

一种高效的三维运动检索方法

向 坚^{1,2}

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

摘 要 随着运动捕获设备的普及,大量的运动数据可以直接得到,从而使得大规模的运动数据库的建立成为可能。在此背景下,研究以检索为核心的运动捕获数据处理技术就显得十分重要了。本文提出了一种对运动捕获数据中的人体的各个关节提取一种基于三维空间变换规律的 3D 时空特征的方法,并基于时空运动连续性引入了关键空间的概念。针对各关节点时空特征相对保持独立的特性,本文用每个关节点作为索引,并通过数据驱动决策树的学习方法去分析关节点对运动相似的不同影响,最终实现了一个高效的运动检索系统。

关键词 运动捕获, 三维时空特征, 决策树, 运动检索

Motion Retrieval Based on 3D Spatial-Temporal Features and Decision Tree

XIANG Jian^{1,2}

(School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023)¹

(College of Computer Science, Zhejiang University, Hangzhou 310027)²

Abstract With the development of Motion Capture techniques, more and more 3D motion libraries become available. In this paper, we present a novel approach for motion retrieval based on data-driven decision tree with spatial-temporal features. First the 3D time spatial features of each human joint are extracted with the help of key-space. Given the assumption of the features of each joint is independence, data-driven decision tree 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 spatial-temporal feature, Decision tree, Motion retrieval

1 前言

近年来,由于设备技术的进步,生成了大量的三维人体运动捕获数据,在计算机动画、游戏、医学仿真等领域得到了广泛的应用^[1,2]。为了更加有效利用大规模的三维运动库所包含的运动数据,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。

在对三维运动数据处理检索时,主要存在如下问题:

(1)运动是各关节点信号的和谐组合,在检索过程中需要合理的运动特征描述机制。因此,提取什么样的运动特征,以及用什么样的表达方式来表达运动特征,对运动处理的效果和效率会有极大的影响。

(2)从运动数据中提取的特征维数通常都非常高,每个数据之间的距离将因为中央极限定律而变得近乎相同,无法区分彼此,产生高维灾难。因此,从高维特征中寻找“有意义”的低维特征,可以避免导致较高的时间和空间复杂度,提高检索效率。

(3)运动时序信号之间的相似度度量十分困难。在度量运动数据相似性时,既要考虑数据在空间上相似性,也要考虑其时间上的相似性。

为了解决如上挑战,本文首先提出了一种描述各关节点之间三维空间关系的时空特征。与以往在运动特征提取中一般提取关节点的坐标、速度和角速度等二维数学特征不同^[3~5],本文提取的时空特征是基于各关节点三维空间和时间关系的三维特征。传统运动特征是二维的,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整地表征每一个运动。而三维时空特征则不然,其各关节点的时空特征是三维的,因此相对独立,每一个关节点都能独立地体现整体运动的一个局部。更进一步,传统的运动特征因为直接从原始数据中提取,通常维数都比较高,会增加较高时空开销,于是大多需要一个降维的过程。而三维空间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。

最后,衡量运动数据之间相似度大多数采用弹性匹配算法(DTW)^[6],但是 DTW 的时间复杂度相当高。本文通过数据驱动的决策树学习,得到关节点对运动影响的权重。这样,在检索中先将运动例子 Q 和运动库中运动 A 内权重最大的关节点的索引表进行相似度计算,如果结果不相似,就可以跳过 A,直接和下一个运动 B 进行相似度计算。只有当权重大的关节点的相似度在某一阈值范围之内的时候,才对其后续关节点继续进行比较。这样的检索过程避免了很大一部分无意义的相似度计算。

2 特征提取

2.1 运动模型

本文所采用的人体骨骼模型如图 1 所示。共有 16 个关节点,各关节点的组织呈树形结构,由从 Root 节点往各末端关节点(比如 Hand, Ankle)的子树组成。Root 关节点决定人体的当前位置和朝向,其他关节点位置则由其和父节点的相对旋转角度确定,整个人体运动由围绕 Root 的平移和旋转组成。我们通过运动捕获设备得到的是这 16 个关节点相对于

其父节点的相对空间位置的角度(除了Root节点之外),通常我们都要把它转换成世界坐标系或者其他参考坐标系里的坐标,再来进行数据处理。各关节的世界坐标可以由骨架长度和旋转向量求出,于是得到一个三维变换公式:

$$\vec{p}_i^{(j)} = T_i^{(root)} R_i^{(root)} \dots T_o^{(grandparent)} R_i^{(grandparent)}(t) T_o^{(parent)} R_i^{(parent)} \vec{p}_o^{(j)} \quad (1)$$

其中, $\vec{p}_i^{(j)}$ 表示第 i 帧中关节点 G_i 的世界坐标, $T_i^{(root)}$, $R_i^{(root)}$ 表示第 i 帧中 Root 关节点的平移和旋转向量, $T_o^{(k)}$ 表示初始时关节 N_k (N_k 为树形人体骨架中从根结点到结点 N_j 之间任意结点)在其父关节点所在局部坐标系下偏移量生成的平移变换矩阵; $R_i^{(k)}$ 表示 i 时刻关节点 N_k (N_k 取值同上)的旋转变换矩阵,由 r_i^k 生成; $\vec{p}_o^{(j)}$ 表示初始时在 N_j 其父结点所在局部坐标系下的偏移量。

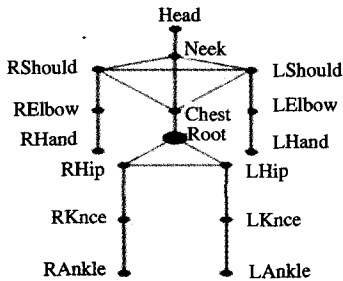


图1 人体骨架

2.2 时空特征的提取

我们利用式(1)计算出各个关节点的世界坐标,得到一个五十一维数据,除去Root节点,这样就是十六个关节点、四十八维数据。

运动表示为

$$M_i = (F_1, F_2, \dots, F_i, \dots, F_n)$$

$$F_i = (p_{i1}, p_{i2}, \dots, p_{ij}, \dots, p_{in}); p_{ij} = (x, y, z) \quad (2)$$

其中 n 为运动数据的帧数, p_{ij} 为第 i 帧中第 j 个节点的世界坐标。

利用这四十八维数据生成运动各关节点的空间变换。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 $S_{ki} \in S_{up}, i=1, 2, \dots, m, S_{lj} \in S_{down}, j=1, 2, \dots, m, m$ 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_{ki}, S_{lj} 为上、下空间集合中的独立空间。以Root节点为基准,将Root以上的关节点对应于 S_{up} , Root以下的关节点对应于 S_{down} 。当某上肢关节点进入空间 S_{ki} 的时候,其空间变换对应于 S_{ki} 的值。下面定义几个空间分划规则:

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

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

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

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

其中 front, left 和 high 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,而 far 规则是根据关节点运动的幅度,即和 N_j 的距离是否大于某个阈值 λ 而计算出来的。通常,我们的规则中 front, left 中的 N_j 都是 Root 节

点,而 high, far 规则视上下身体有所不同。 N_i, N_j 都是同一帧中的关节点。

定义运动的空间变换 $B = (b_1, b_2, \dots, b_n)'$, $b_i = (s_{i1}, s_{i2}, \dots, s_{i16})$, 其中 b_i 为运动中关节点 i 的空间变换, s_{ij} 表示关节点 i 中第 j 帧的空间变换。假设 s_{aj} 表示一个上半肢体的关节点 a 的空间变换,可以推出,当 $s_{aj} = s_{kj} \Leftrightarrow$ 规则 $front(N_{a1}, N_{ar}) \wedge left(N_{a1}, N_{ar}) \wedge high(N_{a1}, N_{as}) \wedge far(N_{a1}, N_{as})$

...

$s_{aj} = s_{kj} \Leftrightarrow$ 规则

$$\neg front(N_{a1}, N_{ar}) \wedge \neg left(N_{a1}, N_{ar}) \wedge$$

$$\neg high(N_{a1}, N_{as}) \wedge \neg far(N_{a1}, N_{as})$$

而 front, left, high 等规则,则是由式(2)中的四十八维数据计算得到,因为规则计算都是在同一帧的不同节点之间,所以计算量很小。由此我们对所有关节点进行计算,从运动原始数据中得到了人体运动的空间变换。这个空间变换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。

到目前为止提取的空间变换,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,它首先就是一组时序数据,所以运动特征中如果没有时间特性的体现,就会漏掉运动的很多关键地方,因此我们将提取运动的时间特性,作为整个时空特征的一个组成部分,使得整个特征能很好地体现出运动的时空性。

这里利用空间变换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,也是最常用的时间特性。由于各关节点空间特征的相对独立性,因此各关节点的空间变换速度也是相对独立的。这个空间速度实际反映的是运动的某个关节点从一个空间到另一个空间变换的速度,所有关节点的空间速度的加权和构成了整个运动的空间变换速度。然后我们求解空间噪音;在相似匹配中,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,所以可以计算出空间噪音来度量其对整个匹配过程中的影响程度。

得到十六个关节点的空间变换、空间速度以及空间噪音之后,就形成了运动数据所包含的完整语义特征。将其合并,得到了每一个关节点的三维时空特征。

因为时空的连续性,所以一个关节点在连续的几个帧中一般都会在同一时空中,可以把具有相同时空的帧合并,生成一段关键时空,这样就将一段有很多帧的运动剪切成关于每个关节点的一个个时空。这实际上也类似于传统的提取关键帧的方法,但是关键时空是基于单个关节点的,实际上是个局部关键帧,不同于以往意义上的关键帧。

这样,我们将运动的时空特征简化为每个关节点的关键时空,运动的相似度计算也直接在关键时空上进行,大大降低了计算量。

我们将十六个关键点的关键时空作为十六个索引建立起检索运动库。对于一个检索的运动例子,我们提取十六个关键时空和检索运动库中的运动分别进行相似度匹配,实现对运动的检索。

3 决策树

在检索中,我们通过运动的时空特征对一系列运动数据进行相似度计算。因为运动的时空特征对于各个关节点来说相对独立,所以我们可以先对两个运动的各个关节点的时空特征单独进行相似度计算,然后组合计算整体的相似度。在

单词类型以及出现位置记录下来。在提取过程中,如果在固定匹配位置出现了杂质单词,即匹配〈word, feature〉中的第一项相同,则直接将其过滤掉。若出现了与杂质单词不相同,但与杂质类型相同的单词,则也将其判断为杂质并过滤掉。

因为本文针对的都是半结构化的教育类网站和资源,网页内容比较规范,很多时候,冗余信息是相似的或者相同的,因此,将针对 word 本身的判断作为首要判断标准,如果发现多余信息与已有记录相同或相似,则可以简单地做出判断,而不需要额外的工作。

5 性能评测

衡量一个元数据提取系统好坏的指标主要有以下三个:

$$\text{查全率(Recall)} \quad R = \frac{\text{正确提取结果数}}{\text{所有可能提取结果数}}$$

$$\text{查准率(Precision)} \quad P = \frac{\text{正确提取结果数}}{\text{所有提取结果数}}$$

$$\text{F度量(F-measure)} \quad F = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}$$

F度量让用户在查全率和查准率上求得平衡。而我们认为元数据提取的查准率比查全率要重要一些,因为准确的数据是服务质量的保证,而数据的完备应该建立在数据准确的基础之上。因此,选择查准率和查全率的重要程度值 $\beta=0.5$,代表 P 的重要程度是 R 的 2 倍。

表 1

	OCW	UIUC	WISC
课程 id'	0.995	0.998	0.996
课程名称	0.995	0.998	0.990
课程描述	0.890	0.936	0.997
参考读物信息	0.821	—	0.897
教师信息	0.857	0.890	—
学时信息	0.902	—	0.953
教学大纲	0.798	—	—

测试数据包括 MIT 的 OCW 网站中 1900 门课程信息;美国伊利诺斯大学(UIUC)的 3000 余门课程资源;以及威斯康星大学麦迪逊分校(WISC)的 1000 门课程资源。表 1 中列出了不同元数据信息域在三所大学课程网页中提取效果的 F 度量值。

(上接第 86 页)

表 2 KF 代表基于八段骨骼特征和关键帧提取的检索算法, SF&DT 是基于时空特征和检索树学习的检索算法

运动序列	Recall		Precision	
	KF	SF&DT	KF	SF&DT
Walk	0.85	0.89	0.89	0.91
Run	0.70	0.85	0.83	0.93
Punch	0.40	0.75	0.59	0.81
kick	0.50	0.75	0.66	0.84

总结和展望 为了有效地对通过运动捕获设备得到的大规模的三维运动数据库进行分析和处理,我们针对运动捕获数据中的人体运动的特征进行了研究,提出了一种在三维空间考虑运动特征的方法,并在关节的时空特征相对独立的特性前提下生成有关各关节重要性的决策树,然后在决策树的指导下完成检索过程。实验证明,本文提出的空间特征高效而且有效地反映了运动的特性,并且通过决策树学习的方法也使检索过程简单和快速。

由于不是所有本项目中需要的元数据域都能在各个大学课程网页中涉及,因此上表中有些数值无法统计,例如 UIUC 的课程网页中没有包含学时信息和教学大纲等信息,而 WISC 网页中没有包含教师信息和教学大纲。

基于以上图表可以看出,对于一些格式比较固定的信息域,如课程 id,课程名称,课程描述等,本项目的提取效果能够保证在 90% 以上,为实际应用提供了非常好的元数据资源。对于一些变化较大的信息域,如参考读物信息,教师信息,教学大纲等,在描述格式较为固定的课程网站中(如 UIUC 和 WISC),本项目仍然能够达到 90% 左右的准确效果,而在格式不是非常固定的情况下,提取效果也能够达到 80%~85% 的准确程度,同样能够满足实际应用的需要,完成绝大多数网站元数据的整合任务。

总结与展望 本文在单一规则匹配提取方法的基础上,设计实现了一种按照优先级处理多种规则的元数据提取方法,并且对抽取的结果进行信息精化。本项目的创新之处在于,能够全自动地按照不同规则对不同网页进行元数据提取,而不需要用户了解并修改程序内部。实际应用的结果表明,该方法能够很好地处理半结构化网页。

现有系统仍然有需要进一步完善的地方,例如如何实现自动发现规则,以适应海量环境多种异构数据源的快速集成处理。

参考文献

- 1 刘世杰,唐世渭,杨冬青,王腾蛟,李立宇. 基于 XML 技术的 Web 信息提取和集成. 见:第二届全国数据库学术会议,2003
- 2 Crescenzi V, Mecca G, Grammars have Exceptions. Information Systems 1998,23 (8); 539~565
- 3 Garcia-Molina H, Papakonstantinou Y, Quass D, et al. The TSIMMIS Approach to Mediation; Data Models and Languages (extended abstract). In NGITS, 1995
- 4 Arocena G, Mendelzon A. WebOQL: Restructuring Documents, Databases, and Webs. In: Proc. ICDE '98, Feb. 1998
- 5 Huck G, Fankhauser P, Aberer K, Neuhold E.J. Jedi; Exchanging and Synthesizing Information from the Web. Coopis,1998
- 6 Califf M E, Mooney R.J. Relational Learning of Pattern-Match Rules for Information Extraction. In: Proceedings of the Sixteenth National Conference on Artificial Intelligence and Eleventh Conference on Innovative Applications of Artificial Intelligence, Orlando, Florida, 1999. 328~334
- 7 Freitag D. Machine Learning for Information Extraction in Informal Domains. Machine Learning, 2000,39(2-3);169~202

参考文献

- 1 Zhuang Yueting, Rui Yong, Huang T S. Adaptive Key-Frame Extraction Using Unsupervised Clustering [A]. In: IEEE ICIP'98, Chicago, USA, Oct. 1998
- 2 Wolf W. Key frame selection by motion analysis [J]. Acoustics, Speech, and Signal Processing, 1996,2: 1228~1231
- 3 Lim I S, Thalmann D. Key-Posture Extraction Out of Human Motion Data by Curve Simplification [C]. In: Proc. EMBC2001, 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2001. 1167~1169
- 4 Lee Jehce, Chai Jinxiang, Reitsma P S A, et al. Interactive Control of Avatars Animated with Human Motion Data [A]. In: Proceedings: SIGGRAPH 2002 [C], San Antonio, Texas, 2002. 491~500
- 5 Chui Y, Chao S P, Wu M Y, et al. Content-based Retrieval for Human Motion Data [J]. Journal of Visual Communication and Image Representation, 2004,16(3): 446~466
- 6 Liu F, Zhuang Y, Wu F, et al. 3d Motion Retrieval with Motion Index Tree [J]. Computer Vision and Image Understanding, 2003, 92(2-3): 265~284
- 7 Hunt E B, Marin J, Stone P T. Experiments in induction. New York: Academic Press,1966
- 8 Quilan J R. Generating Production rules from decision tree. In: Proceeding of IJCAI-87,1987
- 9 Quilan J R. C4. 5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufmann,1993