姿势序列有限状态机动作识别方法

林水强[1)]，吴亚东[1,2)*]，余 芳[1)]，杨永华[1)]

[1)](西南科技大学计算机科学与技术学院 绵阳 621010)

[2)](西南科技大学 核废物与环境安全国防重点学科实验室 绵阳 621010)

(wuyadong@swust.edu.cn)

摘 要：肢体动作分析与识别是实现体感交互的重要前提.为提高用户自然动作识别的效率与通用性，提出姿势序列有限状态机方法.首先，以用户为中心建立肢体节点坐标系，将描述用户动作的肢体节点数据从设备空间变换到用户空间，并建立三维网格划分模型，以尽可能消除用户个体差异；其次，在肢体节点坐标系定义肢体节点特征向量，借助关节点空间运动矢量、关节点运动时间间隔、关节点空间距离描述肢体动作特征，对预定义肢体动作序列进行采样分析；最后，采用关节运动正则表达式表示肢体动作轨迹，构造姿势序列有限状态机，实现对预定义动作的在线识别.针对 17 种预定义动作的实验结果表明，文中方法识别率高，具有良好的扩展性和通用性，能够满足体感交互应用需求.

关键词：动作识别；姿势序列；有限状态机；正则表达式；Kinect 传感器

中图法分类号：TP391

Posture Sequence Finite-State Machine Method for Motion Recognition

Lin Shuiqiang[1)], Wu Yadong[1,2)*], Yu Fang[1)], and Yang Yonghua[1)]

[1)](*School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang* 621010)

[2)](*Fundamental Science on Nuclear Wastes and Environmental Safety Laboratory, Southwest University of Science and Technology, Mianyang* 621010)

Abstract: Limb motion analysis and recognition is critical for somatosensory interactions. In this paper, a new posture sequence finite-state machine (FSM) method is proposed for improving the efficiency and versatility of natural motion recognition. Firstly, a user-centric limb joint coordinate system is setup to transform the skeletal data of the user motions from the device space to the user space, and a three-dimensional grid model is built to eliminate the users' individual differences. Secondly, the joint motion sequences of predefined motions are sampled and analyzed using the limb joint feature vectors, which are defined in the limb joint coordinate system, with spatial motion vectors, motion time intervals, and spatial distances of joints being employed to depict the limb motion features. Finally, the limb motion trajectory is represented by the joint movement regular expressions, and the posture sequence FSM is constructed to recognize the predefined motions online. Experimental results on 17 kinds of predefined motions show that the proposed method has the merits of high recognition rate, good scalability and versatility, and is suitable for somatosensory interaction applications.

Key words: motion recognition; posture sequence; finite-state machine (FSM); regular expression; Kinect sensor

---

收稿日期：2013-08-18；修回日期：2014-03-20. **基金项目**：国家自然科学基金(61303127)；国防基础研究项目(10ZG6102，13ZXNK12)；四川省科技厅项目(2011JQ0041，11ZS2009)；中国科学院"西部之光"人才培养计划(13ZS0106)；四川省教育厅重点项目(11ZA130，13ZA0169)；绵阳市网络融合实验室(12zxwk05).**林水强**(1988—)，男，硕士研究生，CCF 学生会员，主要研究方向为虚拟现实、人机交互；**吴亚东**(1979—)，男，博士，教授，CCF 会员，硕士生导师，论文通讯作者，主要研究方向为图像图形处理、可视化、人机交互等；**余 芳**(1992—)，女，硕士研究生，主要研究方向为人机交互；**杨永华**(1989—)，女，硕士研究生，主要研究方向为设计心理学、用户评估.

在人机交互领域，动作识别是体感交互的前提，动作识别和行为理解逐渐成为人机交互领域的研究热点[1-3]. 为达到有效的交互目的，必须对不同的交互动作，包含肢体运动、手势以及静态姿势进行定义和识别[4]. 近年来，基于 Kinect 体感技术的动作识别应用开发十分丰富，但在这些应用中，虽然能够有效跟踪人体运动轨迹[5-7]，但识别动作比较单一且识别方法不利于扩展[8-10]，亟待研究和开发一种具有扩展性和通用性的动作识别模型.

目前，基于 Kinect 的动作识别方法较多，如事件触发、模板匹配、机器学习等方法. 文献[11]提到的灵活动作和铰接式骨架工具包(flexible action and articulated skeleton toolkit, FAAST)是介于 Kinect 开发包和应用程序之间的体感操作中间件，其主要采用事件触发方式进行识别，如角度、距离、速度等事件. 该方法计算量小、实时性和准确率高，但事件触发方法本身具有局限性，对连续动作的识别较困难. Ellis 等[12]探讨了行为识别准确率与延迟之间的权衡，从动作数据序列中确定关键帧实例，从而推导出动作模板，但动作模板仅保存同一类行为的形态和模型，忽略了变化. Wang 等[13]给出了一种动作子集组合的方法，对关节点子集进行分类，识别率高；但其着重于事先分割的数据流级别，不能用于从未分割数据流中进行在线识别. Zhao 等[14]提出了一种结构化流骨骼(structured streaming skeletons, SSS)特征匹配的方法，通过离线训练建立一个特征字典和手势模型，为未知动作的动作数据流的每一帧数据分配标签. 通过提取 SSS 特征在线预测动作类型，该方法能够有效地解决错误分割和模板匹配不足的问题，从未分割数据流中进行在线识别；但其计算复杂，识别反馈时间不稳定，且对每个动作识别需要特征字典库，对于扩展动作类型识别时，需要收集大量动作数据进行离线训练，特定动作识别与训练集耦合度较高.

为了避免以往肢体动作识别方法不易扩展和识别效率低等不足，本文提出一种姿势序列有限状态机(finite state machine, FSM)动作识别方法. 特定的肢体动作可以看作由多个姿势在时间轴上的一组运动序列，即姿势序列来描述. 本文方法主要采用肢体节点特征向量描述肢体动作特征数据，通过对预定义的肢体动作序列进行采样分析，建立肢体动作的轨迹正则表达式，通过轨迹正则表达式构造姿势序列 FSM，从而实现对肢体动作的分析和识别.

1 姿势序列 FSM 动作识别方法框架

姿势序列 FSM 动作识别方法框架如图 1 所示. 首先，将体感交互设备获得的肢体节点数据变换到以用户为中心的肢体节点坐标系，通过定义肢体节点特征向量对肢体动作序列进行采样分析，建立预定义动作轨迹正则表达式，以构造姿势序列 FSM，从而实现对预定义肢体动作的解析和识别.

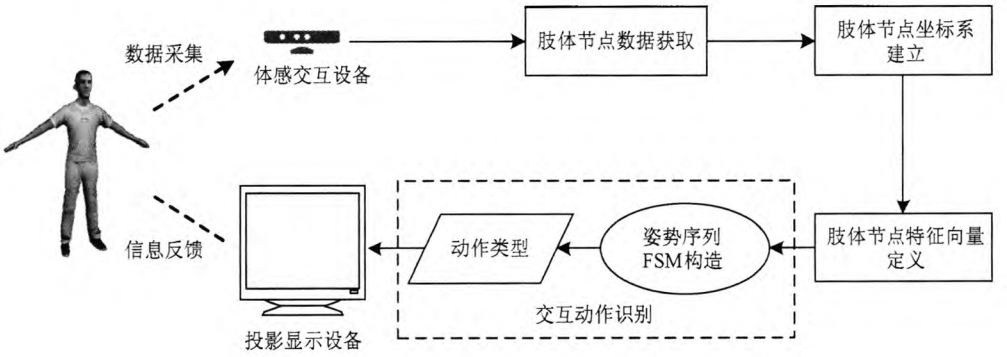
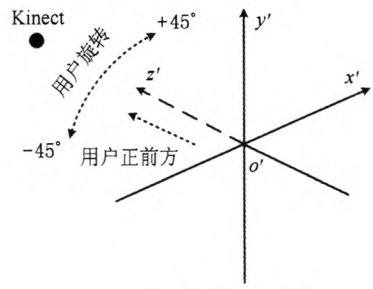


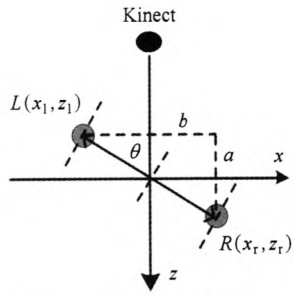
图 1 本文方法框架

1.1 肢体节点坐标系建立

为了尽可能地消除用户个体差异，需要将用户动作的空间描述从设备空间坐标系转换到用户空间坐标系，建立符合用户个体属性的交互动作特征. 本文定义用户空间坐标系如下：以用户右手方向为 x 轴正方向，头部正上方为 y 轴正方向，面向交互设备正前方为 z 轴正方向，两肩中心为坐标原点. 在动作识别过程中，由于用户身体正方向不一定与交互设备平面垂直，因此，需要对获取的用户肢体节点数据进行变换，建立用户肢体节点坐标系. 空间坐标系变换如图 2 所示，图 2 a 描述了用户空间坐标系，O' 表示用户空间坐标系 $o'x'y'z'$ 的原点；图 2 b 描述了 Kinect 空间坐标系下用户绕 y 轴进行旋转的俯视图，其中 $L(x_l, z_l)$ 表示 Kinect 空间坐标系中的用户左肩映射坐标点，$R(x_r, z_r)$ 表示用户右肩映射坐标点，θ 表示用户相对于设备 xoy 平面的旋转角度



a 用户空间坐标系



b Kinect 坐标系下用户旋转俯视图

图 2　空间坐标系转换

$(-45°<\theta<+45°)$.

由于获取的肢体节点数据是镜面对称的[15]，因此，Kinect 空间坐标系 $oxyz$ 下的坐标点 $P(x,y,z)$ 与用户空间坐标系 $o'x'y'z'$ 下的坐标点 $P'(x',y',z')$ 的变换关系可描述为

$$(x',y',z',1)=(x,y,z,1)\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&-1&0\\-x_0&-y_0&z_0&1\end{pmatrix}\cdot\begin{pmatrix}\cos\theta&0&-\sin\theta&0\\0&1&0&0\\\sin\theta&0&\cos\theta&0\\0&0&0&1\end{pmatrix}\tag{1}$$

其中，$O'(x_0,y_0,z_0)$ 表示用户空间坐标系 $o'x'y'z'$ 的原点坐标，θ 为用户相对于传感器 xoy 平面的旋转角度，$\theta=\arctan((x_r-x_l)/(z_r-z_l))$，其中 $x_r>x_l$，$-45°<\theta<+45°$.

用户坐标系下单位度量描述为单位立方体网格，对于不同身高的用户需要考虑其高度和肢体长度的比例对应关系，并用统一的方式对肢体动作进行描述. 通过式(1)的坐标系变换后，在用户坐标系中建立当前用户特有的空间网格模型，本文中将网格模型划分为 w^3 的三维立方体网格(w 为一维网格划分数，本文经验取值 $w=11$)，空间网格在 xoy 平面的截面示意图如图 3 所示.

在用户坐标系下，以原点为中心，分别对三维方向的网格进行比例划分，x 轴正负方向比例为 1∶1，y 轴正负方向比例为 3∶8，z 轴正负方向比例为 6∶5. 通过计算出单位网格的边长 d，用统一的方式对动作类型进行描述，本文根据用户相对身高比例定义网格边长，单位网格边长 d 可描述为 $d=h/(w-1)$；其中，h 表示当前用户在用户坐标系下的相对高度，w 为一维网格划分数.

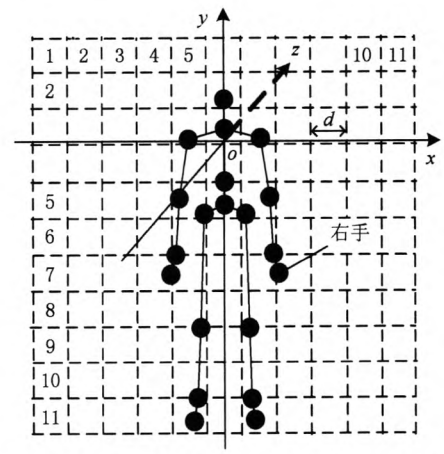


图 3　空间网格划分在 xoy 平面的截面示意图

建立三维网格划分模型后可以对用户坐标系中的区域以立方体网格形式进行描述，以保证始终以用户为中心建立独立的用户肢体节点坐标系，从而尽可能地消除用户个体差异.

1.2　肢体节点特征向量定义

动作在某一时间点的状态为静态姿势，人体某一关节或多个关节点在空间的运动序列为动态行为[16]. 识别动作之前，需要在用户空间坐标系下描述通用特征数据，它一般包括相对关节点的三维坐标信息、关节点空间运动矢量、关节点间空间距离等，肢体动作特征数据描述如图 4 所示.

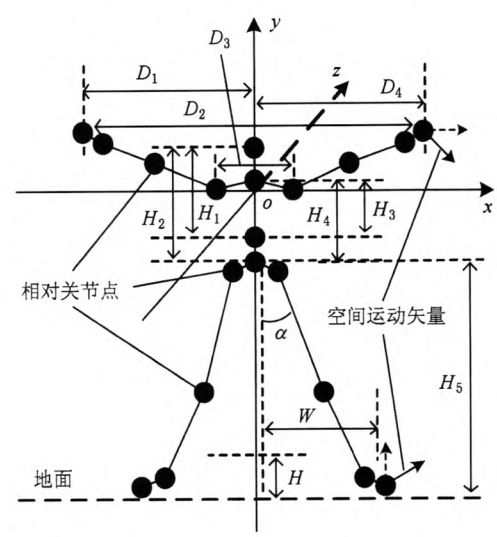


图 4　用户空间坐标系下的特征数据表示

本文定义肢体节点特征向量来描述动作特征数据，通过对肢体节点特征向量参数的计算和分析，实现对多个特定姿势组合形成的动态序列的识别，即对肢体动作的识别. 肢体节点特征向量包括关节点空间运动矢量、关节点运动时间间隔和关节点空间距离，肢体节点特征向量定义为

$$\boldsymbol{V}(T,k)=\left[\bigcup_{i=0}^{s-1}\boldsymbol{J}_k^i\boldsymbol{J}_k^{i+1},\Delta t_k^s,|P_mP_n|\right] \tag{2}$$

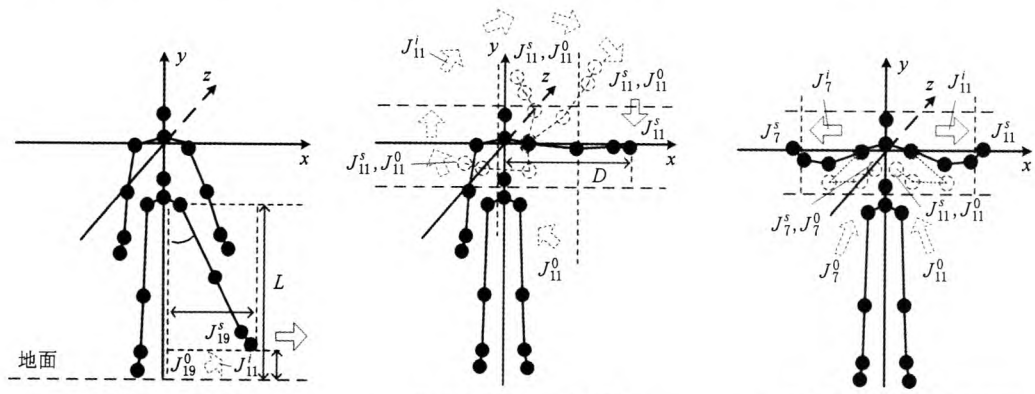
其中，T 表示动作类型，$k(0\leqslant k\leqslant 19)$ 表示关节点索引，$i(i=0,1,\cdots,s)$ 表示当前采样帧，s 表示对应关节点到达下一个特定采样点的结束帧，$\boldsymbol{J}_k^i\boldsymbol{J}_k^{i+1}$ 表示关节点 k 从当前采样帧 i 运动到下一帧 $i+1$ 的空间运动矢量，J_k^i 表示关节点 k 在第 i 帧的空间坐标点 (x_k^i,y_k^i,z_k^i)，Δt_k^s 表示关节点 k 从 J_k^0 坐标点通过轨迹运动到 J_k^s 坐标点的时间间隔，$|P_mP_n|$ 表示人体特定关节点之间的空间距离，该距离作为网格模型中的比例特征校验量.

每个关节点定义空间运动矢量 $\boldsymbol{J}_k^i\boldsymbol{J}_k^{i+1}$，计算出肢体节点的运动方向和轨迹，动作的每一步采样点转移所用时长可以通过时间间隔 $\Delta t_k^s=t_k^s-t_k^0$ 进行描述；其中，t_k^0 和 t_k^s 分别对应关节点 k 在每组起始采样帧和结束采样帧的时刻. 由式(2)定义可知，$J_k^i=(x_k^i,y_k^i,z_k^i)$，$J_k^{i+1}=(x_k^{i+1},y_k^{i+1},z_k^{i+1})$，则空间运动矢量 $\boldsymbol{J}_k^i\boldsymbol{J}_k^{i+1}$ 表示为 $\boldsymbol{J}_k^i\boldsymbol{J}_k^{i+1}=(x_k^{i+1}-x_k^i,y_k^{i+1}-y_k^i,z_k^{i+1}-z_k^i)$.

在 $|P_mP_n|$ 中，P_m 和 P_n 分别表示人体肢体部位的两端关节点，m 和 n 分别表示将关节点集合起始和终止索引号，其中 $m<n$，点 (x_j,y_j,z_j) 表示人体肢体部位对应关节点的空间坐标，$j(m\leqslant j\leqslant n-1)$ 表示在计算时对应关节点的索引变量，则肢体关节点之间的空间固定距离计算公式为

$$|P_mP_n|=\sum_{j=m}^{n-1}\sqrt{(x_j-x_{j+1})^2+(y_j-y_{j+1})^2+(z_j-z_{j+1})^2}.$$

根据上述定义的肢体节点特征向量参数，可以定义各种交互动作. 根据躯干部位和运动特性的不同，将动作类型进行分类阐述，3 类代表动作的肢体节点特征向量如图 5 所示. 图 5a 所示为右腿侧踢的肢体节点特征向量示意图，图 5b 所示为右手划圆的肢体节点特征向量示意图，图 5c 所示为双手水平展开的肢体节点特征向量示意图.



a 腿部动作——右腿侧踢　b 单手动作——右手划圆　c 双手动作——双手水平展开

图 5 3 类代表动作肢体节点特征向量示意图

根据式(2)，定义可表示出 3 类代表动作的肢体节点特征向量 $\boldsymbol{V}(T,k)$. 当关节点 k 到达下一个特定采样点时，结束帧 s 被确定，将当前的特征向量作为状态转移函数的输入参数进行分析，将采样帧 i 置零；等待下一次的结束帧，再分析，再置零，直到最后一个采样点. 具体定义如下：

1) 对于图 5a 中的右腿侧踢动作，提取右脚关节点($k=19$)的通用特征数据，从而定义肢体节点特征向量 $\boldsymbol{V}(\text{“右腿侧踢”},19)=\left[\bigcup_{i=0}^{s-1}\boldsymbol{J}_{19}^i\boldsymbol{J}_{19}^{i+1},\Delta t_{19}^s,L\right]$，其中，$L$ 为腿长度.

2) 对于图 5b 中的右手划圆动作，提取右手关节点($k=11$)的通用特征数据，从而定义肢体节点特征向量 $\boldsymbol{V}(\text{“右手划圆”},11)=\left[\bigcup_{i=0}^{s-1}\boldsymbol{J}_{11}^i\boldsymbol{J}_{11}^{i+1},\Delta t_{11}^s,D\right]$，其中 D 为手臂长度.

3) 对于图 5c 中的双手水平展开动作，提取左/右手关节点($k=7,11$)的通用特征数据，从而定义肢体节点特征向量 $\boldsymbol{V}(\text{“双手水平展开”},7)=\left[\bigcup_{i=0}^{s-1}\boldsymbol{J}_7^i\boldsymbol{J}_7^{i+1},\Delta t_7^s,D\right]\wedge\boldsymbol{V}(\text{“双手水平展开”},11)=\left[\bigcup_{i=0}^{s-1}\boldsymbol{J}_{11}^i\boldsymbol{J}_{11}^{i+1},\Delta t_{11}^s,D\right]$.

以此类推，采用此方法可以为其他肢体动作定义肢体节点特征向量；然后通过姿势序列 FSM 对它们进行分析，从而实现动作识别.

1.3　姿势序列 FSM 构造

针对人体自然交互动作具有多样性和多变性特点[17]，需要一种通用且高效的方法来识别动作. 每个动作由对应的肢体关节点的连续运动轨迹构成，连续的运动轨迹可以由离散的关键点进行拟合，每个关键点对应特定的姿势状态，通过识别每个状态的转移变化过程可以实现动作的判定. 基于上述思想，本文提出姿势序列 FSM 方法识别预定义的肢体动作. 姿势序列表示一个动作由多个姿势在时间轴上描述的一组运动序列，姿势序列 FSM 描述了每个动作的有限个状态以及各个状态之间的转移过程. 本文定义了姿势序列 FSM 为 Λ，其五元组表示为

$$\Lambda=(S,\Sigma,\delta,s_0,F) \tag{3}$$

其中，S 表示状态集$\{s_0,s_1,\cdots,s_n,f_0,f_1\}$，其对动作的每个特定的姿势状态进行描述；$\Sigma$ 表示输入的肢体节点特征向量集和限制参数字母表$\{u,\neg p,\neg t\}$，其中符号“$\neg$”表示逻辑否定；δ 为转移函数，定义为 $S\times\Sigma\rightarrow S$，表示姿势序列 FSM 从当前状态转换到后继状态；s_0 表示开始状态；$F=\{f_0,f_1\}$ 为最终状态集合，分别表示识别成功状态和识别无效状态.

字母表 Σ 中，变量 u 代表某个动作类型对应的所有肢体节点特征向量 $\boldsymbol{V}$ 的集合，特征向量表示动作轨迹在空间网格中的离散点域规则，通过点域规则可以构造出动作的轨迹正则表达式.

路径限制 $p=\{xyz\mid x\in[x_{\min},x_{\max}],y\in[y_{\min},y_{\max}],z\in[z_{\min},z_{\max}]\}$ 对特定的姿势进行关键点的范围控制，在任何情况下超出预定义路径范围，即 $\neg p$ 为真，则被标记为无效状态.

时间戳 $t\in[t_{\text{start}},t_{\text{end}}]$ 规定了动作在当前状态到后继状态转移所需要的时间，若动作的某个状态在规定的时间内未转移到后继有效状态，即 $\neg t$ 为真，则跳转到无效状态.

每个动作由几个典型的静态姿势构成，静态姿势对应已定义的状态量，每种状态量由关键点特征数据在空间网格中计算得到，动作状态转移必须满足路径限制 p 和时间戳 t 的条件，从而识别动作类型、理解用户交互意图. 通过五元组可以描述姿势序列 FSM 的各项属性特性以及每一步转移过程，姿势序列 FSM 运行过程的状态图模型如图 6 所示.

在初始状态 s_0 下，按照预定义的动作达到第一个有效状态 s_1，如果下一时刻的姿势仍然在预定义的范围内，则达到后继有效状态 s_k. 以此类推，直至达到成功状态 f_0，即识别动作成功. 在任意有效状态下，如果行为超出路径限制或时间戳范围，则直接

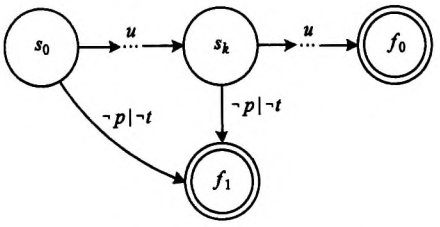


图 6　姿势序列 FSM 原型

标记该序列动作为无效状态，即识别动作失败. 在达到任意结束状态后，当前的姿势序列 FSM 运行完毕，重新初始化进行下一组肢体动作的识别. 通过姿势序列 FSM 的状态图可以得到状态转移表 1.

表 1　姿势序列 FSM 状态转移表

S	Σ		
	u	$\neg p$	$\neg t$
s_0	s_k	f_1	f_1
s_k	$s_{(k+x)}$	f_1	f_1
$s_{(k+x)}$	f_0	f_1	f_1
f_0	—	—	—
f_1	—	—	—

表 1 中，$k,x=0,1,2,\cdots,n$，且 $k\neq k+x\leqslant n$，n 表示识别动作所需要的中间有效状态总数. 该姿势序列 FSM 接受的点域规则为 $L(\Lambda)=\{(u^n\mid u^*\ \neg p\mid u^*\ \neg t)\mid n\geqslant 1\}$.

由式(3)定义，采用姿势序列 FSM 对 1.2 节中的动作进行描述. 输入字母表 $\Sigma=\{a_i,b_i,c_i,d_i,e_i,f_i,g_i,h_i,m_i\}$，其中 $i=0,1$，每个字母变量描述了采样点在空间网格模型中所关联的空间范围，即点域. 运动轨迹可采用多个点域进行拟合，点域字符串

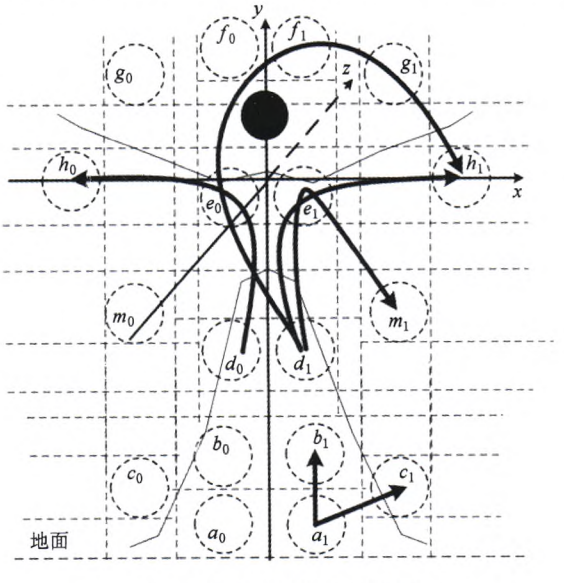


图 7　部分动作轨迹示意图

构成了对某个动作轨迹的离散化描述.

针对图 5 所示 3 类代表动作的部分动作轨迹示意图如图 7 所示,其中,x 负方向空间中的点域用下标 $i=0$ 表示,x 正方向空间中的点域用下标 $i=1$ 表示. 表 2 给出了 3 类代表动作的特征向量点域字符串.

表 2 3 类代表动作的特征向量点域字符串

代表动作	特征向量点域字符串
右腿侧踢	a_1c_1
右手划圆	$d_1e_0f_1g_1h_1$
双手水平展开	$(d_0 \wedge d_1)(e_0 \wedge e_1)(h_0 \wedge h_1)$

通过动作的特征向量点域字符串可以提取到动作的轨迹正则表达式. 在字母表 Σ 中,$\Sigma_{i(1-i)}=\Sigma_i \wedge \Sigma_{1-i}$,其中 $i=0,1$;表示沿 yoz 平面对称的空间点域同时成立. 动作的轨迹正则表达式整理简化后如

$$R=a_ic_i \mid d_ie_{1-i}f_ig_ih_i \mid d_{i(1-i)}e_{i(1-i)}h_{i(1-i)} \quad (4)$$

根据式(4)得出 3 类代表动作及其对称动作类型的 FSM 图,如图 8 所示;其中省略了 f_1 无效状态,s_0 为初始状态,s_k 为过渡有效状态,f_0 状态代表可接受点域字符串的成功状态. 在初始和任意有效状态下,如果动作姿势超出路径限制或时间戳范围,则直接标记该系列动作为无效状态 f_1,即识别动作失败. 在达到任意结束状态后,当前动作的姿势序列 FSM 运行完毕,重新初始化进行下一组动作行为的识别.

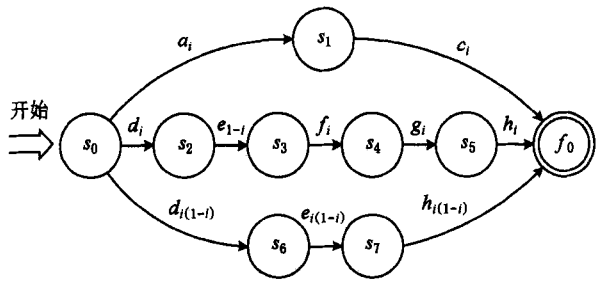


图 8 3 类代表动作的姿势序列 FSM

在姿势序列 FSM 运行过程中需要同步对动作集合进行优化处理,优化算法步骤如下:

Step1. 初始化所有可能动作的集合 $\{T\}=\{$"右腿侧踢","右手划圆","双手水平展开",…$\}$,其中每个特定的动作对应了多个点域字符串表达式.

Step2. 在姿势序列 FSM 运行过程中,每一步状态转移后将所有不可能的动作从集合中排除,所有可能的动作在集合中进行保留.

Step3. 当到达结束状态时,如果最终状态为无效状态,则无任何输出,并重新开始;如果为可接受的成功状态,则当前集合的唯一元素即为动作类型,进行动作类型输出,跳转到 Step1,循环识别.

最后,由用户定义交互动作的语义和用途,用户通过得到的动作类型赋予新的语义,如左右抬腿代表场景漫游、左右手滑动代表文档翻页、双手水平展开代表拉开帷幕等,从而实现体感交互应用.

2 实验结果与分析

2.1 实验测试及结果

根据本文提出的姿势序列 FSM 模型和算法实现了 17 种肢体动作的识别,并在 Intel Xeon CPU (2.53 GHz) X3440,4 GB 内存的 Windows 7 x64 系统下进行了测试. 肢体动作定义如表 3 所示,根据运动特性和躯干部位的不同,将动作类别分为腿部动作、单手动作和双手动作.

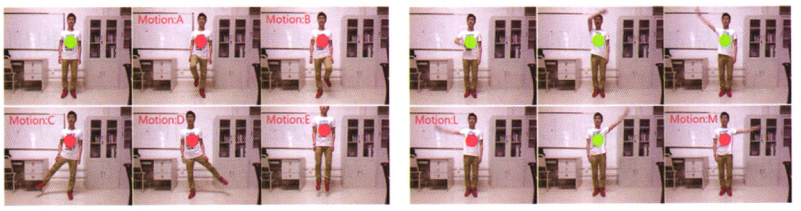
表 3 肢体动作定义表

动作分类	动作名称	对应标识	动作详细说明
腿部动作	左抬腿	A	左脚向上垂直提起
	右抬腿	B	右脚向上垂直提起
	左腿侧踢	C	左腿伸直向左侧踢出
	右腿侧踢	D	右腿伸直向右侧踢出
	起跳	E	双脚同时起跳
单手动作	左手滑动	F	左手从胸前向左滑出
	右手滑动	G	右手从胸前向右滑出
	左手向上举	H	左手向上举起
	右手向上举	I	右手向上举起
	左手向下按	J	左手抬起后快速向下按
	右手向下按	K	右手抬起后快速向下按
	左手划圆	L	左手在身体前方逆时针划圆弧
	右手划圆	M	右手在身体前方顺时针划圆弧
双手动作	双手斜向上推	N	双手从胸前向左右斜上方推出
	双手斜向下推	O	双手从胸前向左右斜下方推出
	双手水平展开	P	双手从胸前向左右外侧水平推出
	双手水平收缩	Q	双手从左右两侧向胸前合拢

图 9 展示了 17 种预定义肢体动作动态识别过程,图中人体胸前的绿点表示识别过程的部分状态,红点表示识别成功状态.

对 30 名不同身高和体形的志愿者进行实验测

试，每名待测者对每一种肢体动作进行 3 次重复样本测试，总共 1 530 个动作实例；动作识别测试结果混淆矩阵如表 4 所示，其中 None 表示未检测到任何动作.



a 腿部动作

b 左/右手划圆

c 左/右手向上举

d 左/右手向下按

e 双手斜向上推

f 双手斜向下推

g 双手水平展开

h 双手水平收缩

i 左/右手水平滑动

图 9　动作识别实现效果

表 4　动作识别测试结果混淆矩阵

T	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	None
A	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	98.9	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D	0.0	0.0	0.0	98.9	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.0	94.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6
F	0.0	0.0	0.0	0.0	0.0	97.8	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1
G	0.0	0.0	0.0	0.0	0.0	0.0	97.8	0.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1
H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
I	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
J	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
K	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
L	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	0.0	0.0	96.7	0.0	0.0	0.0	0.0	0.0	1.1
M	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0	95.6	0.0	0.0	0.0	0.0	1.1
N	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	96.7	0.0	0.0	0.0	3.3
O	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	94.4	0.0	0.0	5.6
P	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	95.6	0.0	2.2
Q	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.9	1.1

测试结果表明，本文方法的动作识别率均在94%以上，并且大部分动作识别率达到100%，平均识别率达98%，能够满足体感交互应用的需求. 其中，双脚起跳(E)在没有明显起跳幅度的情况下无法识别，左/右划圆(L,M)动作可能会被误识别为左/右手向上举(H,I)的动作，双手斜向上和向下推(N,O)动作可能出现未识别的情况，双手水平展开(P)可能会被误判为双手斜向下推(O)的动作. 在总体上，本文方法能够很好地识别上述肢体动作，实际测试识别反馈时间在0.060～0.096 s之间，满足实时交互要求.

2.2 实验对比及分析

表5给出了人体动作识别方法比较，其中，多实例学习方法[12]、动作子集组合方法[13]和SSS特征匹配方法[14]采用的数据集为MSR-Action3D Dataset，本文采用的测试数据集为30名实际用户的1530个动作实例. 实验中识别准确率在0.6以下为“低”，在0.6～0.8之间为“中”，在0.8～1.0之间为“高”.

表5 人体动作识别方法比较

识别方法	基本技术	识别准确率	鲁棒性	计算复杂度	识别反馈时间/s	扩展性	离线训练	能否处理未分割数据流	适合动作类别
FAAST动作识别方法[11]	事件触发	>0.95	低	低	<0.1	强	否	能	简单动作
多实例学习方法[12]	机器学习、模板匹配	0.657	中	中	<1.0	中	是	能	简单、连续动作
动作子集组合方法[13]	机器学习、模板匹配	0.817	高	高	<1.5	中	是	否	简单、连续动作
SSS特征匹配方法[14]	机器学习、模板匹配	0.882	高	高	−1.5～+1.5	中	是	能	简单、连续动作
本文方法	姿势序列有限状态机	>0.94	中	中	<0.1s	强	否	能	简单、连续动作

多实例学习方法[12]从动作数据序列中确定关键帧实例，从而推导出动作模板；但是，动作模板仅保存同一类行为的形态和模型，忽略了变化，在鲁棒性和实时性上表现一般. 动作子集组合方法[13]对关节点子集进行分类，识别率较高，但其着重于是事先分割的数据流级别，不能用于从未分割数据流中进行在线识别，虽然鲁棒性较高，却计算复杂，实时性较差. SSS特征匹配方法[11]通过离线训练建立一个特征字典和手势模型，为未知动作的动作数据流的每一帧数据分配标签，通过提取SSS特征在线预测动作类型，能够从未分割数据流中进行在线识别，鲁棒性较高；但其计算复杂，识别反馈时间不稳定(−1.5 s表示提前1.5 s识别，+1.5 s表示延后1.5 s识别). 以上3种方法都采用了机器学习和模板匹配技术实现，它们对每个动作识别均需要特征字典库，对于扩展动作类型识别时，需要收集大量动作数据进行离线训练，对特定动作识别与训练集耦合度较高，因此，扩展性一般. FAAST动作识别方法[11]采用角度、距离、速度等事件触发方式进行识别，其计算量小、实时性好、扩展性较强，对于已定义的简单动作识别准确率高；但由于事件触发技术本身具有局限性，鲁棒性较低，且对连续动作识别较困难.

本文方法的优势主要有：1)动作识别准确率高. 通过对30名不同身高和体形的用户进行重复3次样本测试，识别准确率在94%以上；2)动作识别反馈时间快. 实际测试识别反馈时间在0.060～0.096 s之间，小于0.1 s；3)在扩展动作类型时不需要收集大量动作数据进行离线训练，对于特定动作只需要定义肢体动作的轨迹正则表达式，通用性和扩展性强；4)本文提出的姿势序列FSM模型中定义了初始和结束状态，通过肢体节点特征向量进行实时分析，能够处理未分割的动作数据流；5)姿势序列FSM是一种以离散姿势序列来拟合连续动作轨迹的方法，因此，其适用于简单和连续动作的识别. 但本文方法在鲁棒性上表现不够好，主要是由于在识别过程中任一状态不符合预定义规则即被视为无效状态，因此，姿势序列识别较为敏感，需要用户动作在个人风格范围内尽量规范.

3 结 语

本文提出了一种姿势序列FSM动作识别方法，实现了对预定义动作的快速识别. 本文方法采用肢体节点特征向量描述肢体动作特征数据，对预定

义肢体动作序列进行采样分析，建立肢体动作轨迹正则表达式，构造出姿势序列 FSM，以实现肢体动作识别. 本文方法可对任意动作或手势进行整体描述，不需要离线训练和学习，通用性和扩展性较强，且对简单和连续动作的识别准确率高，实时性好，并满足体感交互应用需求.

本文方法的不足之处在添加新动作时必须考虑动作定义之间的排斥与相容情况，其在鲁棒性上表现不够好. 在后续的研究工作中将对规则和特征进行优化，提升鲁棒性，并提供可视化界面来支持自定义修改动作特征数据参数，以满足用户个人风格交互的定义需求.

致谢　在论文写作过程中，澳大利亚昆士兰大学(The University of Queensland, Australia) Li Xue 博士提供了宝贵的论文资料，同时与 Zhao Xin 博士就实验测试结果进行多次交流讨论，在此一并致谢!

参考文献(References)：

[1] Yu Tao. Kinect application development and actual combat: in the most natural way to dialogue with the machine [M]. Beijing: China Machine Press, 2012: 46-47 (in Chinese)
(余　涛. Kinect 应用开发实战：用最自然的方式与机器对话 [M]. 北京：机械工业出版社，2012：46-47)

[2] Wang J, Xu Z J. STV-based video feature processing for action recognition [J]. Signal Processing, 2013, 93(8): 2151-2168

[3] Xu Guangyou, Cao Yuanyuan. Action recognition and activity understanding: a review [J]. Journal of Image and Graphics, 2009, 14(2): 189-195 (in Chinese)
(徐光祐，曹媛媛. 动作识别与行为理解综述[J]. 中国图象图形学报，2009，14(2)：189-195)

[4] van den Bergh M, Carton D, de Nijs R, *et al*. Real-time 3D hand gesture interaction with a robot for understanding directions from humans [C] //Proceedings of the 20th IEEE International Symposium on Robot and Human Interactive Communication. Los Alamitos: IEEE Computer Society Press, 2011: 357-362

[5] Zhang Q S, Song X, Shao X W, *et al*. Unsupervised skeleton extraction and motion capture from 3D deformable matching [J]. Neurocomputing, 2013, 100: 170-182

[6] Shotton J, Sharp T, Kipman A, *et al*. Real-time human pose recognition in parts from single depth images [J]. Communications of the ACM, 2013, 56(1): 116-124

[7] El-laithy R A, Huang J D, Yeh M. Study on the use of Microsoft Kinect for robotics applications [C] //Proceedings of Position Location and Navigation Symposium. Los Alamitos: IEEE Computer Society Press, 2012: 1280-1288

[8] Oikonomidis I, Kyriazis N, Argyros A. Efficient model-based 3D tracking of hand articulations using Kinect [C] // Proceedings of the 22nd British Machine Vision Conference. British: BMVA Press, 2011: 1-11

[9] Shen Shihong, Li Weiqing. Research on Kinect-based gesture recognition system [C] //Proceedings of the 8th Harmonious Human Machine Environment Conference CHCI. Beijing: Tsinghua University Press, 2012: 55-62 (in Chinese)
(沈世宏，李蔚清. 基于 Kinect 的体感手势识别系统的研究 [C] //第 8 届和谐人机环境联合学术会议论文集 CHCI. 北京：清华大学出版社，2012：55-62)

[10] Soltani F, Eskandari F, Golestan S. Developing a gesture-based game for deaf/mute people using Microsoft Kinect [C] // Proceedings of the 6th International Conference on Complex, Intelligent and Software Intensive Systems. Los Alamitos: IEEE Computer Society Press, 2012: 491-495

[11] Suma E A, Krum D M, Lange B, *et al*. Adapting user interfaces for gestural interaction with the flexible action and articulated skeleton toolkit [J]. Computers & Graphics, 2013, 37(3): 193-201

[12] Ellis C, Masood S Z, Tappen M F, *et al*. Exploring the trade-off between accuracy and observational latency in action recognition [J]. International Journal of Computer Vision, 2013, 101(3): 420-436

[13] Wang J, Liu Z C, Wu Y, *et al*. Mining actionlet ensemble for action recognition with depth cameras [C] //Proceedings of Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 1290-1297

[14] Zhao X, Li X, Pang C Y, *et al*. Online human gesture recognition from motion data streams [C] //Proceedings of the 21st ACM International Conference on Multimedia. New York: ACM Press, 2013: 23-32

[15] Biswas K K, Basu S K. Gesture recognition using Microsoft Kinect® [C] //Proceedings of the 5th International Conference on Automation, Robotics and Applications. Los Alamitos: IEEE Computer Society Press, 2011: 100-103

[16] Zhang Yi, Zhang Shuo, Luo Yuan, *et al*. Gesture track recognition based on Kinect depth image information and its applications [J]. Application Research of Computers, 2012, 29(9): 3547-3550 (in Chinese)
(张　毅，张　烁，罗元等. 基于 Kinect 深度图像信息的手势轨迹识别及应用[J]. 计算机应用研究，2012，29(9)：3547-3550)

[17] Chaquet J M, Carmona E J, Fernandez-Caballero A. A survey of video datasets for human action and activity recognition [J]. Computer Vision and Image Understanding, 2013, 117(6): 633-659

姿势序列有限状态机动作识别方法

作者： 林水强， 吴亚东， 余芳， 杨永华， Lin Shuiqiang， Wu Yadong， Yu Fang， Yang Yonghua

作者单位： 林水强,余芳,杨永华,Lin Shuiqiang,Yu Fang,Yang Yonghua(西南科技大学计算机科学与技术学院 绵阳 621010)， 吴亚东,Wu Yadong(西南科技大学计算机科学与技术学院 绵阳621010;西南科技大学核废物与环境安全国防重点学科实验室 绵阳621010)

刊名： 计算机辅助设计与图形学学报 ISTIC EI PKU

英文刊名： Journal of Computer-Aided Design & Computer Graphics

年，卷(期)： 2014,26(9)