

基于四元数 3D 骨骼表示的人体行为识别

徐海洋¹, 孔军^{1,2}, 蒋敏¹

¹江南大学物联网工程学院, 江苏 无锡 214122;

²新疆大学电气工程学院, 新疆 乌鲁木齐 830047

摘要 为了准确地描述人体骨骼的运动细节以及 3D 骨骼间的几何关系, 提出一种基于四元数 3D 骨骼表示的人体行为识别方法。首先, 在已捕获的关键帧集合的基础上, 对普通关键帧和变速关键帧分别采用线性插值和二次多项式插值, 获得相同帧数的骨骼序列; 然后, 针对所得的骨骼序列, 采用四元数对每帧中 3D 骨骼间的几何关系进行描述, 获得四元数骨骼特征描述子; 最后, 采用支持向量机分类器对这一系列特征描述子进行训练和测试, 得到最终的识别结果。在 3 个标准数据库上的实验结果均显示, 四元数骨骼特征描述子对噪声、运动速度变化、视角变化和时域不对齐都具有很好的稳健性, 可以显著提高人体行为识别的准确率。

关键词 图像处理; 人体行为识别; 四元数特征描述子; 关键帧; 动态时间规整算法; 支持向量机

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP55.021002

Human Action Recognition Based on Quaternion 3D Skeleton Representation

Xu Haiyang¹, Kong Jun^{1,2}, Jiang Min¹

¹School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China;

²College of Electrical Engineering, Xinjiang University, Urumqi, Xinjiang 830047, China

Abstract We propose a new human action recognition method based on quaternion three-dimensional (3D) skeleton representation, in order to accurately describe the movement details of human skeletons and 3D geometric relationship of skeletons. Firstly, we obtain skeletal sequences with the same frame quantity by applying linear interpolation and quadratic polynomial interpolation to normal key frames and variable key frames, respectively, on the basis of the captured key frames. Secondly, we use quaternions to represent 3D geometric skeletal relationship of the obtained skeletal sequences to generate quaternion feature descriptors. Finally, we use the support vector machine classifier to train and test the quaternion feature descriptors to realize recognition. Experimental results based on three standard datasets prove that the quaternion feature descriptor is robust to noise, changes of moving rate and viewpoint, and time domain misalignment, and it is able to improve the identification accuracy of human behavior significantly.

Key words image processing; human action recognition; quaternion feature descriptor; key frames; dynamic time warping algorithm; support vector machine

OCIS codes 100.4993; 100.4994; 100.5010

收稿日期: 2017-07-19; 收到修改稿日期: 2017-08-16

基金项目: 国家自然科学基金(61362030, 61201429)、中国博士后科学基金(2015M571720, 2016M606360)、江苏省博士后科学基金(1601416C)、中国公安部技术科研项目(2014JSYJB007)

作者简介: 徐海洋(1992—), 男, 硕士研究生, 主要从事人体行为识别和机器学习方面的研究。

E-mail: 1041922650@qq.com

导师简介: 孔军(1974—), 男, 博士, 副教授, 主要从事人脸定位、目标跟踪、行为识别方面的研究。

E-mail: kongjun@jiangnan.edu.cn(通信联系人)

1 引言

人体行为识别是指计算机对人体行为的描述和理解,是应用于视频监控、体感游戏和人机交互等领域的关键技术,主要研究内容有目标检测、特征提取与表示和行为识别^[1-3],其中基于 RGB 视频序列的人体行为识别方法已经较为成熟^[4-5]。近年来,随着传感器技术(如 Kinect 传感器)的发展,基于深度图和骨骼视频序列的人体行为识别方法应运而生。相较于传统的彩色摄像头捕捉系统,新兴的捕捉系统对光照、场地、遮挡具有一定的稳健性,能提供骨骼关节的三维坐标信息,并且可以有效地保护采集者的个人隐私。基于骨骼的人体行为识别方法采用骨骼关节模型来描述行为。文献[6]已证明人体行为模式可用若干个人体关节表示。基于此理论基础,基于骨骼视频序列的人体行为识别研究近来得到广泛关注。

现有的基于骨骼视频序列的人体行为识别方法大致可以分为:基于关节的方法和基于身体部位的方法。基于关节的方法灵感来自于 Johansson 经典的移动灯显示实验^[6],将人体骨骼简单地看作点集,由此产生各种基于人体骨骼关节或者关节组合的特征提取方法^[7-8],如: Luvizon 等^[7]提出了局部聚集描述子向量算法和分类池模型,利用骨骼关节子集提取局部时空特征向量,将特征向量与 LMNN 算法相结合,最后采用无参 K-NN 分类器进行分类; Mahasseni 等^[8]提出基于卷积神经网络(CNN)的长短期记忆模型(LSTM),训练时采用骨骼的标准 LSTM 编码,其中 LSTM 编码使用标准的反向传播优化梯度下降算法,最后采用 CNN 进行分类。基于身体部位的方法将人体骨骼视为一组连接的刚性段(身体部位),主要描述各个身体部位的运动^[9]或者身体部位间关节角的变化^[10],如: Li 等^[10]提出关节空间图形,采用顶点表示人体关节,边缘表示两两关节之间的相对变化,最后通过比较 2 张图片之间的相似性判断类别。

近几年,学界对四元数的研究取得一定突破。黎云汉等^[11]提出了应用于人脸识别的四元数主成分分析模型,首先用四元数矩阵模型表示彩色人脸图像,然后求该四元数矩阵的协方差矩阵及其特征,最后与已知人脸特征空间进行对比来识别人脸; Xu^[12]提出了基于四元数的判别分析方法,首先用四元数向量表示彩色图像像素,然后使用线性判别分析算法将四元数转换为一个低维度的四元数向量,

并在这个四元数向量空间里进行分类。

本文提出基于四元数 3D 骨骼表示的人体行为识别方法。首先,提出基于骨骼关键帧分类的插值方法,该方法根据帧间骨骼对应位置的夹角变化速度将序列分为两类,分别使用线性和二次多项式的插值方法获得相同帧数的运动序列;其次,采用四元数表示每帧骨骼向量间的相对空间位置,即将每帧的骨骼表示为骨骼向量,将骨骼由世界坐标系转为局部坐标系,通过四元数表示每对骨骼向量之间的相对空间位置,从而获得视频序列特征集;然后,采用动态时间规整(DTW)和傅里叶时域金字塔方法增强视频序列特征集;最后,采用线性支持向量机(SVM)分类器在 3 个标准数据库上进行训练和测试。

2 原理

图 1 是本文所采用的基于四元数的人体行为识别算法流程图。基于插值与四元数的行为识别方法的步骤包括数据预处理、特征提取、特征增强等。在训练阶段:首先,为了获得相同帧数的骨骼序列,采用关键帧分类的插值方法,即根据骨骼间夹角变化的速度采用相应的插值方法;其次,对获得的骨骼序列的每帧骨架采用四元数表示骨骼向量之间的位置关系,从而获得骨骼特征向量;然后,采用动态时间规整和傅里叶时域金字塔的方法解决速率不同、时域不对齐、噪声等问题,从而获得增强的骨骼特征向量;最后,采用 SVM 分类器训练获得每类的 SVM 模型。在测试阶段,前几步骤与训练阶段相同,获得增强的骨骼特征向量,将增强的骨骼特征向量经过每一类的 SVM 模型获得概率值,将最后值最大的作为测试样本的类别。

2.1 数据预处理

采用人体骨骼数据作为训练和测试数据,以人体骨骼拓扑结构及骨骼旋转特性为依据,将 Kinect 检测出的 20 个关节点(图 2)进行分层^[13]。

第一层,身体躯干关节点,支撑整个人体的躯干。该层由头、左右肩、脊椎等 8 个关节点(图 2 中括号里标记数字 1 的关节点)组成。这层关节点可以提供人体某些动作所需的特征信息。

第二层,四肢关节点。人体大部分动作(如挥手、出拳、踢腿等)是由人的四肢来完成的,因此包含了大量动作的特征信息。该层由左右肘、腕,左右膝、踝 8 个关节点(图 2 中括号里标记数字 2 的关节点)组成。

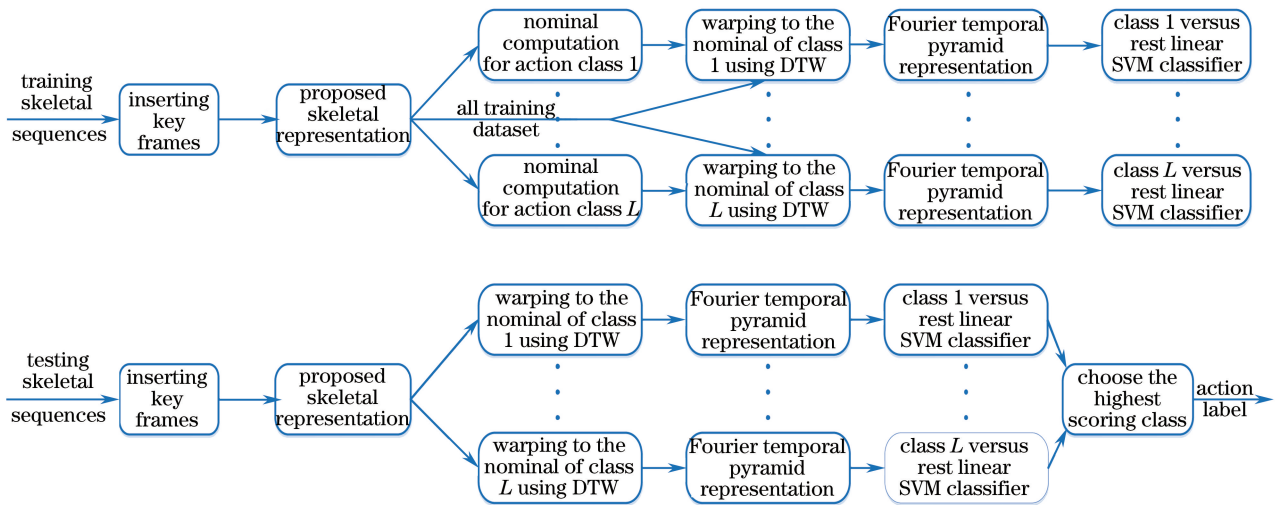


图 1 训练和测试流程图

Fig. 1 Flow chart of training and testing process

第三层,手与脚,包括左右手、左右脚这 4 个关节点(图 2 中括号里标记数字 3 的关节点)。在所研究的人体动作中,手掌和脚的特征信息非常微小,为了降低特征数据维数,手掌和脚的特征信息可以忽略。

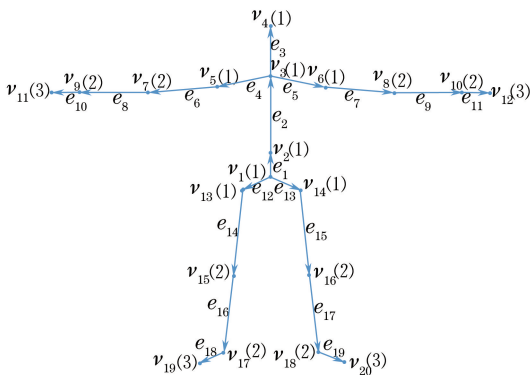


图 2 分层关节点与定义骨骼向量

Fig. 2 Hierarchical skeleton points and definition of skeletal vectors

通过分层,对其中 16 个重要的关节点进行操作。为了保证实验数据的准确性,以每个视频序列中每一帧的人体骨骼臀部作为坐标原点。

为了方便之后的特征提取和分类器训练,对连续的视频序列进行预处理,通过插值获取相同帧数的视频序列。为了解决关键帧提取中序列时序性和数据冗余度的问题,本文根据关节的运动特性将关键帧分为普通关键帧和变速关键帧^[14]。首先,计算序列中每个关节前后旋转的夹角,根据相邻两帧间夹角相差与阈值 θ_f 的比较结果,对关键帧进行分类,其中 θ_f 为经验值。

1) 普通关键帧

对小于阈值 θ_f 的关键帧运用球面线性插值的方法进行等角度的向量插值,将所获得的小于阈值所提取的关键帧称为普通关键帧,主要表示关节的匀速运动。

对普通关键帧采用通用的球面线性插值公式:

$$F_{IN}(e_i^k, e_i^{k+1}, t) = \frac{e_i^k \sin[(1-t)\theta] + e_i^{k+1} \sin(t\theta)}{\sin \theta}, \quad (1)$$

式中 t 为插值因子, θ 为第 i 个骨骼向量在第 k 帧和第 $k+1$ 帧的关节向量 e_i^k, e_i^{k+1} 的方向夹角。

$$t = \frac{T_c}{T}, \quad (2)$$

式中 T_c 为当前的时间, T 为总的插值时间。

2) 变速关键帧

在已有的关键帧集合中,存在一部分序列,是由关节进行大角度变换形成的关键帧,此序列中部分关节的旋转是先加速再减速。

对于人体运动十分丰富的四肢,若采用简单的线性插值将减少骨骼运动的细节,所以将表示加速到减速的关键帧作为第 2 种类型的关键帧,称为变速关键帧,主要用于表示快速运动中关节运动的细节。

对于变速关键帧,插值方法同(1)式所示,其中插值因子

$$t = \begin{cases} 2 \left(\frac{T_c}{T} \right)^2, & T_c < \frac{T}{2} \\ - \left(\frac{T_c}{T} - 1 \right)^2 + 1, & T_c \geq \frac{T}{2} \end{cases} \quad (3)$$

经过预处理,可以获得相同帧数的骨骼序列。

2.2 特征提取

分别计算了每帧中每对 3D 骨骼的几何关系, 对于每对骨骼的几何关系采用四元数的表示方式。在进行四元数表示时, 首先算出空间旋转角, 然后计算旋转轴, 这样就可以准确表示骨骼间的空间旋转, 即空间几何关系。

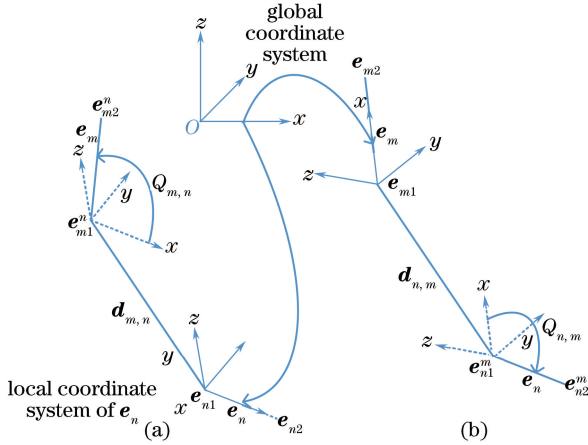


图 3 (a) 骨骼 e_m 在相对 e_n 的局部坐标系中的表示;

(b) 骨骼 e_n 在相对 e_m 的局部坐标系中的表示

Fig. 3 (a) Representation of body part e_m in local coordinate system of e_n ; (b) Representation of body part e_n in local coordinate system of e_m

四元数^[15]用复数形式表示: $r + a \times i + b \times j + c \times k$ (r, a, b 和 c 是实数), 满足 $i \times i = j \times j = k \times k = i \times k \times j = -1$ 的规则。根据欧拉旋转定理, 在三维空间中, 刚体的任意一组旋转可等价于在空间中绕某一轴 (u_x, u_y, u_z) 旋转某一角度 θ , 这一信息可用四元数表达为 $q = \cos(1/2\theta) + (u_x \times i, u_y \times j, u_z \times k) \cdot \sin(1/2\theta)$, 即四元数 (a, b, c, d) 可以用来表示旋转轴和旋转角度。其中, $\theta = 2 \cos a, u_x = b / \sin(1/2\theta), u_y = c / \sin(1/2\theta), u_z = d / \sin(1/2\theta)$ 。

根据以上原理, 采用四元数对骨骼序列进行描述。假设 $S = \{V, E\}$ 表示序列一帧中的骨骼, 其中 $V = \{v_1, \dots, v_N\}$ 表示一组关节点, $E = \{e_1, \dots, e_M\}$ 表示一组刚性的人体骨骼。如图 2 所示的示例框架, 包含有 20 个关节点和 19 个骨骼 (图中显示全部骨骼, 但在后面的实验中省略第三层骨骼)。用 $e_{n1} \in \mathbf{R}^3, e_{n2} \in \mathbf{R}^3$ 分别表示骨骼 e_n 的起点和终点^[16]。 l_n 表示骨骼 e_n 的长度。

假设一对骨骼 e_m 和 e_n , 为了描述其相对几何关系, 使用局部坐标系表示另外一个骨骼向量 (图 3), 即世界坐标系中的骨骼 e_n 通过在局部坐标系中旋转 (最小旋转) 和平移获得, 其中 e_{n1} 作为起点, 并且 e_n 的方向与 x 轴方向一致 [图 3 (a)]。

用 $e_{m1}^n(t), e_{m2}^n(t) \in \mathbf{R}^3$ 分别表示骨骼 e_m 在局部坐标系中相对于骨骼 e_n 在时刻 t 的起点和终点。

对于人体不同的动作而言, 通过某一个关节点连接的两段骨骼之间有着不同的空间夹角关系, 因此, 可通过图 2 定义的 15 个骨骼向量两两之间的 3D 几何关系来描述人体行为。

图 4 中两实线所成的夹角 θ 用来表示两段骨骼之间的空间夹角。以某两段骨骼向量为例, 计算其空间夹角的余弦值。在空间中, 设骨骼向量 e_n 用坐标表示为 (x_1, y_1, z_1) , 骨骼向量 e_m 用坐标表示为 (x_2, y_2, z_2) , 通过计算得到空间中这两段骨骼的夹角余弦, 公式为

$$\cos \theta = \frac{x_1 x_2 + y_1 y_2 + z_1 z_2}{\sqrt{x_1^2 + y_1^2 + z_1^2} \sqrt{x_2^2 + y_2^2 + z_2^2}}, \quad (4)$$

式中 $\sqrt{x_1^2 + y_1^2 + z_1^2} \neq 0$, 且 $\sqrt{x_2^2 + y_2^2 + z_2^2} \neq 0$ 。

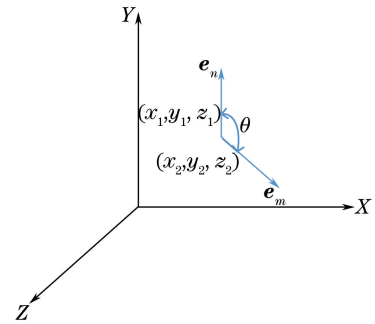


图 4 骨骼之间夹角的余弦值计算

Fig. 4 Angle cosine calculation between body parts

根据上述算法, 依次计算每帧骨骼中每对骨骼之间的夹角。接下来, 计算旋转轴。假设骨骼向量 e_n 和骨骼向量 e_m 是平面上不平行的向量, 则

$$\mathbf{r} = \mathbf{e}_n \times \mathbf{e}_m, \quad (5)$$

$$\text{式中 } \mathbf{e}_n \times \mathbf{e}_m = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix} =$$

$\left(\begin{vmatrix} y_1 & z_1 \\ y_2 & z_2 \end{vmatrix}, - \begin{vmatrix} x_1 & z_1 \\ x_2 & z_2 \end{vmatrix}, \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \right)$, \mathbf{r} 向量表示骨骼向量 e_n 旋转到骨骼向量 e_m 的旋转轴, 法向量的方向可由右手准则确定。

根据旋转轴 \mathbf{r} 与旋转角度 θ 以及四元数与旋转之间的关系, 得到单位四元数 q :

$$\mathbf{v} = (x, y, z) = \sin(\theta/2) \times \mathbf{r}, \quad (6)$$

$$\omega = \cos(\theta/2), \quad (7)$$

$$\mathbf{q} = (\mathbf{v}, \omega) = (x, y, z, \omega) = [\sin(\theta/2) \times \mathbf{r}, \cos(\theta/2)]. \quad (8)$$

最后用四元数 $Q(3) \times \dots \times Q(3)$ 表示骨骼之间

的几何关系,每个骨骼表示成 7×182 的矩阵,对应的一个序列就是 $7 \times 182 \times n$,其中 n 是一个骨骼序列的帧数。

2.3 特征增强

获得四元数骨骼特征集后,由于序列中存在速率变化、时域不对齐、噪声等问题,不宜进行分类操作。参考文献[17],采用 DTW^[18] 来处理速率变化问题。同时,为了计算出标准的类特征,所有的训练样本都必须具有相同数量的采样,为此,在前面预处理阶段采用插值算法,将插值之后的骨骼序列映射到四元数上。为了解决时间偏差和噪声的问题,采用傅里叶时域金字塔来表示提取的四元数特征集^[16],从而去除高频系数,将傅里叶时域金字塔单独运用在每个维度并且连接所有的傅里叶系数来获得最终的骨骼特征向量(图4)。

3 实验结果

3.1 数据库介绍

为验证本文所提出方法的有效性,采用3个标准的行为数据库来进行测试:MSR-Action3D 数据库^[19]、UTKinect-Action 数据库^[9]和 Florence3D-Action 数据库^[20]。

MSR-Action3D 数据库是使用类似于 Kinect 的深度传感器采样获得的,由10个个体完成20类动作,每个个体执行每一类动作2~3次,总共有557个动作序列,每个序列提供了20个关节的3D位置。这个数据库的主要难点在于动作类型较多,且有些动作相似。

UTKinect-Action 数据库使用固定视角的 Kinect 深度传感器采样获得,由10个个体完成10类动作,每个个体执行每一类动作2次,总共有199个动作序列(数据库中有一动作序列丢失),每个序列提供了20个关节的3D位置。这个数据库的主要难点在于采样视角的变化和高内类变化。

Florence3D-Action 数据库使用固定视角的 Kinect 深度传感器采样获得,由10个个体完成9类动作,每个个体执行每一类动作2~3次,总共有215个动作序列,每个序列提供了15个关节的3D位置。这个数据库的主要难点在于高内类变化(对于某些相同操作,在某些序列上使用左手而在其他一些序列中使用右手)和相似动作(比如喝水和接电话)。

数据预处理。在 MSR-Action3D 和 UTKinect-

Action 数据库中,每个骨架都有19个部分和20个关节,而在 Florence3D-Action 数据库中,每个骨架只有14部分和15个关节。为了让人体骨骼数据的绝对位置不变,所有的3D关节坐标从世界坐标系转换为以髋关节为中心的坐标系。对于每个数据库,本文选择一个骨架作为参考,并规范化其他所有骨骼(不改变其关节角),这样其他的身体部分的长度等于相应的参考框架的长度。本文将骨架旋转至从左髋到右髋部的骨骼向量平行于坐标系 x 轴,使骨架视角不变。

3.2 参数设置

对于 MSR-Action3D 数据库,本文分别采用了文献[16]和文献[19]的交叉测试设置。对于 UTKinect-Action 和 Florence3D-Action 数据库,本文依据文献[21]进行交叉测试设置,即一半的个体用于训练,剩下的用于测试。本文使用每段1/4长度的三级傅里叶时域金字塔作为低频系数,支持向量机参数值 C 被设置为1。对于每一个数据库,所有的曲线 $Q(3) \times \dots \times Q(3)$ 被重新采样达到相同长度。参考长度是选择采样前的数据库样本的最大帧数。

3.3 实验结果

表1显示在 MSR-Action3D 数据库上各种骨骼表示的识别率,标*的表示数据库采用文献[19]的实验设置的平均识别率,其余为采用文献[16]的实验设置的平均识别率。可以看出,本文所提出的描述子优于其他常用的描述子,表明四元数描述子可以很好地描述关节之间的3D几何关系。实验设置按照文献[19]的混淆矩阵如图5所示,可以看出,主要识别错误在于高挥手、画圆和画X。在高挥手和看手表与画X之间存在部分识别错误,这是由于这些动作都是手臂左右来回摆动,即多次看手表或者画X将会和挥手的动作相似,其中画圆和画钩的整个过程很相似,画X和画钩只有部分序列不同,因而被识别错误的可能性较高。

表2显示在 MSR-Action3D 数据库上采用关键帧分类插值的四元数表示和仅用四元数表示的实验数据,识别率采用文献[19]的实验设置,时间采集的是生成测试和训练的特征描述子总时间。实验结果表明,采用关键帧分类插值方法比仅用四元数表示识别率提高了1.05%,说明通过分类关键帧插值可以准确地描述运动细节,有利于四元数的表示。

表 1 在 MSR-Action3D 数据库上的描述子比较

Table 1 Comparison of different descriptors different descriptors on MSR-Action3D dataset

Method	Recognition rate /%
Ref. [22]	91.21 *
Ref. [10]	92.03 *
Ref. [10]	92.20 *
Ref. [17]	92.49 *
Proposed	95.17 *
Ref. [16]	88.20
Ref. [23]	90.00
Ref. [17]	89.48
Proposed	91.88

表 2 MSR-Action3D 数据库上基于关键帧的识别结果

Table 2 Recognition rate based on key frames on MSR-Action3D dataset

Method	Recognition rate /%
Quaternion without classification of key frames	94.12
Quaternion with classification of key frames	95.17

表 3 显示在 UTKinect-Action 和 Florence3D-Action 数据库上本文方法与其他方法的识别率。在 UTKinect-Action 数据库上,本文提出的描述子的识别率比文献[22]描述子高 8.99%,比文献[24]描述子高 2.99%,比文献[17]描述子高 0.41%。这些结果表明,本文提出的骨骼描述子优于现有的其

他描述子,四元数在描述骨架中骨骼之间的关系时优于旋转矩阵,且需要的参数较少。

表 3 在 UTKinect-Action 和 Florence3D-Action 数据库上不同描述子的识别效果比较

Table 3 Comparison of different descriptors on database of UTKinect-Action and Florence3D-Action

Dataset	Method	Recognition rate /%
UTKinect-Action dataset	Ref. [22]	88.50
	Ref. [24]	94.50
	Proposed	97.08
Florence3D-Action dataset	Ref. [20]	82.00
	Ref. [25]	87.00
	Ref. [17]	90.88
	Proposed	92.14

在 Florence3D-Action 数据库上的结果显示,本文方法的识别率比文献[20]方法高 10.14%,比文献[23]方法高 5.14%,比文献[17]方法高 1.26%。这表明基于四元数的描述子对骨架中骨骼之间 3D 几何关系的描述更为准确。图 6 显示了 Florence3D-Action 数据库对应的混淆矩阵,错误主要集中在喝水、回电话和看手表这 3 个动作上,喝水被看作回电话或回电话被看作喝水的识别错误率较高,这是由于两者唯一的区别在于手最后放置的位置不同,动作的相似度极高,导致了识别错误的发生。

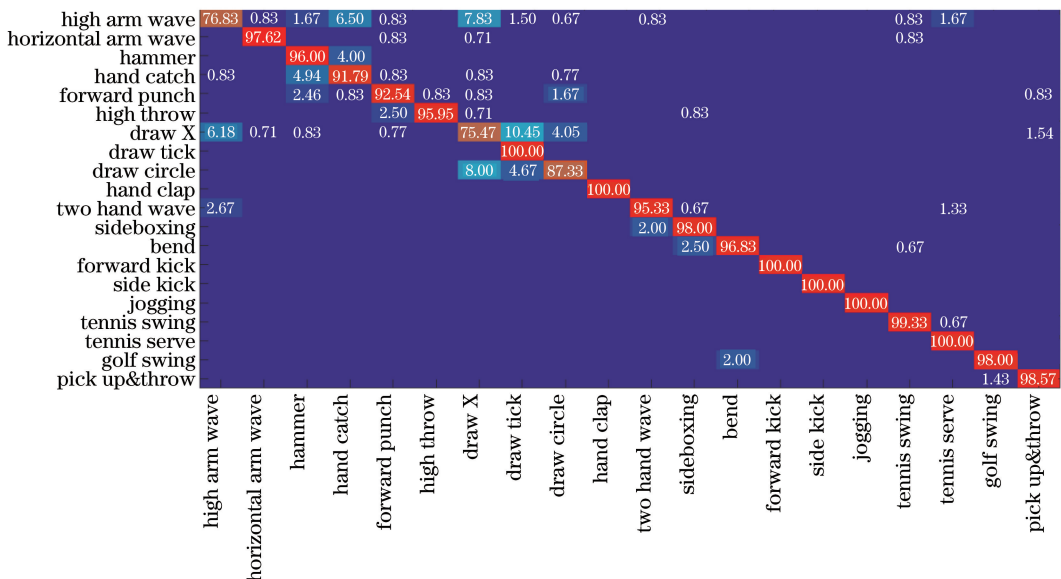


图 5 MSR-Action3D 数据库的混淆矩阵

Fig. 5 Confusion matrix of MSR-Action3D database

最近在 MSR-Action3D 数据中,一些方法^[20,22]的识别率大约 94.5%,是通过结合骨骼特性与额外特性

取得的。本文的重点不在结合多个特性上,因而只使用基于骨骼的结果与文献[20,22]进行比较。

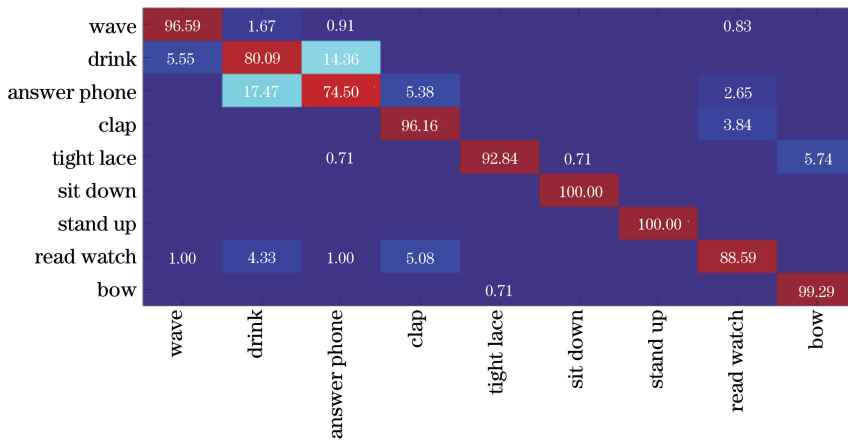


图6 Florence3D-Action 数据库的混淆矩阵

Fig. 6 Confusion matrix of Florence3D-Action database

4 结 论

提出一个将关键帧分类插值与四元数结合的方案用于人体行为识别。首先,对传感器采样获得的关键帧进行分类,分别进行相应的插值,从而获得具有相同帧数的新骨骼序列;然后,用四元数 $Q(3) \times \dots \times Q(3)$ 表示人体骨骼,通过旋转和平移的方式对各个身体部位之间显式地建模 3D 几何关系。使用 $Q(3) \times \dots \times Q(3)$ 表示一个行为过程,对每一帧进行四元数表示,最后形成一个向量空间,然后进行时序建模和分类。在 3 个数据库上的实验表明,该描述子优于许多现有的骨骼表示方法。在本文工作中,采用插值预处理方法详细地描述了人体动作姿势,同时,为了有效地提取骨骼姿势的细节信息,使用四元数描述子对所有成对骨骼间的相对几何关系进行描述;由于每个动作通常表现为一组特定身体部位的相互作用,在对所有姿势的所有骨骼进行描述时,存在局部相同的特征。因此,计划去探索能自动识别给定动作的身体部位的策略。

参 考 文 献

- [1] 陈战胜, 王辉, 贺萍, 等. 增强现实中跟踪预测方法的研究[J]. 河南科技大学学报(自然科学版), 2013, 34(2): 57-60.
- [2] 裴道方, 钟悦. 考虑行为和眼动跟踪的用户兴趣模型[J]. 河南科技大学学报(自然科学版), 2014, 35(1): 49-52.
- [3] Zhang X G, Liu C X, Zuo J Q. Small scale crowd behavior recognition based on causal network analysis [J]. Acta Optica Sinica, 2015, 35(8): 0815001. 张旭光, 刘春霞, 左佳倩. 基于因果网络分析的小规模人群行为识别[J]. 光学学报, 2015, 35(8):

0815001.

- [4] Cai J X, Feng G C, Tang X, *et al.* Human action recognition based on local image contour and random forest [J]. Acta Optica Sinica, 2014, 34(10): 1015006. 蔡加欣, 冯国灿, 汤鑫, 等. 基于局部轮廓和随机森林的人体行为识别[J]. 光学学报, 2014, 34(10): 1015006.
- [5] Cai J X, Feng G C, Tang X, *et al.* Human action recognition by leaning pose dictionary [J]. Acta Optica Sinica, 2014, 34(12): 1215002. 蔡加欣, 冯国灿, 汤鑫, 等. 基于姿势字典学习的人体行为识别[J]. 光学学报, 2014, 34(12): 1215002.
- [6] Johansson G. Visual perception of biological motion and a model for its analysis [J]. Perception Psychophysics, 1973, 14(2): 201-211.
- [7] Luvizon D C, Tabia H, Picard D. Learning features combination for human action recognition from skeleton sequences [J]. Pattern Recognition Letters, 2017, 99: 13-20.
- [8] Mahasseni B, Todorovic S. Regularizing long short term memory with 3D human-skeleton sequences for action recognition [C]. Computer Vision and Pattern Recognition, IEEE, 2016: 3054-3062.
- [9] Yacoob Y, Black M J. Parameterized modeling and recognition of activities [J]. Computer Vision and Image Understanding, 1999, 73(2): 232-247.
- [10] Li M, Leung H. Graph-based approach for 3D human skeletal action recognition [J]. Pattern Recognition Letters, 2017, 87: 195-202.
- [11] Li Y H, Zhu S A, Zhu L. Quaternion principal component analysis algorithm for face recognition [J]. Journal of Signal Processing, 2007, 23(2): 214-216. 黎云汉, 朱善安, 祝磊. 基于四元数主成分分析的人

- 脸识别算法[J]. 信号处理, 2007, 23(2): 214-216.
- [12] Xu Y. Quaternion-based discriminant analysis method for color face recognition [J]. Plos One, 2012, 7(8): e43493.
- [13] 曹林. 人脸识别与人体动作识别技术及应用[M]. 北京: 电子工业出版社, 2015.
- [14] Kong D H, Wang L C, Zheng C Y. A key frame interpolation method enhancing motion details of skeletal animation[J]. Journal of Beijing University of Technology, 2011(8): 1255-1261.
孔德慧, 王立春, 郑重雨. 增强骨骼动画运动细节的关键帧插值方法[J]. 北京工业大学学报, 2011(8): 1255-1261.
- [15] Xiao S B. The method of quaternion and its application[J]. Advances in Mechanics, 1993, 23(2): 249-260.
肖尚彬. 四元数方法及其应用[J]. 力学进展, 1993, 23(2): 249-260.
- [16] Wang J, Liu Z, Wu Y, *et al.* Mining actionlet ensemble for action recognition with depth cameras [C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2012: 1290-1297.
- [17] Vemulapalli R, Arrate F, Chellappa R. Human action recognition by representing 3D skeletons as points in a lie group [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 588-595.
- [18] Müller M. Information retrieval for music and motion [M]. New York: Springer Verlag, 2007.
- [19] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3D points [C]. Computer Vision and Pattern Recognition Workshops, IEEE, 2010: 9-14.
- [20] Seidenari L, Varano V, Berretti S, *et al.* Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses [C]. 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013: 479-485.
- [21] Zhu Y, Chen W, Guo G. Fusing spatiotemporal features and joints for 3D action recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE Computer Society, 2013: 486-491.
- [22] Slama R, Daoudi M, Daoudi M, *et al.* Accurate 3D action recognition using learning on the Grassmann manifold[J]. Pattern Recognition, 2015, 48(2): 556-567.
- [23] Ben A B, Su J, Srivastava A. Action recognition using rate-invariant analysis of skeletal shape trajectories [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(1): 1-13.
- [24] Ding W, Liu K, Cheng F, *et al.* Learning hierarchical spatio-temporal pattern for human activity prediction [J]. Journal of Visual Communication & Image Representation, 2016, 35(C): 103-111.
- [25] Devanne M, Wannous H, Berretti S, *et al.* 3-D human action recognition by shape analysis of motion trajectories on Riemannian manifold [J]. IEEE Transactions on Cybernetics, 2015, 45(7): 1340-1352.