 等. 结合低维运动模型和逆运动学的风格化

人体运动合成[J]. 计算机辅助设计与图形学学报,2010,22(1): 145-151.

[8] Rose C, Cohen M F, Summons R E, et al. Verbs and adverbs: Multidimensional motion interpolation [J]. IEEE Computer Graphics and Applications, 1999, 18(5): 32-40.

[9] Jinxiang C, HJessica K H. Performance Animation from Low-dimensional Control Signals[J]. Siggraph 2005, 2005.

[10] Sidenblad H, Black M J, Sigal L. Implicit probabilistic models of human motion for synthesis and tracking[C]//European Conference on Computer Vision, 2002: 784-800.

[11] Jamie S, Andrew F, Mat C, et al. Real-time human pose recognition in parts from single depth images[J]. Institute of Electrical and Electronics Engineers, 2011.

[12] Xu G, Zhang Z. Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach[J]. Kluwer, 1996.

[13] Qi Xiqing, Chen S Y, Liu Sheng. A Novel 3D Reconstruction Method Based on Rotational Stereo [C]//2nd International Conference on Computer Vision Theory and Applications(VISAPP 2007), Barcelona, Spain: 8 - 11 March, 2007: 1-6.

(上接第 68 页)

表 4 归纳待证式的验证步骤

情景	验证步骤
情景 1	1, 3, 6, 7, 8, 9, 10, 11, 12, 13
情景 2	1, 3, 5, 17
情景 3	1, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13
情景 4	1, 3, 4, 5, 17
情景 5	1, 2
情景 6	14, 15, 16

由于奠基待证式和归纳待证式均得到验证,待证性质得到验证。当所有性质均验证完毕后,泛型订单管理服务组合的规范确定为完备的,则此泛型服务组合的可靠性得以保证。

如果将泛型订单管理服务组合的规范实例化,不必进行重复的验证。下面是一个实例化的例子——图书订单管理服务组合的定义。

Abbreviation : Book Order Management is
Order Management[with Book Order as Order,
Book Order Sequences as Order Sequences, ...]

5 相关工作比较

目前已有一些形式化方法在 Web 服务组合的规范和验证领域进行了有益探索,取得了一定成果。下面将基于概念的方法和一些同样描述过程定义的形式化方法进行比较说明异同。如表 5 所示。

表 5 不同形式化方法的功能比较

形式化方法	变量类型重用	控制流程	错误处理	补偿处理	服务规范重用
Petri-net	no	yes	no	no	no
Pi-Calculus	no	yes	no	no	no
ALBERT	no	yes	yes	yes	no
with Concepts	yes	yes	yes	yes	yes

基于 Petri-net^[6]的过程验证方法针对过程定义中控制流、数据

流和资源三维元信息相结合的冲突检测问题,研究了过程的语义验证方法。但没有考虑异常情况下的错误处理和补偿处理。

Pi-Calculus^[7]理论的主要内容包括语法定义、结构等价规则、操作语义、互相似理论、代数理论等。基于 Pi-Calculus 的组合规范和验证方法中提出了 OWL-S 涉及的八种控制结构,但不包括事件处理结构。另外也未考虑异常情况下的错误处理和补偿处理。

ALBERT^[8]是为 BPEL 过程设计的断言语言,通过模型检查支持工作流中设计时间的验证。该方法考虑了控制流程、错误处理、补偿处理。

以上几种方法均未实现变量类型的重用和服务规范的重用。

6 结 语

本文将基于概念的思想引入 Web 服务组合的规范和验证,其优点是提高规范的可重用性和简化验证的需要。通过基于概念的 Tecton 语言创建了泛型 Web 服务组合的形式化规范,规范包括服务变量及类型描述、通用函数描述、控制流程描述、异常情况处理描述。然后介绍如何应用结构化归纳法验证泛型 Web 服务组合的规范是否完备。并应用泛型订单管理服务组合的例子贯穿全篇,详细说明规范、验证、实例化的过程,易于理解。

Violet^[9]系统是基于 Tecten 语言的自动验证工具,其主要目的是辅助用户发现和理解 Tecten 语言规范的验证,有助于验证泛型 Web 服务组合中的一般规范。但如果泛型 Web 服务组合规范包括并行、迭代等控制流程以及异常情况处理的描述,情况比较复杂,目前难以通过 Violet 系统实现自动验证。因此今后研究的方向是完善 Violet 系统,实现任意情况下泛型 Web 服务组合规范的自动验证,并希望建立经过验证的泛型 Web 服务组合库供应用者分享。如果以已验证的泛型 Web 服务组合概念为基础构建新的泛型组合概念,可靠性将得以提高。

参 考 文 献

[1] 梁爱虎. SOA 思想、技术与系统集成应用详解[M]. 北京: 电子工业出版社, 2007.

[2] Michael P Papazoglou. Web 服务原理与技术[M]. 北京: 机械工业出版社, 2010.

[3] Musser D R, Shao Z. Concept use or concept refinement: an important distinction in building generic specifications[C]//Proc. 4th International Conference on Formal Engineering Methods, LNCS 2495. Berlin: Springer-Verlag, 2002: 132-143.

[4] Musser D R, Shao Z. The Tecton concept description language (revised version). Technical Report TR-02-03 [R]. Computer Science Department, Rensselaer Polytechnic Institute, Troy, New York, 2002.

[5] Burstall R. Proving properties of programs by structural induction[J]. Computer Journal, 1969, 12(1): 41-48.

[6] 周建涛, 史美林, 叶新铭. 一种基于 Petri 网化简的工作流过程语义验证方法[J]. 软件学报, 2005, 16(7): 1242-1251.

[7] Yong-Lian Wang, Xue-Li Yu. Formalization and Verification of Automatic Composition Based on Pi-Calculus for Semantic Web Service [C]//Second International Symposium on Knowledge Acquisition and Modeling. IEEE Computer Society Digital Liabrary, 2009: 103-106.

[8] Baresi L, Bianculli D, Ghezzi C, et al. Validation of Web service compositions[J]. IET Software, 2007, 1(6): 219-232.

[9] 翟洁, 邵志清. 基于 Tecten 的验证系统——Violet[J]. 华东理工大学学报, 2005, 31(2): 198-202.