

基于改进的数据驱动决策树分析的 3维人体运动检索

向坚^{1),2)} 徐劼¹⁾ 郭同强¹⁾ 吴飞¹⁾ 庄越挺¹⁾

¹⁾(浙江大学计算机学院, 杭州 310027) ²⁾(浙江科技学院信息与电子工程学院, 杭州 310023)

摘要 随着大量3维人体运动捕获数据库的出现,使得如何对人体运动数据进行高效分析和处理,从而有效利用运动捕获数据库成为一个新的挑战。为了高效地进行3维人体运动检索,首先通过从人体运动中提取一种基于3维空间变换特征规律的空间变换特征和运动的一些关键的时间特性来得到人体运动的3维时序特征;然后针对不同的训练需求,通过改进的数据驱动决策树的学习方法来分析关节点对运动相似的不同影响,并在检索过程中按照不同影响程度依次对关键点进行相似度计算;最终实现了一个高效的运动检索仿真系统。

关键词 运动捕获 3维时序特征 决策树 检索

中图分类号: TP391.3 **文献标识码**: A **文章编号**: 1006-8961(2007)10-1761-05

Human Motion Retrieval With Time-sequence Features Based on Data-driven Decision Tree Analysis

XIANG Jian^{1),2)}, XU Jie¹⁾, GUO Tong-qiang¹⁾, WU Fei¹⁾, ZHUANG Yue-ting¹⁾

¹⁾(College of Computer Science, Zhejiang University, Hangzhou 310027)

²⁾(School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023)

Abstract With the development of motion capture techniques, more 3D motion libraries become available. In this paper, space transformation property and time property are extracted from human motion capture data, which compose 3D motion time sequence features. Given the assumption of the features of each joint is independent, a revised data-driven decision tree of anytime induction is automatically constructed to present the influence of each point during the comparison of motion similarity. Experiment results show that our approaches are effective for motion data retrieval.

Keywords motion capture, 3D time sequence feature, decision tree, retrieval

1 引言

20世纪90年代以来,随着运动捕获技术的兴起,以及设备技术的进步,大量的3维人体运动捕获数据生成,并被广泛应用在计算机动画、游戏、医学仿真、电影特技^[1]等领域。在此背景下,基于运动捕获数据的研究,已经成为近年来计算机视觉、图形学研究领域和动漫应用领域的一大热点。为了高效地进行3维人体

运动检索,需要研究一种高效的运动数据检索技术,以支持对运动进行编辑、变形和合成等处理。

为了应对如上挑战,本文首先提出了一种描述各关节点之间3维空间关系的空间变换特征。与以往在运动特征提取中一般提取关节点的坐标、速度和角速度等2维数学特征不同^[2,3],本文提取的空间变换特征是基于各关节点3维空间关系的3维特征^[4]。由于传统运动特征是2维的,所以需要所有关节点或者多个关节点的运动特征结合起来才能完

基金项目:国家自然科学基金项目(60525108,60533090);浙江省科技计划项目重大科技项目(2005C13032,2006C13097);长江学者和创新团队发展计划(IRT0652)

收稿日期:2007-06-20;改回日期:2007-07-18

第一作者简介:向坚(1976~),男,2007年6月获浙江大学博士学位,现任教于浙江科技学院。主要研究领域为多媒体分析与检索、计算机动画、统计学习等。E-mail:freexiang@gmail.com

整地表征每一个运动。而 3 维空间变换特征则相对独立,即每一个关节点都能独立地体现整体运动的一个局部。进行变换时,首先通过时间特性和空间变换特征的结合来得到人体运动完整的 3 维时序特征;然后分别处理每一个关节点的局部时序特征,并利用决策树学习各关节点的相似程度来得到其对整体相似的影响,这就在检索过程中避免了很大一部分无意义的相似度计算。更进一步,传统的运动特征因为直接从原始数据中提取,所以通常维数都比较高,由于会增加较高时序开销,于是大多需要一个降维的过程,而 3 维时序特征则避开了和原始数据的直接接触,从而也在很大程度上避免了维数灾难。

2 3 维时序特征

本文所采用的人体骨骼模型如图 1 所示。

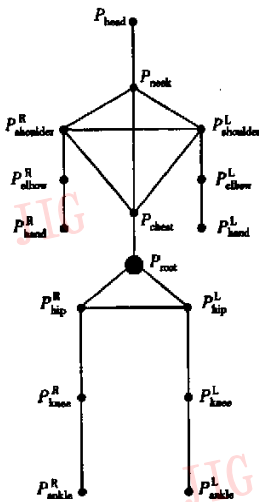


图 1 人体骨骼模型

Fig. 1 Human skeleton model

各关节点的世界坐标可以由骨架长度和旋转向量求出。于是可得到以下 3 维变换公式:

$$P_i^{(j)} = T_i^{(root)} R_i^{(root)} \dots T_0^{(grandparent)} R_i^{(grandparent)}(t) T_0^{(parent)} R_i^{(parent)} P_0^{(j)} \quad (1)$$

其中, $P_i^{(j)}$ 为第 j 个关节点在时刻 i 的世界坐标, $T_i^{(j)}$, $R_i^{(j)}$ 为时刻 i 针对关节点 j 的平移矩阵和旋转矩阵。

由式(1)可计算出各个关节点的世界坐标,即可得到一个 51 维数据,除去根节点(root),这样就是 16 个关节点,48 维数据。基于这 48 维人体运动数据,首先针对人体的上半部分和下半部分分别定

义一个空间集合 S_{up} 和 S_{down} ,其中 $s_{k,i} \in S_{up}, i = 1, 2, \dots, m; s_{l,j} \in S_{down}, j = 1, 2, \dots, m; m$ 为空间集合的空间数,现在将上下部分划分为等数量的空间集合, $s_{k,i}, s_{l,i}$ 为上,下空间集合中的独立空间。以根节点为基准,将根以上的关节点对应于 S_{up} ,根以下的关节点对应于 S_{down} ,当某上肢关节点进入空间 $s_{k,i}$ 的时候,其空间变换特征对应于 $s_{k,i}$ 的值。

下面定义几个空间划分规则:

$$Fr(N_i, N_j) = \begin{cases} 1 & N_i \text{ in front of } N_j \\ 0 & N_i \text{ behind of } N_j \end{cases}$$

$$L(N_i, N_j) = \begin{cases} 1 & N_i \text{ left to } N_j \\ 0 & N_i \text{ right to } N_j \end{cases}$$

$$H(N_i, N_j) = \begin{cases} 1 & N_i \text{ above } N_j \\ 0 & N_i \text{ below } N_j \end{cases}$$

$$Fa(N_i, N_j) = \begin{cases} 1 & N_i \text{ distance from } N_j > \lambda \\ 0 & N_i \text{ distance from } N_j < \lambda \end{cases}$$

其中,Fr, L 和 H 规则产生于关节点 N_i 相对于关节点 N_j 的上下左右的空间关系, Fa 规则是根据关节点运动的幅度,即根据和 N_j 的距离是否大于某个阈值 λ 而计算出来的。通常规则中, Fr, L 规则中的 N_i 都是根节点,而 H, Fa 规则则视上下身体有所不同。 N_i, N_j 是同一帧中的关节点。

定义运动的空间变换特征 $B = (b_1, b_2, \dots, b_n)$, $b_i = (s_{i,1}, s_{i,2}, \dots, s_{i,16})$, 其中 b_i 为运动中关节点 i 的空间变换特征向量, $s_{i,j}$ 表示第 j 帧中关节点 i 的空间变换特征。假设 $s_{a,j}$ 表示一个上半肢体的关节点 a 的空间变换特征。表 1 是利用以上定义的规则生成的空间变换特征 $s_{a,j}$ 的空间规则,由此表可以推出,用 $s_{a,j} = s_{k,1}$ 可产生以下规则:

$$\begin{aligned} & Fr(N_{a,1}, N_{a,r}) \wedge L(N_{a,1}, N_{a,r}) \\ & \wedge H(N_{a,1}, N_{a,r}) \wedge Fa(N_{a,1}, N_{a,r}) \\ s_{a,j} = s_{k,1} & \text{ 可产生以下规则:} \\ & \neg Fr(N_{a,1}, N_{a,r}) \wedge \neg L(N_{a,1}, N_{a,r}) \\ & \wedge \neg H(N_{a,1}, N_{a,r}) \wedge \neg Fa(N_{a,1}, N_{a,r}) \end{aligned}$$

表 1 空间规则

Tab. 1 The space rules

$s_{a,j}$	$Fr(N_{a,j}, N_{r,j})$	$L(N_{a,j}, N_{r,j})$	$H(N_{a,j}, N_{r,j})$	$Fa(N_{a,j}, N_{r,j})$
$s_{a,j} = s_{d,1}$	1	1	1	1
$s_{a,j} = s_{d,2}$	0	1	1	1
...
$s_{a,j} = s_{d,m}$	0	0	0	0

而 Fr, L, H 等规则,则是通过原始运动数据中的 48 维数据计算得到,因为规则都是在一帧的不同节点之间计算,所以计算量很小。由此通过对所有关节点进行计算,即可从运动原始数据中得到人体运动的空间变换特征。不仅这个空间变换特征的每一个局部是针对人体的每个关节点的,而且也是相对独立的。

综上所述,到目前为止提取的空间变换特征,完全是运动空间特性的反映。但是对于一个完整的运动来说,由于它首先就是一组时序数据,如果运动特征中没有时间特性的体现,那么就会漏掉运动的很多关键地方,因此可将提取运动的时间特性作为整个时序特征的一个组成部分,使之能更好地体现出运动的时序性。

这里利用空间变换特征速度来提取空间速度相似度,这是度量两个运动快慢的决定性因素,这也是最常用的时间特性。由于各关节点的空间变换特征具有相对独立性,所以各关节点的空间特征变换速度也是相对独立的。算法如下所示:

Procedure $V_{space}()$

Input: 运动的第 k 个节点的局部空间变换特征 $s_k = (s_{1,k}, s_{2,k}, \dots, s_{n,k})$, n 为运动的帧数

Output: V_k^{space} 运动的第 k 个节点的空间特征变换速度。

- (1) 初始化 $N_j = 0, i = 1, j = 0$
- (2) if $s_{i,k} \neq s_{(i+1),k}$ $V_{i,k}^{space} = N_j, l = l + 1, j = j + 1$
else $N_j = N_j + 1;$
- (3) $i = i + 1$, if 到了帧尾则转到步骤(4), 否则转到步骤(2)
- (4) return V_k^{space}

这个空间变换速度实际上反映的是运动的某个关节点从一个空间到另一个空间的变换速度,所有关节点的空间变换速度的加权和就构成了整个运动的空间特征变换速度。

噪声空间定义如下:在相似匹配中,由于可能会有一些零星的空间无法得到匹配,从而影响匹配结果,但是事实上由于运动的不规则和偶发性总会导致这样的空间的出现,所以可以通过计算出噪声空间来度量其对整个匹配过程中的影响程度,算法如下:

Procedure $N_{space}()$

Input: 运动的第 k 个节点的局部空间变换特征 $s_k = (s_{1,k}, s_{2,k}, \dots, s_{n,k})$, n 为运动的帧数

- Output: N_k^{space}
- (1) 初始化 $N_j = 0, i = 1, j = 0, l = 1;$
 - (2) if $s_{i,k} \neq s_{(i+1),k}$

$V_{i,k}^{space} = N_j, j = j + 1,$

if $\frac{N_j}{n} < \epsilon$ $N_{i,k}^{space} = s_{i,k}, l = l + 1;$

else

$N_j = N_j + 1;$

(3) $i = i + 1$, if 到了帧尾则转到步骤(4), 否则转到步骤(2)

(4) return N_k^{space}

在得到关节点的空间特征变换速度和噪声空间之后,就形成了运动数据所包含的完整语义特征,加上本文得到的 3 维空间变换特征就得到了每一个关节点的 3 维时序特征。

3 改进的决策树

对于决策树的学习,最初的系统是亨特等人设计的 CLS (concept learning system)^[5-7]。现在普遍采用的是基于概率论的分类属性方法,其中最著名的一种就是 Quilan 在 1986 年提出的决策树生成算法 ID3 (induction of decision tree)^[6],之后还提出了 C4.5^[7], C5 等改进算法。

然而以上这些决策树生成算法通常需要提供一个局部最优假设的固定时间和结果。而当用户需要更多的时间去训练,以便得到更精确结果的时候,这些算法的弱点就凸显无疑。因此本文采用了一种 ID3 的扩展算法 LSID3 (lookahead by stochastic ID3)^[8],用来生成任意时间以及精度需求的决策树,以提高对运动分析的精确度。算法如下:

Procedure LSID3-Choose-Attribute(E, A, r)

- (1) if $r = 0$
Return ID3-Choose-Attribute(E, A)
- (2) for each $a \in A$
for each $v_i \in a(E)$
 $E_i = \{e \in E \mid a(e) = v_i\}$
 $min_i = \infty$
for $i = 1: r$
 $min_i = \min(min_i, ID3(E_i, A - \{a\}))$
 $total_i = \sum_{i=1}^r min_i$
return a for which $total_i$ is minimal

通过对运动库中 100 多种类型,500 ~ 600 个运动数据的训练,得到的一棵关于关节点对运动相似计算影响程度的决策树如图 2 所示。

由训练数据得到的检索决策树可以判断,双足的空间变换特征对人体运动的影响最大,其次是手,

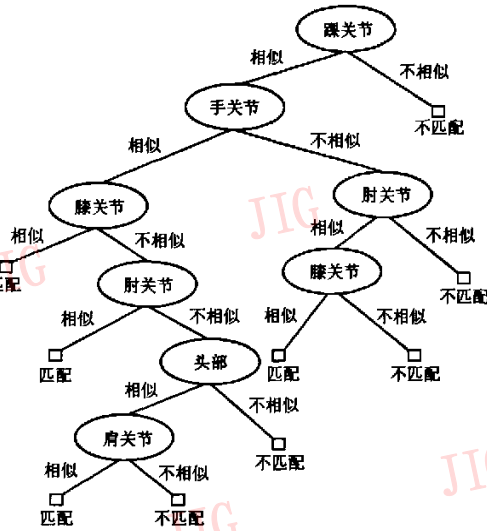


图 2 关节的检索决策树
Fig. 2 Decision tree of each joint for retrieval

而之后是肘部和膝盖,依次类推。这样在检索中就可以首先对两个运动的双足空间变换特征进行相似度计算,只有当其相似的时候,才继续对双手的空间变换特征的相似度进行计算,如果不相似,则再计算肘部的相似度,如果不相似,则直接终止计算过程。通过实验发现,根据决策树的检索过程可以避免大量没必要的相似计算,从而使得检索效率大大提高,对加速整个运动数据的处理有着很大的意义。

4 实验结果和分析

本文在 matlab 上实现了基于时序特征的运动检索,实验是在实验室的 Mocap 设备捕获得到的一系列运动数据的基础上展开的。其中用作测试的运动库包含有 1 300 多个运动序列,上百种运动类型。

表 2 是针对不同关节加上不同权重,采用对人体所有关节进行相似度计算的方法和利用改进的决策树学习方法所提高的检索效率的结果比较,由表 2 可以看到,基于决策树学习的检索方法显著地提高了检索的性能,比传统的将所有人体关节的特征全部进行相似度比较的检索方法具有很大的优越性。

图 3 显示了传统的决策树^[4]和改进的决策树分别对运动分析之后完成运动检索的精度比较,可以看出改进的决策树分析之后的检索精度得到了明显

表 2 检索时间比较

Tab. 2 Comparison of retrieval time

运动序列	基于权重的全部关节相似度计算的方法		基于决策树的学习方法	
	$N=200$ 的检索时间(s)	$N=800$ 的检索时间(s)	$N=200$ 的检索时间(s)	$N=800$ 的检索时间(s)
A(51)	1.651 2	4.741 3	0.734 2	1.876 1
B(97)	1.902 2	5.102 1	0.821 3	2.197 5
C(300)	2.241 3	5.633 2	1.123 3	2.321 1

注:A(51)代表一个有 51 帧数据的运动序列

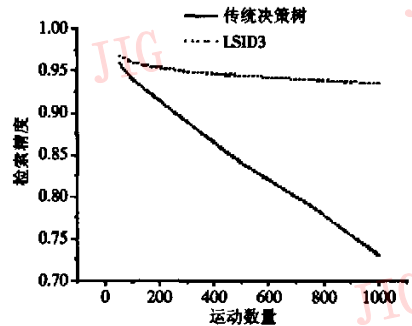


图 3 传统和改进决策树分析的检索精度比较

Fig. 3 Comparison of retrieval accuracy by conventional decision tree with LSID3

的提高。而表 3 则列出了对于相同的数据库本文算法所得到的和采用目前普遍使用的基于八段骨骼特征和关键帧提取的检索算法的查全率(recall)和查准率(precision)。由此可以发现,基于时序特征和决策树学习的运动检索有着更理想的效果和更好的效率,而且对复杂一些的运动也有着很好的实用性。

表 3 不同检索算法检索结果比较

Tab. 3 Comparison of retrieval performance

运动序列	查全率(%)		查准率(%)	
	KF	SF&DT	KF	SF&DT
Walk	85	91	90	96
Punch	40	79	60	88
Kick	50	80	70	89

注:KF 代表基于八段骨骼特征和关键帧提取的检索算法, SF&DT 是基于时序特征和决策树学习的检索算法

5 结论

本文提出了一种包含运动空间变换特征和时间特性的 3 维时序特征。并在关节的时序特征相对

独立的特性前提下采用了数据驱动的方法生成有关各关节点重要性的改进决策树,然后在改进决策树的指导下完成检索过程。实验证明,本文提出的时序特征高效而且有效的反映了运动的特性,并且通过决策树学习的方法也使检索过程简单和快速。

参考文献 (References)

- 1 Zhuang Y T, Rui Y, Huang T S. Adaptive key frame extraction using unsupervised clustering[A]. In: Proceedings of IEEE International Conference on Image Processing[C], Chicago, IL, USA, 1998: 886 ~ 870.
- 2 Luo Zhong-xiang, Zhuang Yue-ting, Pan Yun-he, et al. Video based motion capture[J]. Journal of Image and Graphics, 2002, 7(8): 752 ~ 756. [罗忠祥, 庄越挺, 潘云鹤等. 基于视频的运动捕获[J]. 中国图象图形学报, 2002, 7(8): 752 ~ 756.]
- 3 Jehee L, Jinxiang C, Paul S A R, et al. Interactive control of avatars animated with human motion data [A]. In: Proceedings of SIGGRAPH[C], San Antonio, Texas, USA, 2002: 491 ~ 500.
- 4 Xiang J, Zhuang Y T, Wu F. Data-driven automatic generation of decision tree for motion retrieval with temporal-spatial features[A]. In: Proceedings of International Conference on Machine Learning and Cybernetics[C], Dalian, China, 2006: 2718 ~ 2725.
- 5 Hunt E B, Marin J, Stone P T. Experiments in Induction[M]: New York: Academic Press, 1966.
- 6 Quilan J R. Generating Production rules from decision tree[A]. In: Proceedings of International Joint Conference on Artificial Intelligence [C], Milan, Italy, 1987: 304 ~ 307.
- 7 Quilan J R. C4.5: Programs for Machine Learning[M]. San Mateo, CA, USA: Morgan Kaufmann, 1993.
- 8 Esmeir S, Markovitch S. Lookahead-based algorithms for anytime induction of decision trees[A]. In: Proceedings of The Twenty-First International Conference on Machine Learning[C], Banff, Canada, 2004: 257 ~ 264.