

## 基于谱聚类的运动捕获数据分割

胡晓雁, 孙 波, 朱小明, 魏云刚

(北京师范大学信息科学与技术学院 北京 100875)  
(huxy@bnu.edu.cn)

**摘要:** 为将长运动中所包含的不同运动自动分割出来, 提出一种基于谱聚类的长运动数据分割算法. 首先将运动捕获数据分解成长度相等的运动数据片段, 并基于主成分分析来计算这些小片段之间的相似度, 得到运动数据相似度矩阵; 然后用谱聚类算法将相似度矩阵转换为相应的拉普拉斯矩阵, 计算其前若干个特征向量, 并采用  $K$  均值算法获得聚类结果; 由于上述相似度矩阵直接实施谱聚类算法所得到的类别标签序列包含大量噪声, 采用统计滤波算法对分类结果进行了处理, 获得最终的分割点. 在 14 个运动数据上进行自动分割测试, 验证了文中算法的有效性.

**关键词:** 运动捕获数据; 谱聚类; 运动分割; 主成分分析  
**中图分类号:** TP391.41

## Motion Capture Data Segmentation Based on Spectral Clustering

Hu Xiaoyan, Sun Bo, Zhu Xiaoming, and Wei Yungang

(College of Information Science and Technology, Beijing Normal University, Beijing 100875)

**Abstract:** A long motion capture data often contains several different motions or the same motion repeats several times. It is an important topic to segment a long motion capture data into different motions. In this paper, we propose a method to segment long motions into several different motions by using a spectral clustering algorithm. When computing similarity matrices, we first cut the original motion capture data into mocap clips and each clip contains  $k$  frames mocap data. We then apply PCA dimension reduction technique on each mocap clips and compute similarities between these clips. By doing so, our similarity measurement takes the motion continuity into account. Moreover, we also greatly improve the efficiency by avoiding frame by frame similarity computation which is much more time consuming. When applying spectral clustering on the similarity matrix directly, the resulting classification labels are with serious noises. To address this problem, we propose to use median filter to remove the noises, and get good segmentation points. The automatic segmentation results on 14 motion data demonstrate the effectiveness of the proposed method.

**Key words:** motion capture data; spectral clustering; motion segmentation; principal component analysis

随着运动数据采集技术的发展, 运动数据采集的成本正在逐渐降低, 研究人员已经构建了若干规模比较大的运动数据库, 比较知名的有卡耐基梅隆大学图形实验室构建的运动数据库(CMU

graphics lab motion capture database), 以及德国马普研究所构建的运动数据库(HDM05). 市场上运动捕获设备的种类越来越丰富, 价格也逐渐降低到可接受的水平, 这使得很多小型科研机构或

收稿日期: 2015-08-15; 修回日期: 2016-02-04. 基金项目: 国家自然科学基金(61103086, 61170186); 中央高校基本科研业务费专项资金(105583GK); 全国教育科学规划课题(DCA140229). 胡晓雁(1978—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为虚拟现实、计算机图形学、计算机动画; 孙 波(1966—), 男, 博士, 教授, 博士生导师, 主要研究方向为图像处理、虚拟现实; 朱小明(1959—), 男, 硕士, 高级工程师, 主要研究方向为虚拟现实; 魏云刚(1981—), 男, 博士, 高级工程师, 主要研究方向为虚拟现实.

商业团体也能够购买运动捕获设备来采集他们所需要的数据。这种局面使得运动数据的采集比以前更容易了。而将大量采集到的运动数据分割成具体的不同运动是一项繁杂的工作, 通常需要在运动捕获完成后手动完成。研究开发运动数据分割算法将有益于减少这方面的工作量, 提高运动数据处理的效率。

运动数据分割的困难之处在于问题本身是 NP 难的。如果没有先验知识, 该运动数据包含多少个不同的运动, 以及每个运动从哪里开始到哪里结束都是未知量, 也即运动数据的每个帧都可能是分割点, 且分割点的总数预先并不知道。对于如何分割运动数据里面包含的不同运动, 以及获得这些运动的分割点而言, 一个重要的步骤在于设计合理且易于计算的运动数据相似性度量准则。在处理运动数据的相关问题中, 常使用基于帧-帧之间的相似度判断准则, 但基于帧-帧比较的相似性度量并不适合于比较 2 段动作之间的相似性, 因为运动是一段在时间上连续的帧序列, 比较运动之间的相似性需要将组成该 2 段运动的帧序列综合考虑; 而且参与比较的 2 段运动之间有可能长度是不一样的, 即包含的运动数据帧数可能是不一样的, 因而需要考虑对齐和裁剪, 这也对设计运动之间的相似性度量准则带来一定的困难。一般而言, 可以使用动态时间规整技术对时间序列进行对齐并计算相似度。但对运动分割问题而言, 在确定分割方案之前, 参与比较的运动序列本身的起始点和结束点是未知的。本文采用一种折中的策略在连贯运动的相似性度量计算和帧-帧相似度计算之间取得平衡, 即在预处理阶段将长运动数据分割成长度相等的小运动片段, 每个运动片段包含  $k$  帧, 帧数  $k$  的取值应当使得每个小运动片段包含适当的运动连贯性, 但不能太长, 大约相当于 0.1~0.2 s 的运动时长。本文实验中取  $k=15$ 。由于每个运动片段的长度相同, 因此不需要进行时间对齐预处理, 而且相似度矩阵的大小也变成原来的  $1/k^2$ , 提高了计算效率。

本文基于谱聚类算法对所获得的基于运动片段的相似度矩阵进行分割。谱聚类算法并不直接在原始数据上工作, 而是对数据集合的相似度矩阵进行分析, 得到分类结果。直接对基于运动片段进行两两相似度计算而得到的相似度矩阵是一个稠密矩阵, 而实际中, 距离较远的 2 个运动片段仍然处于同一个运动的概率是非常低的。例如, 如果

采集了一段运动数据中依次包含了走路、跳跃、跑步、走路的动作, 那么处于第一次走路和最后一次走路中的 2 个运动片段虽然相似度非常高, 但是依然需要被分割开。因此, 本文采用高斯核函数对相似度矩阵进行预处理, 根据时间上的距离来调整不同运动片段的相似度, 充分保留时间上相邻或者相近的运动片段之间的相似度值, 大幅降低时间上距离较远的 2 个运动片段之间的相似度值, 直至为 0。

本文在实验中发现, 采用多路谱聚类算法得到分割结果包含有较大的类似于椒盐噪声的噪声, 导致所得到分割结果碎片化。为解决这个问题, 通过引入中值滤波算法对分割结果进行后处理, 改善了运动分割效果。

## 1 相关工作

### 1.1 运动相似性度量策略

大多数运动数据处理的问题都涉及到采用一种合适的方法来衡量运动之间的相似性, 对运动数据进行分割也离不开运动之间相似度的计算。总体而言, 运动数据相似性度量的计算大致可分为 4 类: 基于关节角度或者关节位置的相似度计算策略, 基于主成份分析(principal component analysis, PCA)降维和投影的相似度计算策略, 基于几何特征序列对齐和检测, 基于子空间匹配和比较的相似度算法。下面简要叙述这些相似度计算策略。

1) 基于关节角度和关节位置的相似度计算。这类方法着眼于比较 2 个运动帧对应关节角向量或关节位置的欧氏距离, 用以度量帧-帧相似程度。Lee 等<sup>[1]</sup>和 Arikani 等<sup>[2]</sup>采用关节角及关节角速度的欧氏距离进行加权, 求帧-帧相似性; Kovar 等<sup>[3]</sup>使用帧窗口内的点云计算帧之间的相似度; 蓝荣祎等<sup>[4]</sup>、潘红等<sup>[5]</sup>提取关键帧, 并计算关键帧之间的相似度; Tang 等<sup>[6-7]</sup>计算关节相对位置关系, 统计相似帧数量。

2) 基于 PCA 降维和投影的相似度计算。这类方法的主要特点是利用 PCA 对运动数据降维, 再用降维后的数据计算相似度或聚类。文献[8-12]采用 PCA、加权 PCA 或概率 PCA 对运动数据降维, 然后分别对降维后的数据进行拟合, 用马氏距离计算相似度或进行聚类。Numaguchi 等<sup>[13]</sup>基于双向子空间投影方法, 将其中 2 个运动数据相互投影到对方 PCA 空间计算误差, 并综合成这 2 个运动

数据的相对误差.

3) 基于几何特征序列对齐和检测. 这类方法为运动数据生成对应的几何特征序列. 肖俊等<sup>[14]</sup>使用人体四肢骨骼到中心骨骼的角度作为特征向量, 并用其  $L_2$  范数生成一维特征序列, 依此特征序列对运动数据进行分割; Wu 等<sup>[15]</sup>基于自组织映射模型将运动片段映射到 motion string, 并用 Smith Waterman 算法计算相似度; Müller 等<sup>[16]</sup>基于二值几何特征, 对运动数据进行索引; Lin<sup>[17]</sup>基于重要骨骼的几何特征对运动曲线的尖峰模式匹配; Kovar 等<sup>[18]</sup>基于 match web 比较运动相似性.

4) 基于子空间匹配和比较的相似度算法. 这类算法也对运动数据进行 PCA, 但不执行数据的子空间投影, 而是利用 PCA 主轴方向进行比较, 计算运动数据的相似度. Yang 等<sup>[19]</sup>提出的 Eros (extended frobenius norm, Eros)方法是基于 PCA 主轴内积之加权和多维时序数据的相似性计算算法; Li 等<sup>[20-21]</sup>将权值设置为对应主轴的方差的平均值, 得到  $k$ WAS ( $k$  weighted angular similarity,  $k$ WAS)算法, 并将该方法用于运动数据的分割和识别中. 胡晓雁等<sup>[22]</sup>在  $k$ WAS 方法基础上采用四元数形式的关节角, 并使用辗转法求主轴内积之加权和, 参与相似度计算的运动数据可以保留不同数量的主轴方向.

基于子空间匹配和比较的相似度算法可以在不对数据降维的情况下直接计算运动数据的相似度, 而对运动数据进行的 PCA 可以分别在各自的预处理阶段完成. 因此在线阶段仅需要完成很少的矢量内积运算就可以完成, 简单且高效.

## 1.2 运动分割方法

在三维人体运动数据分割方面, Arikan 等<sup>[23]</sup>采用手工标注方式生成训练样本, 并学习一个支持向量机分类器, 可用于分割人体运动. Kahol 等<sup>[24]</sup>基于贝叶斯分类器, 使用层次数据结构表示和分割人体运动. Barbič 等<sup>[12]</sup>基于 PCA 和概率 PCA 对运动数据降维, 分析不同运动的内在维度变化, 进而分割运动; 他同时提出使用混合高斯模型对运动数据建模, 不同的运动属于不同的高斯模型, 从而得到运动数据的聚类分析, 据此可进行运动分割. Xiao 等<sup>[25]</sup>提出基于非线性流形学习的三维人体运动数据自动分割算法, 在低维流形上采用聚类方法分割运动数据.

朱登明等<sup>[26]</sup>采用线性时不变系统对运动数据降维, 在低维空间中定义姿态相似性度量, 并采用误差平方和准则对低维数据点集进行运动分割.

杨跃东等<sup>[10]</sup>使用加权窗计算帧-帧相似度, 进而使用谱聚类和最大值滤波法生成运动串, 通过后缀树分析提取运动串中的静态子串和周期子串, 从而实现分割. 肖俊等<sup>[14]</sup>使用人体各主要骨骼夹角作为原始运动数据的几何特征表示, 提出一种启发式方法自动检测潜在分割点, 这些潜在分割点可用于后续的交互式界面中, 以辅助用户进行精确的运动分割.

Fod 等<sup>[27]</sup>和 Jenkins 等<sup>[28]</sup>通过检测角速度的过零点来分割运动; Li 等<sup>[29]</sup>将运动数据建模为 2 层线性动态系统, 使用 EM 算法将运动分割为不同的线性动态子系统; Beaudoin 等<sup>[30]</sup>开发基于字符串 motion motif 检索和运动分割系统; Lu 等<sup>[31]</sup>提出一种双阈值多维分割算法, 将人体运动序列分解为一个简单动态线性模型序列, 适合对含周期性的运动数据进行分割; Souvenir 等<sup>[32]</sup>基于流形聚类方法对运动数据进行分割.

Zhou 等<sup>[33-34]</sup>提出对齐聚类分析(aligned cluster analysis, ACA)和层次对齐聚类分析(hierarchical aligned cluster analysis, HACA)算法, 该方法对标准的 kernel  $K$  均值法进行了两方面的扩展: 一是聚类均值包含的特征数量是可变的; 二是使用动态时间对齐核(dynamic time alignment kernel, DTAK)来达到时间序的不变性. ACA 和 HACA 在运动数据分割上达到非常好的效果, 同时这 2 种方法还可用于视频中的运动分割.

## 2 运动片段的相似度计算

### 2.1 运动片段

如前所述, 本文将运动数据等分成若干长度为  $k$  帧的运动片段, 选取合适的  $k$  值使得每个运动片段能够一定程度的保留运动连贯性. 设某待分割运动数据  $M$  共含有  $N$  帧运动数据, 等分后将得到  $\lfloor N/k \rfloor + 1$  段运动片段(最后的运动片段长度  $< k$ ).

将这些运动片段理解为长度为  $k$  的运动数据帧序列,  $M_{\text{clip}}=(m_1, m_2, m_3, \dots, m_k)$ , 每帧采用关节角度来表示人体的姿态. 与文献[12,33-34]一样, 本文选取其中对人体姿态比较重要的 14 个关节, 将其关节角度从欧拉角表示转换成四元数表示, 转换后每一帧含有 56 个自由度. 每个运动片段  $M_{\text{clip}}$  含有  $k$  帧, 则运动片段可以表示为  $56 \times k$  矩阵:  $X_{56 \times k}=(x_1, x_2, \dots, x_k)$ , 其中  $x_i$  ( $i=1, 2, \dots, k$ )是 56 维列向量, 表示第  $i$  帧  $m_i$  的姿态数据.

## 2.2 运动片段相似性计算

本文将一段运动片段视为一个由  $k$  个高维空间中的点组成的集合,  $X_{56 \times k} = \{x_1, x_2, \dots, x_k\}$ , 其中点  $x_i (i=1, 2, \dots, n)$  是 56 维的向量. 这样, 计算 2 个运动片段相似性, 可以视为考虑 2 个高维空间中点集的相似性问题. 如图 1 中 2 个二维平面的点集  $A$  和  $B$ , 绘制出这 2 个点集的主方向用虚线箭头表示, 箭头的长度表示点集在该主方向上的投影方差; 如果这 2 个点集非常相似的话, 那么它们各自的主方向朝向和对应的方差应该非常接近. 因此, 点集的整体相似性计算分为 2 步:

Step1. 计算点集的主方向(主轴)

Step2. 计算对应主方向的近似程度, 通过计算对应主轴的内积并依据该主轴方向所对应的数据方差进行加权来得到相似度.

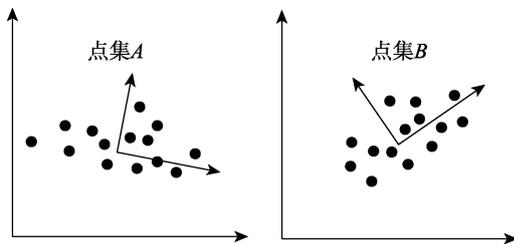


图 1 平面点集的主方向示意图

### 2.2.1 PCA 分解

为了得到每个运动片段的主轴信息, 对运动片段数据进行 PCA 处理

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

$$C = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T \quad (2)$$

$$C = Q\Lambda Q^T \quad (3)$$

式(1)~(3)中,  $\mu$  是运动数据的样本均值;  $C$  是  $56 \times 56$  的运动数据协方差矩阵; 式(3)对  $C$  进行特征值分解, 得到特征向量矩阵  $Q$ ,  $Q$  是一个正交矩阵, 其中的每一列为  $C$  的一个特征向量; 对角矩阵  $\Lambda$  中保存着  $Q$  中特征向量对应的特征值. 对运动数据的特征值进行排序, 得到  $\sigma_1 \sigma_2 \dots \sigma_{56} 0$ , 其对应的特征向量分别为  $(q_1, q_2, \dots, q_{56})$ , 观察  $q_i (i=1, 2, \dots, 56)$ , 它是一个 56 维的向量, 与运动数据具有相同的维度;  $q_i$  的每个分量都与原始运动数据的一个自由度相对应, 分量的值则表示该自由度对主方向  $q_i$  的贡献, 这可以解释为该运动中主要是哪些关节在起作用, 以及这些关节运动对该运动的重要程度.

### 2.2.2 基于 PCA 的运动片段相似性计算

对于 2 个运动数据片段  $M_a$  和  $M_b$ , 将数据组织成矩阵形式  $X_a$  和  $X_b$ , 对二者作上述 PCA, 得到特征向量矩阵  $Q_a$  和  $Q_b$ , 以及特征值矩阵  $\Lambda_a$  和  $\Lambda_b$ ; 对特征值从大到小进行排序, 并将特征向量做相应排序, 设对于  $X_a$ , 有特征值  $\sigma_1 \sigma_2 \dots \sigma_{56} 0$  及特征向量  $(p_1, p_2, \dots, p_{56})$ ; 对  $X_b$ , 有特征值  $\rho_1 \rho_2 \dots \rho_{56} 0$  及特征向量  $(q_1, q_2, \dots, q_{56})$ . 其中特征向量  $p_i, q_i (i=1, 2, \dots, 56)$  都是单位向量, 如果  $M_a$  和  $M_b$  非常相似, 则它们的各主轴应接近重合, 即对应的主轴之间夹角非常小. 本文使用矢量内积来定义 2 个矢量  $p$  和  $q$  夹角余弦值

$$\cos \theta = \frac{p \cdot q}{\|p\| \|q\|}$$

Safonova 等<sup>[35]</sup>指出, 人体运动数据可以使用 5~10 个主轴方向近似表示, 且误差非常小. 因此可以只使用前若干个主轴方向来计算 2 个运动数据的整体相似性, 本文取前若干个主方向, 保证投影数据误差在 10% 以内; 另一方面, 以投影数据方差来衡量运动数据沿某主轴方向投影所保留的原始数据的量, 这能反映每个主轴方向在对高维数据降维过程中的重要程度, 可以将归一化的方差作为每个主轴的权重应用到相似性计算上去. 但是, 对于 2 个运动数据  $M_a$  和  $M_b$ , 其对应的主方向  $p_s, q_t$  具有不同的归一化方差  $\hat{\sigma}_s = \sigma_s / \sum_{i=1}^{56} \sigma_i, \hat{\rho}_t = \rho_t / \sum_{i=1}^{56} \rho_i$ . 对此, 本文考虑按照权值将主方向分配给对应主方向计算内积. 例如, 设  $\hat{\sigma}_1 = 0.8, \hat{\rho}_1 = 0.5, \hat{\rho}_2 = 0.3$ , 则有  $0.8p_1 = 0.5p_1 + 0.3p_1$ , 由此可以计算  $M_a$  和  $M_b$  的整体相似度为  $0.5 \times |p_1 \cdot q_1| + 0.3 \times |p_1 \cdot q_2|$ . 基于上述讨论, 本文设计了一种类似辗转相除法的策略来计算加权整体相似度.

算法 1. 计算运动数据  $M_a$  和  $M_b$  的整体相似度  $s$

输入. 运动数据  $M_a$  的前  $r$  个主方向  $p_1, p_2, \dots, p_r$  和归一化方差  $\hat{\sigma}_1, \hat{\sigma}_2, \dots, \hat{\sigma}_r$ , 运动数据  $M_b$  的前  $l$  个主方向  $q_1, q_2, \dots, q_l$  和归一化方差  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_l$ .

输出. 运动数据  $M_a$  和  $M_b$  的整体相似度  $s$ .

初始化.  $s=0, i=j=1$ .

Step1. while  $i \leq r$  and  $j \leq l$

if  $\hat{\sigma}_i > \hat{\rho}_j$

$$s = s + \hat{\rho}_j |p_i \cdot q_j|; \quad \hat{\sigma}_i = \hat{\sigma}_i - \hat{\rho}_j;$$

$j = j + 1;$

else if  $\hat{\sigma}_i < \hat{\rho}_j$

$$s = s + \hat{\sigma}_i |p_i \cdot q_j|; \quad \hat{\rho}_j = \hat{\rho}_j - \hat{\sigma}_i;$$

```

    i = i+1;
else
    s = s + \hat{\sigma}_i |p_i \cdot q_j|;
    i = i+1; j = j+1;
end
end
Step2. while j < l
    s = s + \hat{\rho}_j; j = j+1;
end
Step3. while i < r
    s = s + \hat{\sigma}_i; i = i+1;
end

```

### 3 谱聚类运动分割算法

#### 3.1 谱聚类算法

谱聚类算法是近年来被广泛关注的一类聚类算法,它基于图谱理论,根据节点间的相似度关系来完成对节点的聚类或分割.设数据点之间相似度矩阵为  $W$ ,则  $W$  为对称矩阵.定义拉普拉斯矩阵  $L=D-W$ ,其中  $D$  是对角矩阵,其对角线上每一项  $d_i$  都是  $W$  对应对应的和,即

$$d_i = \sum_{j=1}^n w_{ij}.$$

$L$  具有如下重要性质:首先  $L$  是对称且半正定的;其次  $L$  的所有特征值均大于等于 0,且最小的特征值为 0,零特征值对应分量全为 1 的特征向量. $L$  的第二小特征值所对应的特征向量一般称为 Fiedler 向量.对于所有的  $n$  维实向量  $f$ ,有

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2.$$

设  $f$  是特征值 0 的一个特征向量,则有

$$0 = f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2.$$

假定相似度矩阵中的相似度值  $w_{ij} > 0$ ,对于 2 个节点  $i, j$ ,若其相似度  $w_{ij} > 0$ ,则其对应的特征向量分量相等,即  $f_i = f_j$ .更进一步,如果 2 个节点之间有一条路径可达,且路径上相邻节点之间的相似度都大于 0,则这条路径上的节点所对应的特征向量分量都相等,这表明相似度矩阵所对应的图结构中,同属一个连通子图的节点所对应的特征向量分量相等<sup>[36]</sup>.这个结论构成了谱聚类算法的重要理论依据.拉普拉斯矩阵特征值 0 的重数对应了连通子图数目<sup>[36]</sup>.而这些连通子图则构成对数据的聚类,也就完成了对数据的分割.

#### 3.2 相似度矩阵构建

根据上述谱聚类算法,对数据施行谱聚类算

法,最重要的是要构建合适的相似度矩阵.对于谱聚类而言,理想的相似度矩阵是分块对角矩阵.但在一般应用中很少能够直接获得这样的相似度矩阵,需要采取一些改进策略使得相似度矩阵能够尽可能接近这种形式,以使得聚类结果更稳定.本文对之前的相似度矩阵做两方面的改进:

1) 使用高斯核对其进行处理.由于运动数据是一个时间序列,对于分割运动数据的任务来说,可以抑制时间上相距较远的 2 个运动片段之间的相似度(有可能出现时间上不相邻但相似度很高的运动片段,例如,一长段运动的顺序是“走,跳,走”,但对运动分割任务而言,抑制 2 个“走”之间的相似度并不影响将它们与“跳”动作分割开来).

本文使用一个高斯核  $g_{ij} = e^{-\frac{\|i-j\|^2}{2\sigma^2}}$  来完成这个抑制过程.

2) 使用经典的  $K$  近邻构图法来构造相似度矩阵.对于相似度矩阵  $W$  的每一行,保留最大的  $K$  个元素,而将剩余的元素设为 0.

#### 3.3 谱聚类运动分割

本节给出本文提出的基于谱聚类的运动分割算法的完整流程.该算法要求提供运动数据  $M$  中包含的不同运动的个数  $m$ ,而这其实是未知的.在本文采用的实验数据集中,各运动数据包含 6~12 个运动,其中某些运动是同一类型的运动.因此各运动数据包含 10 种左右的不同类型运动(或者更少).本文设置  $m=10$ .

##### 算法 2. 基于谱聚类的运动分割算法

输入. 运动数据  $M$ (含有  $N$  帧),运动片段长度  $k$ ,高斯核参数  $\sigma$ ,近邻数  $K$ , $M$  包含的运动个数  $m$ .

输出. 分割点  $c_1, c_2, \dots, c_{m-1}$ .

Step1. 将运动数据切成  $n = \lfloor N/k \rfloor + 1$  个运动片段.

Step2. 计算运动片段之间的相似度,乘以高斯核.

Step3. 基于 Step2 的相似度矩阵构建  $K$  近邻图.

Step4. 构建拉普拉斯矩阵  $L = D - W$ .

Step5. 计算  $L$  的前  $m$  个特征向量  $u_1, u_2, \dots, u_m$ (对应于最小的前  $m$  个特征值),以  $u_1, u_2, \dots, u_m$  为列向量组成矩阵  $U$ .

Step6. 令  $y_i (i=1, \dots, n)$  为  $U$  的第  $i$  行向量的转置.

Step7. 对  $y_i (i=1, \dots, n)$  使用  $K$  均值法进行聚类,得到  $m$  个类.

#### 3.4 统计滤波后处理

##### 3.4.1 自适应中值滤波

直接应用算法 2 进行运动分割得到的结果会产生强烈的噪声,这种噪声的性状类似于图像

处理中的椒盐噪声, 可采用中值滤波算法来消除这种噪声. 产生噪声的原因来自于运动片段相似度矩阵的计算, 例如, 一个走路动作中包含“迈左腿”和“迈右腿”的运动片段, 这 2 个片段的相似度较低, 因此走路运动中各运动片段的相似度呈现高低交错的特点, 使得谱聚类算法将这些运动片段交替地归为不同的类, 产生了类似噪声的现象. 因此, 本文使用自适应中值滤波算法对结果进行过滤, 能得到较为干净的结果. 本文设置最大窗口为 7.

#### 3.4.2 “最多出现”滤波

对于部分无法用中值滤波消除的噪声信号, 进一步采用最多出现滤波进行处理. 对于被处理的当前信号, 统计以该信号为中心的窗口内各类别信息出现次数, 并选择出现次数最多的值赋予当前信号. 经过这步处理后, 本文将获得较连续的无噪声的分类信息, 基本可直接用于运动分割.

#### 3.4.3 合并较短的分割段

对于可能出现的较短的分割片段(1~2 s), 将其合并到相邻的分割片段中. 例如, 某分割片段  $s_i$  长度小于设定的阈值, 则将其合并到相邻的分割片段  $s_{i-1}$  或者  $s_{i+1}$  中去. 若  $s_{i-1}$  的长度小于  $s_{i+1}$ , 则将  $s_i$  合并到  $s_{i-1}$  中; 反之, 则将  $s_i$  合并到  $s_{i+1}$  中.

## 4 实验及结果分析

### 4.1 数据集

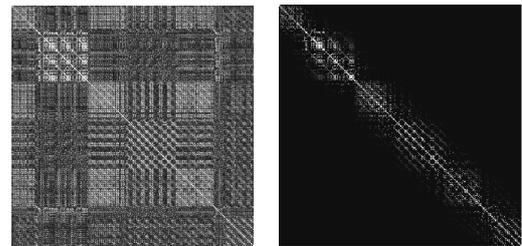
本文采用文献[12]中的数据集进行测试, 该数据集来自 CMU Mocap database 86 号对象动作采集组(Subject 86)的 14 套运动数据. 这些运动数据包含数千帧运动数据, 采集频率为 120 帧/s, 持续时间为 30~100s 不等. 数据集的真实分割点(Ground Truth)由文献[12]给出. 这些真实分割点是由人工观察的方式给出的, 每个真实分割点分别给出了 3 项数据, 分别是最早分割点、最晚分割点和最可能分割点. 例如, 运动数据 9 的第一个真实分割点的 3 个估计分别是 880 帧, 1 020 帧和 1 045 帧.

### 4.2 实验过程

#### 4.2.1 相似度计算

以运动数据 14 为例来说明本文算法运行过程及结果. 对于每个待分割的运动数据, 按算法 1 计算相似度矩阵. 相似度矩阵结果如图 2 所示.

图 2 中, 较黑的颜色表示相似度较低, 而白色表示相似度高. 对角线上的块状子矩阵就代表着运动数据中包含的各个运动. 从图 2a 可以看



a. 相似度矩阵 b. 高斯核及  $K$  近邻图结果

图 2 运动数据 14 的相似度矩阵

出, 对于分割, 比较有用的信息集中在对角线带上, 因此依照第 3.2 节中的方法对上述相似度矩阵使用高斯核及  $K$  近邻图构建法进行处理, 去除不重要的信息, 留下起主要作用的相似度数据, 这样可大幅减少后续计算步骤的数据量. 相似度矩阵的处理结果如图 2b 所示.

#### 4.2.2 谱聚类分割和统计滤波后处理

预处理后的相似度矩阵由算法 2 进行处理. 运动数据 14 的谱聚类结果如图 3 所示. 可以看出, 谱聚类给出的分类信息在大尺度上揭示了运动数据 14 中包含的若干不同运动及其大致的分割点位置, 表明谱聚类算法能够有效地提取出运动数据分割信息. 但算法 2 的谱聚类分类信息颗粒度较细(处理对象为 15 帧的运动片段), 因此相邻运动片段的聚类结果跳跃性很大并呈现出一定的周期性, 这些周期性来自于运动数据中各动作本身的周期性. 例如, 运动数据 14 包含若干段篮球运动中的运球动作和行走动作, 每段运球动作都包含着若干次拍球动作, 而行走动作也呈现周期性. 同时, 图 3 中分类编号的绘图结果在视觉上类似于图像处理中的椒盐噪声, 这提示我们可以采用自适应中值滤波技术对结果进行去噪声处理(见第 3.4.1 节). 图 4 所示为运动数据 14 的自适应中值滤波处理结果.

经过自适应中值滤波后的分类信息形成了明显连续的类别信息, 可用于对运动数据进行分割, 但仍然存在部分不连续的突变信号, 从视觉上, 这些信号很大程度仍然是因为运动周期性引起. 此时使用中值滤波已经无法消除这些突变, 改用最多出现

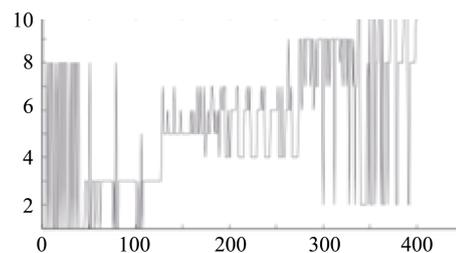


图 3 运动数据 14 谱聚类结果

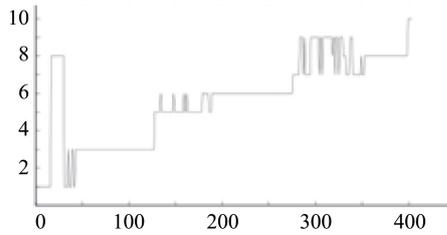


图 4 对运动数据 14 聚类结果使用中值滤波

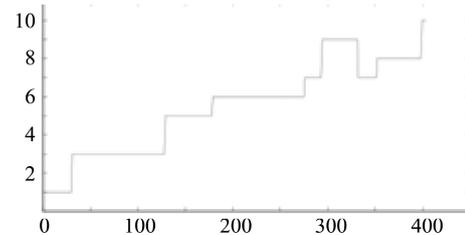


图 5 对运动数据 14 继续使用“最多出现”滤波处理

滤波,即统计窗口内出现次数最多的类别编号,并以此编号赋值给当前处理的类别量(见第 3.4.2 节).经过这一步骤的处理,可以得到比较干净的分割信息.最终依照此信息对原运动数据进行分割,如图 5 所示.

对于某些较短的分割片段(本文中选择长为 20 帧以内的运动片段,即长度为  $20 \times 15=300$  帧以内),将其合并到 2 个相邻片段中较短的一段中去(见第 3.4.3 节).

### 4.3 实验结果

本文对数据集中的 14 套运动数据进行了实验,并将分割点与文献[12]提供的 Ground Truth 进行了对比,如图 6 所示.图 6 中共有 14 个分割点分布图;每个分布图包含上下两部分,上部分为本文算法的实验结果,下部分是 Ground Truth 提供的最可能分割点.实验中,我们认为在最早分割点和最晚分割点之间的值都为可接受的值.

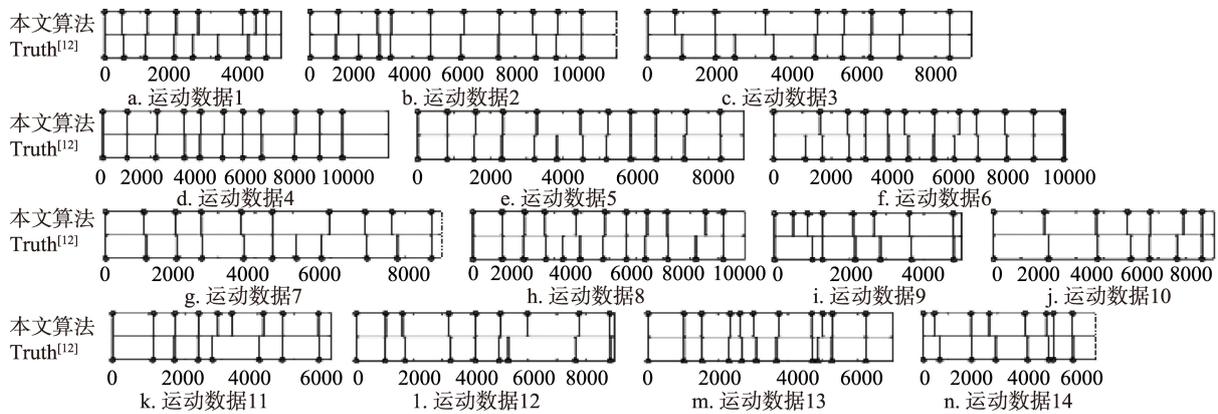


图 6 实验数据集 14 个运动数据的分割结果

从图 6 可以看出,本文算法很好地找到了绝大多数的分割点,且实验结果与 Ground Truth 的位置重合或者非常接近.当然,本文算法有时候会漏检若干分割点,如运动数据 2(如图 6b 所示)的第 2 个分割点等.统计本文结果相对于 Ground Truth 的准确程度,主要包含两方面:一是分割点是否落入真实分割点区间(最早分割点及最晚分割点之间),若是,则认为误差为 0;否则,误差为其距离相应真实分割点区间的距离(帧数).将所有找到的分割点的误差求和,除以该运动数据的总帧数,并记为  $a$ .二是计算漏检的真实分割点,误差为该漏检的真实分割点距离最近分割点的距离(帧数),将所有漏检的真实分割点误差求和,除以该运动数据的总帧数,记为  $e$ .则准确度

$$A = 1.0 - a - e.$$

计算对于该套数据集共 14 个运动数据,本

文算法分割后的平均准确度信息如表 1 所示.图 7 所示为各运动数据分割的准确度.

表 1 算法的平均精确度信息

精度及误差	平均值/%
$A$	91.17
$a$	4.92
$e$	3.91

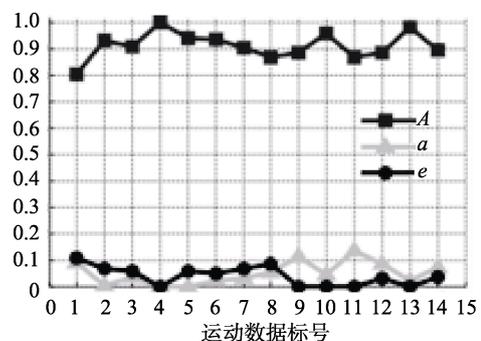


图 7 在实验数据集上的分割精确度

### 4.4 运行效率分析

本文实验在 ThinkPad X220, i5CPU@2.4GHz, 4GB 内存, Windows XP 系统, Matlab 2012a 环境下进行, 算法的运行时间如图 8 所示。

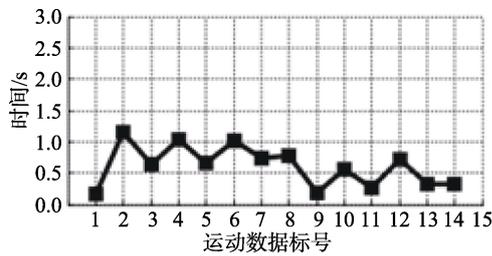


图 8 实验数据集上运行时间

### 4.5 实验结果对比

本文算法的分割结果与 ACA 方法<sup>[33]</sup>和 HACA

方法<sup>[34]</sup>的结果比较如图 9 所示. 其中依次列出了每个运动数据的分割结果对比: 第 1 行为本文算法的分割结果; 第 2 行为文献[34]给出的 Ground Truth; 第 3 行为采用 ACA 算法得到的分割结果; 第 4 行为采用 HACA 算法得到的分割结果. 可以看出, 本文算法获得了与 HACA 及 ACA 相近的分割效果. 在运动数据 4, 9, 13, 14 上的表现要优于这二者. 例如, 在运动数据 14 中, 本文算法给出了与 Ground Truth 相符的倒数第 2 个运动的分割效果, ACA 及 HACA 则没有获得相应的分割效果.

同时我们发现, HACA 方法<sup>[34]</sup>给出的 Ground Truth 有 3 处与文献[12]的不同, 分别发生在运动数据 5, 9, 13. 从图 10(Truth<sup>[12]</sup>用红线条标出的分割点)可以看出, 本文算法成功地将文献[12]给出的分割点(红色线条处)识别出来.

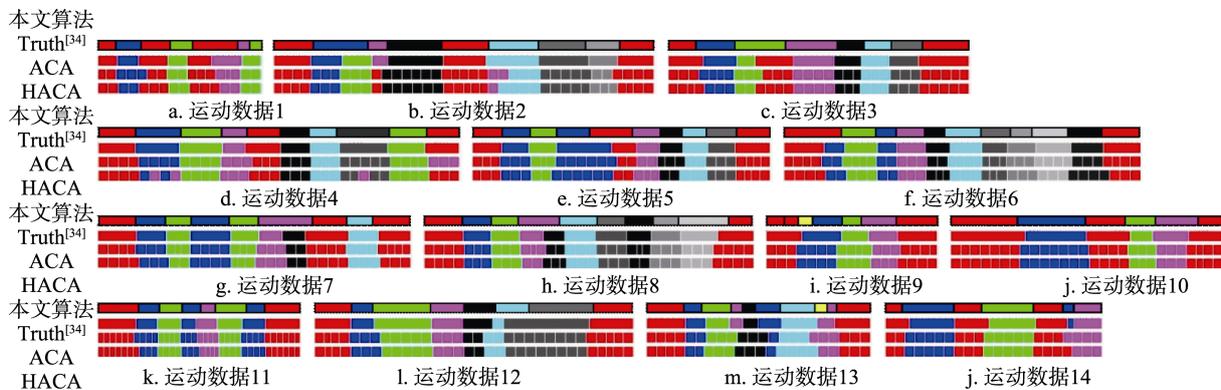


图 9 与文献[33-34]的实验对比图

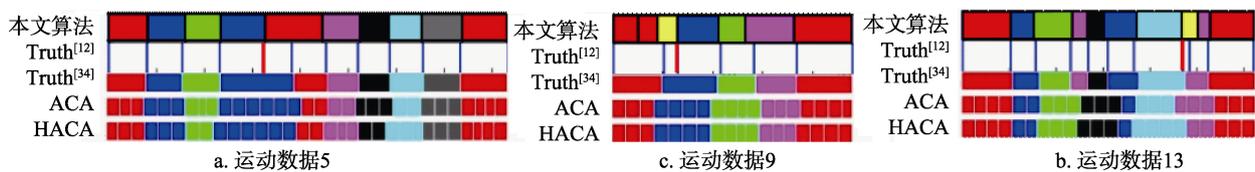


图 10 文献[12,34]在运动数据上的分割点差异

## 5 讨论与进一步工作

本文提出一种简单有效的运动数据分割算法, 该算法基于谱聚类算法, 对运动数据的相似度矩阵进行分割. 不同于已有的主要的运动数据分割方法<sup>[12,33-34]</sup>, 本文采用基于运动片段的相似度计算, 以保留运动的连续特征, 能更好地表达运动之间的相似特征; 同时避免单帧相似度计算耗时, 提高了计算效率, 并取得了较好的分割正确率. 运动片段应该包含运动的连续内容, 因此其长度不应太短; 同时运动片段也不应过长, 以避免一个片段

中包含不同的运动. CMU 运动数据库中的运动是以 120 帧/s 进行采集的, 因此本文将运动片段长度设为  $k = 15$ , 时长约为 0.125 s.

本文基于标准谱聚类算法对运动数据进行分割, 而谱聚类算法只需要提供数据的相似度矩阵即可; 本文基于 PCA 方法计算运动片段的相似度, 虽然可以很好地保留运动片段的主轴信息, 但同时也丢失了运动片段的时序列信息, 这导致在实验结果(如图 9 所示)中出现了一些分割失败的运动片段, 如运动数据 7, 8, 12 处; 而且 PCA 方法计算需要消耗一定的计算量. 我们将在下一步的工作

中研究能够结合运动片段时序特征的相似度量算法。另外,本文工作仅考虑了运动数据分割问题,如何对分割结果进行识别和分类,也将是下一步研究的问题之一。

在第 4.2.2 节对谱聚类结果进行了自适应中值滤波处理,出发点是聚类结果绘图在视觉上呈现出椒盐噪声的特点,但实际上出现这种情况更有可能是由于被聚类运动自身的周期性引起的。如果从这个方面出发,也许分析聚类结果的周期性是一个更好的选择,例如采用傅里叶分析来提取信号的频率信息,并进行滤波。也可考虑使用小波变换来提取信号的不同尺度的信息来进行运动分割。

由于谱聚类算法只要求数据的相似度矩阵,因此其使用范围非常广。图 2 中绘制的运动数据相似度矩阵在视觉上类似于纹理图像,其谱聚类结果带有强烈噪声特征,本文采用统计滤波方法进行了后处理,取得了较好的效果。今后也将探索将本文中对谱聚类的统计滤波后处理方法应用到具有类似相似度矩阵结构的其他数据分割问题中去。

本文实验均在 Matlab 软件中完成,对于具有 10 616 帧约 89 s 的运动数据大约需要 1.1 s 处理时间,基本能够达到实时。我们也将考虑实效率更高的 C++ 实现版本,并在相似度计算及谱聚类 2 个环节进行优化,以进一步提高算法的效率。

致谢:感谢周晓巍博士对本文工作提出的建设性建议!感谢谢顺波在本文修改过程中的大量帮助和建议!

## 参考文献(References):

- [1] Lee J, Chai J X, Reitsma P S A, *et al.* Interactive control of avatars animated with human motion data[J] *ACM Transactions on Graphics*, 2002, 21(3): 491-500
- [2] Arikan O, Forsyth D A. Interactive motion generation from examples[J] *ACM Transactions on Graphics*, 2002, 21(3): 483-490
- [3] Kovar L, Gleicher M, Pighin F. Motion graphs[J] *ACM Transactions on Graphics*, 2002, 21(3): 473-482
- [4] Lan Rongyi, Sun Huaijiang, Lian Heqing, *et al.* Vector space modeling and retrieval of human motion capture data[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2011, 23(8): 1357-1364(in Chinese)  
(蓝荣祎, 孙怀江, 连荷清, 等. 人体运动捕获数据的向量空间建模与检索[J]. *计算机辅助设计与图形学学报*, 2011, 23(8): 1357-1364)
- [5] Pan Hong, Xiao Jun, Wu Fei, *et al.* 3D human motion retrieval based on key-frames[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2009, 21(2): 214-222(in Chinese)  
(潘红, 肖俊, 吴飞, 等. 基于关键帧的三维人体运动检索[J]. *计算机辅助设计与图形学学报*, 2009, 21(2): 214-222)
- [6] Tang J K T, Leung H, Komura T, *et al.* Emulating human perception of motion similarity[J]. *Computer Animation & Virtual Worlds*, 2008, 19(3/4): 211-221
- [7] Tang J K T, Leung H. Retrieval of logically relevant 3D human motions by adaptive feature selection with graded relevance feedback[J]. *Pattern Recognition Letters*, 2012, 33(4): 420-430
- [8] Li Chunpeng, Wang Zhaoqi, Xia Shihong. Motion synthesis for virtual human using functional data analysis[J]. *Journal of Software*, 2009, 20(6): 1664-1672(in Chinese)  
(李淳鹏, 王兆其, 夏时洪. 人体运动的函数数据分析与合[J]. *软件学报*, 2009, 20(6): 1664-1672)
- [9] Zhu Dengming, Wang Zhaoqi. Human animation synthesis based on primitive movement[J]. *Journal of Computer Research & Development*, 2009, 46(4): 610-617(in Chinese)  
(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)
- [10] Yang Yuedong, Wang Lili, Hao Aimin. Motion string: a motion capture data representation for behavior segmentation[J]. *Journal of Computer Research & Development*, 2008, 45(3): 527-534(in Chinese)  
(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)
- [11] Forbes K, Fiume E. An efficient search algorithm for motion data using weighted PCA[C] //Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Aire-la-Ville: Eurographics Association Press, 2005: 67-76
- [12] Barbič J, Safonova A, Pan J Y, *et al.* Segmenting motion capture data into distinct behaviors [C]//Proceedings of Graphics Interface 2004. Waterloo: Human-Computer Communications Society Press, 2004: 185-194
- [13] Numaguchi N, Nakazawa A, Shiratori T, *et al.* A puppet interface for retrieval of motion capture data[C] //Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Aire-la-Ville: Eurographics Association Press, 2011: 157-166
- [14] Xiao Jun, Zhuang Yueting, Wu Fei. Feature visualization and interactive segmentation of 3D human motion[J]. *Journal of Software*, 2008, 19(8): 1995-2003(in Chinese)  
(肖俊, 庄越挺, 吴飞. 三维人体运动特征可视化与交互式运动分割[J]. *软件学报*, 2008, 19(8): 1995-2003)
- [15] Wu S Y, Xia S H, Wang Z Q, *et al.* Efficient motion data indexing and retrieval with local similarity measure of motion strings[J]. *The Visual Computer*, 2009, 25(5-7): 499-508
- [16] Müller M, Röder T, Clausen M. Efficient content-based retrieval of motion capture data[J] *ACM Transactions on Graphics*, 2005, 24(3): 677-685
- [17] Lin Y. Efficient motion search in large motion capture databases[M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2006, 4291: 151-160
- [18] Kovar L, Gleicher M. Automated extraction and parameterization of motions in large data sets[J] *ACM Transactions on Graphics*, 2004, 23(3): 559-568
- [19] Yang K, Shahabi C. A PCA-based similarity measure for multivariate time series[C] //Proceedings of the 2nd ACM

- International Workshop on Multimedia Databases. New York: ACM Press, 2004: 65-74
- [20] Li C J, Prabhakaran B. A similarity measure for motion stream segmentation and recognition[C] //Proceedings of the 6th International Workshop on Multimedia Data Mining: Mining Integrated Media and Complex Data. New York: ACM Press, 2005: 89-94
- [21] Li C J, Prabhakaran B. Indexing of motion capture data for efficient and fast similarity search[J]. Journal of Computers, 2006, 1(3): 35-42
- [22] Hu Xiaoyan, Sun Bo, Cao Le, *et al.* Similarity computation of motion data based on principle component analysis[OL]. [2015-08-15]. <http://d.wanfangdata.com.cn/Conference/775440961-68> (in Chinese)  
(胡晓雁, 孙波, 曹乐, 等. 基于 PCA 的运动数据相似性计算[OL]. [2015-08-15]. <http://d.wanfangdata.com.cn/Conference/775440961-68> )
- [23] Arikan O, Forsyth D A, O'Brien J F. Motion synthesis from annotations[J]. ACM Transactions on Graphics, 2003, 22(3): 402-408
- [24] Kahol K, Tripathi P, Panchanathan S. Automated gesture segmentation from dance sequences[C] //Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2004: 883-888
- [25] Xiao J, Zhuang Y, Wu F. Getting distinct movements from motion capture data[C] //Proceedings of the International Conference on Computer Animation and Social Agents. Geneva: Wiley, 2006: 33-42
- [26] Zhu Dengming, Wang Zhaoqi. Extraction of keyframe from motion capture data based on motion sequence segmentation[J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(6): 787-792(in Chinese)  
(朱登明, 王兆其. 基于运动序列分割的运动捕获数据关键帧提取[J]. 计算机辅助设计与图形学学报, 2008, 20(6): 787-792)
- [27] Fod A, Mataric M J, Jenkins O C. Automated derivation of primitives for movement classification[J]. Autonomous Robots, 2002, 12(1): 39-54
- [28] Jenkins O C, Mataric M J. Deriving action and behavior primitives from human motion data[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and System. Los Alamitos: IEEE Computer Society Press, 2002, 3: 2551-2556
- [29] Li Y, Wang T S, Shum H Y. Motion texture: a two-level statistical model for character motion synthesis[J]. ACM Transactions on Graphics, 2002, 21(3): 465-472
- [30] Beaudoin P, Coros S, van de Panne M, *et al.* Motion-motif graphs[C] //Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. Aire-la-Ville: Eurographics Association Press, 2008: 117-126
- [31] Lu C M, Ferrier N J. Repetitive motion analysis: segmentation and event classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(2): 258-263
- [32] Souvenir R, Pless R. Manifold clustering[C] //Proceedings of the 10th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2005, 1: 648-653
- [33] Zhou F, De la Torre F, Hodgins J K. Aligned cluster analysis for temporal segmentation of human motion[C] //Proceedings of IEEE International Conference on Automatic Face & Gesture Recognition. Los Alamitos: IEEE Computer Society Press, 2008: 1-7
- [34] Zhou F, De la Torre F, Hodgins J K. Hierarchical aligned cluster analysis for temporal clustering of human motion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 582-596
- [35] Safonova A, Hodgins J K, Pollard N S. Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces[J]. ACM Transactions on Graphics, 2004, 23(3): 514-521
- [36] Luxburg U. A tutorial on spectral clustering[J]. Statistics and Computing, 2007, 17(4): 395-416