

文章编号 1004-924X(2018)10-2584-08

# 利用几何特征和时序注意递归网络的动作识别

李庆辉\*, 李艾华, 郑 勇, 方 浩  
(火箭军工程大学 保障学院, 陕西 西安 710025)

**摘要:**为提高基于人体骨架(Skeleton-based)的动作识别准确度,提出一种利用骨架几何特征与时序注意递归网络的动作识别方法。首先,利用旋转矩阵的向量化形式描述身体部件对之间的相对几何关系,并与关节坐标、关节距离两种特征融合后作为骨架的特征表示;然后,提出一种时序注意方法,通过与之前帧加权平均对比来判定当前帧包含的有价值的信息量,采用一个多层感知机实现权值的生成;最后,将骨架的特征表示乘以对应权值后输入一个 LSTM 网络进行动作识别。在 MSR-Action3D 和 UWA3D Multiview Activity II 数据集上该方法分别取得了 96.93% 和 80.50% 的识别结果。实验结果表明该方法能对人体动作进行有效地识别且对视角变化具有较高的适应性。

**关键词:**动作识别;部件相对几何关系;时序注意;LSTM 神经网络

**中图分类号:**TP391 **文献标识码:**A **doi:**10.3788/OPE.20182610.2584

## Action recognition using geometric features and recurrent temporal attention network

LI Qing-hui\*, LI Ai-hua, ZHENG Yong, FANG Hao

(Academy of Operational Support, Rocket Force Engineering University, Xi'an 710025, China)

\* Corresponding author, E-mail: lqhuiul212@126.com

**Abstract:** To improve the accuracy of action recognition based on the human skeleton, an action recognition method based on geometric features and a recurrent temporal attention network was proposed. First, a vectorized form of the rotation matrix was defined to describe the relative geometric relationship between body parts. The vectorized form was fused with joint coordinates and joint distances to represent a skeleton in a video. A temporal attention method was then introduced. By considering the weighted average of the previous frame, a multi-layer perceptron was used to learn the weight of the current frame. Finally, the product of the feature vector and corresponding weight was propagated through three layers of long short-term memory to predict the class label. The experimental results show that the recognition accuracy of the proposed algorithm was superior to that of existing algorithms. Specifically, experiments with the MSR-Action3D and UWA3D Multiview Activity II datasets achieved 96.93 and 80.50% accuracy, respectively.

**Key words:** action recognition; relative geometry of body parts; temporal attention modal; LSTM neural network

收稿日期:2018-02-06;修订日期:2018-04-03.

基金项目:国家自然科学基金资助项目(No. 61501470);陕西省重点研发计划资助项目(No. 2017GY-075)

## 1 引言

人体动作识别在人机交互、智能监控、视频检索和视频理解等领域具有广泛的应用价值,已经成为计算机视觉和模式识别领域的研究热点<sup>[1-2]</sup>。计算机通过动作识别技术可以自动地理解和描述视频中的人体动作,从而实现底层视频数据与高层语义的关联。由于人体骨架的高层次表达能力以及对视角、表观和背景干扰的鲁棒性,基于骨架的动作识别方法近年来受到了广泛地关注<sup>[3]</sup>。此外,Kinect、RealSense 等高性价比深度相机的普及和骨架实时估计算法的成熟也极大促进了这项技术的发展<sup>[4]</sup>。

基于骨架的动作识别方法将人体视为由关节连接身体部件的铰接系统,并利用该系统的空间结构变化来描述运动。因此,利用原始骨架的关节、部件参数建立有效的特征表示是视频动作识别的前提。Rahmani 等<sup>[5]</sup>在全局坐标系下以骨架中心点坐标为参考定义了空间关节位移向量,在此基础上利用关节运动体特征(Joint Motion Volume)编码不同关节对动作识别的重要程度;Yang 等<sup>[6]</sup>基于关节位置变化提出一种融合静态姿势、运动和偏移特征的本征关节(EigenJoints),并采用累积运动能量(Accumulated Motion Energy, AME)筛选骨架序列;Boubou 等<sup>[7]</sup>通过计算关节的运动方向建立了方向速度向量直方图(HOVV)作为骨架表示;Zhang 等<sup>[8]</sup>采用递归神经网络作为时序数据建模方法,提取一组简单的几何特征(如关节坐标、关节距离、关节所构成的平面间角度等)作为网络输入。

获得骨架的特征表示后,传统方法通常采用傅里叶时间金字塔(FTP)、动态时间规整(DTW)或隐马尔科夫模型(HMM)建模骨架序列的时间动态。递归神经网络(Recurrent Neural Net, RNN)由于引入了定向循环,能够处理输入之间前后关联的问题,在多项时序数据处理任务中取得了优于传统方法的效果,如自然语言理解、动作识别和视频摘要等。Du 等<sup>[9]</sup>根据人体物理结构将骨架划分成 5 个部位,分别利用一个 RNN 建模低层身体部件表示,然后通过分层递归神经网络不断融合成骨架整体的表示,最后输入单层感知机进行分类;Zhu 等<sup>[10]</sup>设计了一种端到端全连

接的深度 LSTM(Long Short-Term Memory)网络,采用新的正则化方法对骨架关节的共生特征进行学习;Liu 等<sup>[11]</sup>在 LSTM 中引入信任门(Trust Gate)来学习顺序数据的可靠性,并相应地更新存储单元内保存的长时上下文信息。

研究表明,结合注意力机制的递归神经网络模型能够高度关注与识别任务有关的高辨识性信息,挖掘输入序列中更多的有用信息<sup>[12]</sup>。基于上述思想,本文提出一种利用骨架几何特征与时序注意递归网络的动作识别方法。算法首先利用旋转矩阵的向量化形式描述身体部件对之间的相对几何关系,再与骨架关节坐标、关节距离两种特征级联作为人体骨架的特征向量;然后计算当前帧与之前帧加权平均的差值,作为一个多层感知机的输入来学习当前帧的权值,权值大小反映了该帧信息对识别任务的重要程度;最后将每帧骨架的特征向量乘以对应权值后输入一个 LSTM 网络进行动作识别。

## 2 特征提取

### 2.1 数据预处理

原始骨架序列的所有数据记录在以相机原点的全局坐标系下,相机视角的变化会导致人体骨架呈现不同的形态,从而影响动作识别的结果。为解决这个问题,将骨架序列转换到以人为中心的坐标系中<sup>[13]</sup>,实现视角无关(View-Invariant)。新坐标系以左右髋关节连线中心为原点, $x$ 轴平行于髋关节连线在  $xOy$  平面的投影。选定其中一帧骨架为参考对其他骨架进行归一化,这种归一化操作确保其他骨架的身体部件长度与参考骨架对应相等,实现尺度无关(Scale-Invariant)。

### 2.2 骨架表示

定义  $S=(J, P)$  为一个人体骨架,其中  $J=\{j_1, \dots, j_M\}$  表示关节,  $P=\{p_1, \dots, p_N\}$  表示连接关节的部件,图 1 为一个包含 20 个关节和 19 个部件的人体骨架示例。本文提取每帧骨架的 3 种几何特征,级联后作为该帧骨架的特征向量:

(1) 关节坐标 (Joint Coordinate),  $J_c(j_i) = (x_i, y_i, z_i)$ ;

(2) 关节距离 (Joint-Joint Distance),  $J_d(j_i, j_j) = \|j_i j_j\| = |x_j - x_i, y_j - y_i, z_j - z_i|, 1 \leq i < j \leq M$ ;

(3) 部件相对几何关系 (Relative Geometry of Body Parts),  $P_{rg}(p_i, p_j), 1 \leq i < j \leq N$ .

人体动作可以被解释为不同身体部件的相互作用<sup>[8]</sup>, 提取部件之间的相对几何关系特征可以

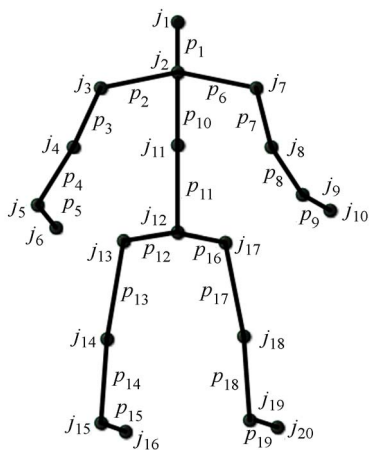


图 1 人体骨架示例

Fig. 1 An example of human skeleton

有效描述动作。给定两个身体部件  $p_i$  和  $p_j$  的位置, 如图 2 所示, 可以认为  $p_j$  是由  $p_i$  经过平移、旋转而来, 先将  $p_i$  平移至与  $p_j$  起点重合位置得到  $p_i'$  再旋转角度  $\theta$  可以得到  $p_j$ , 因此可用旋转矩阵来描述部件  $p_i$  和  $p_j$  的相对几何关系  $P_{rg}(p_i, p_j)$ 。

本文以转轴-角度方法表示旋转矩阵  $R_{i,j}$ , 首先以  $p_i$  的起点为原点、以  $p_i'$  的方向为  $x$  轴正方向建立局部坐标系, 再计算旋转轴的单位向量  $\omega(x_\omega, y_\omega, z_\omega)$  和旋转角度  $\theta$ 。由旋转过程可知, 旋转角所在的平面为  $p_i'$  和  $p_j$  所构成的平面, 那么旋转轴必垂直于该平面, 如公式(1)、(2)所示:

$$\omega = \frac{p_i \times p_j}{\|p_i \times p_j\|}, \quad (1)$$

$$\theta = \arccos(p_i \cdot p_j). \quad (2)$$

根据罗德里格旋转公式 (Rodrigues' rotation formula) 可得旋转矩阵  $R_{i,j}$ :

$$R_{i,j}(\omega, \theta) = I + (1 - \cos \theta) \begin{pmatrix} x_\omega \\ y_\omega \\ z_\omega \end{pmatrix} \begin{pmatrix} x_\omega & y_\omega & z_\omega \end{pmatrix} + \sin \theta \begin{pmatrix} 0 & -z_\omega & y_\omega \\ z_\omega & 0 & -x_\omega \\ -y_\omega & x_\omega & 0 \end{pmatrix} = \begin{pmatrix} \cos \theta + x_\omega^2(1 - \cos \theta) & x_\omega y_\omega(1 - \cos \theta) - z_\omega \sin \theta & y_\omega \sin \theta + x_\omega z_\omega(1 - \cos \theta) \\ z_\omega \sin \theta + x_\omega y_\omega(1 - \cos \theta) & \cos \theta + y_\omega^2(1 - \cos \theta) & -x_\omega \sin \theta + y_\omega z_\omega(1 - \cos \theta) \\ -y_\omega \sin \theta + x_\omega z_\omega(1 - \cos \theta) & x_\omega \sin \theta + y_\omega z_\omega(1 - \cos \theta) & \cos \theta + z_\omega^2(1 - \cos \theta) \end{pmatrix}. \quad (3)$$

公式(3)可以进一步简化为:

$$R_{i,j}(\omega, \theta) = \exp \begin{pmatrix} 0 & -z_\omega \theta & y_\omega \theta \\ z_\omega \theta & 0 & -x_\omega \theta \\ -y_\omega \theta & x_\omega \theta & 0 \end{pmatrix}. \quad (4)$$

由公式(4)可知, 对于任意两个身体部件之间的旋转矩阵, 指数部分是固定形式的斜对称矩阵, 并且只与旋转轴的单位向量  $\omega(x_\omega, y_\omega, z_\omega)$  和旋转角度  $\theta$  有关, 因此本文直接取旋转矩阵  $R_{i,j}$  的向量化形式为:

$$\text{vec}(R_{i,j}) = [x_\omega \theta, y_\omega \theta, z_\omega \theta]. \quad (5)$$

部件  $p_i$  和  $p_j$  的相对几何关系可以表达为 3 维向量  $P_{rg}(p_i, p_j) = \text{vec}(R_{i,j}) = [x_\omega \theta, y_\omega \theta, z_\omega \theta]$ 。对于一个包含  $N$  个身体部件的骨架, 所有部件对的相对几何关系组成  $3N(N-1)/2$  维的特征向量  $[P_{rg}(p_1, p_2), P_{rg}(p_1, p_2), \dots, P_{rg}(p_{N-1}, p_N)]$ 。

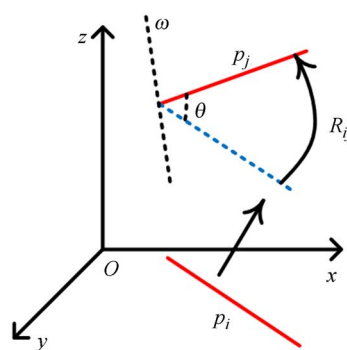


图 2 部件相对几何关系描述

Fig. 2 Relative geometry of body parts

### 3 时序注意递归网络结构

#### 3.1 时序注意

对于一段动作骨架序列, 不同帧包含的有价

值的信息量是不相等的,通常只有部分帧提供最具辨识性的信息,而其余帧提供上下文信息<sup>[14]</sup>。研究发现,人类通过观察部分关键帧即可识别骨架序列中的动作,所以序列中存在着大量的冗余帧。时序注意机制通过给不同的帧分配不同的权重引导分类器高度关注序列中包含强辨识性信息的帧。本文认为这些包含强辨识性信息的帧应当是与之前帧存在较大差异的帧,并通过比较当前帧与之前帧的加权平均来确定这种差异,采用一个多层感知机(Multilayer Perceptron, MLP)来实现权值的生成。最后将每帧骨架的特征向量乘以对应权值后输入一个 LSTM 网络进行动作识别,网络架构如图 3 所示。

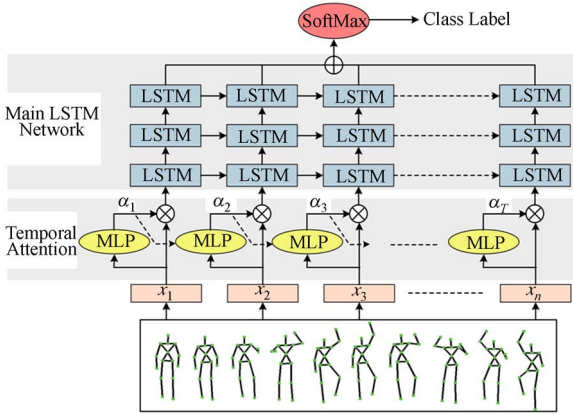


图 3 时序注意 LSTM 网络架构

Fig. 3 Architecture of LSTM network with temporal attention

设级联 3 种特征向量后得到的骨架序列表示为  $\mathbf{X}=[x_1, x_2, \dots, x_T]$ ,对每帧骨架分别分配一个权重  $\alpha_i \in [0, 1]$ ,权重大小取决于该帧骨架包含的对动作识别任务有价值的信息量。采用一个 3 层 MLP 学习生成权重,网络由输入层、隐含层和输出层组成。对于第  $t$  帧骨架,网络输入为当前帧特征向量  $x_t$  与之前帧加权平均  $\mathbf{I}(t-1)$  的差,隐含层和输出层的激活函数分别为  $\tanh(\cdot)$  和  $\text{sigmoid}(\cdot)$ ,则:

$$\beta_t = \tanh\{W_1[x_t - \mathbf{I}(t-1)] + b_1\}, \quad (6)$$

$$\alpha_t = \sigma(W_3 \beta_t + b_2), \quad (7)$$

其中:  $W_1$  和  $b_1$ 、 $W_2$  和  $b_2$  分别为隐含层、输出层的权重矩阵和偏置项,  $\beta_t$  为隐含层的输出向量,  $\sigma$  为 sigmoid 激活函数。生成权重后,可以利用递归公式(8)计算  $\mathbf{I}(t)$  用于计算下一帧骨架的权重  $\alpha_{t+1}$ 。

$$\mathbf{I}(t) = \frac{\alpha_t x_t + \mathbf{I}(t-1) \sum_{i=1}^{t-1} \alpha_i}{\sum_{i=1}^t \alpha_i}. \quad (8)$$

### 3.2 LSTM 网络

传统 RNN 在利用 BP 算法进行训练时会出现梯度消失现象, LSTM 的引入解决了这个问题。如图 4 所示, LSTM 使用记忆单元  $c_t$  存储和输出信息,能够学习输入序列的长期依赖关系。LSTM 单元的输入为当前数据  $x_t$  和上一时刻输出状态  $h_{t-1}$ ,并通过输入门(Input Gate)、遗忘门(Forget Gate)、输出门(Output Gate)3 个门来选择性添加当前信息和删除历史信息,单元输出为当前时刻状态  $h_t$ 。LSTM 的计算原理如公式(9)~(14)所示:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_i), \quad (9)$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f), \quad (10)$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o), \quad (11)$$

$$g_t = \sigma(W_{xc} x_t + W_{hc} h_{t-1} + b_c), \quad (12)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t, \quad (13)$$

$$h_t = o_t \circ \tanh(c_t), \quad (14)$$

式中:  $i_t$ 、 $f_t$  和  $o_t$  分别为 LSTM 单元  $t$  时刻输入门、遗忘门和输出门的输出,  $\sigma$  为 sigmoid 激活函数,  $W_*$  和  $b_*$  分别为对应门的权重矩阵和偏置项,表示向量对应位置元素乘运算。

为充分挖掘多层网络的优势,本文使用了一个 3 层 LSTM 网络,如图 4 所示。第 1 层以每帧骨架的特征向量与对应权值的乘积为输入,第 2、3 层分别以其下一层的输出  $h_t$  为输入,最后将第 3 层的输出送入一个 softmax 层进行分类识别。这种多层结构可以学习输入序列更长时间的依赖关系。

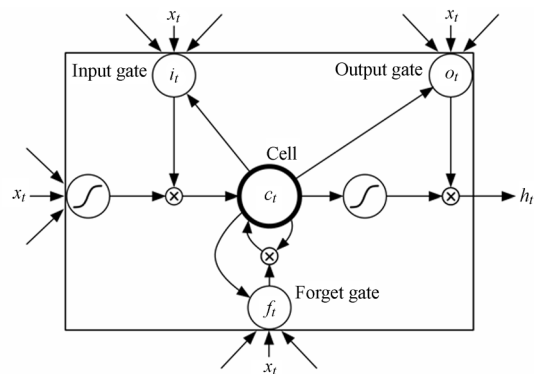


图 4 LSTM 单元结构图

Fig. 4 Schema of a long short-term memory (LSTM) unit

### 3.3 损失函数

本文利用一个交叉熵损失 (Cross Entropy Loss) 项和两个正则化项组成时序注意 LSTM 网络的损失函数, 定义如下:

$$L = - \sum_{i=1}^C y_i \log \hat{y}_i + \frac{\lambda}{T} \sum_{i=1}^T |\alpha_i| + \gamma \|T\|_2, \quad (15)$$

式中:  $\mathbf{y} = (y_1, \dots, y_C)^T$  为实际动作类别标签 (One-Hot Label Vector),  $\hat{y}_i$  表示序列被预测为第  $i$  类的概率,  $C$  为动作类别总数,  $\lambda$  为注意力惩罚系数,  $\gamma$  为权值衰减系数,  $T$  代表模型所有参数。第 2 项是一个 L1 正则化项, 目的是产生稀疏权值矩阵, 确保模型尽可能关注骨架序列中最具判别力的帧; 第 3 项是一个 L2 正则化项, 可以防止模型过拟合 (Overfitting)。

## 4 实验结果与分析

### 4.1 实验设置

本文在 MSR-Action3D 和 UWA3D Multiview Activity II 数据集上验证提出的动作识别方法。实验中, 注意力惩罚系数、权值衰减系数  $\gamma$  分别取  $1e-3$ 、 $5e-4$ , 网络训练过程中采用 Adam 自动调整学习速率<sup>[15]</sup>, 加权平均  $\mathbf{I}(t)$  的初始值取为第一帧的特征向量。

为充分验证算法性能, 设计 3 种实验方案。第 1 种 LSTM(JC+JJD) 作为对比基准, 特征采用关节坐标 (JC) 和关节距离 (JJD) 两种, 网络采用常规 LSTM 网络; 第 2 种 LSTM(JC+JJD+RGBP) 验证部件相对几何关系 (RGBP) 特征的作用; 第 3 种 TA-LSTM(JC+JJD+RGBP) 进一步验证本文提出的时序注意 (Temporal Attention) 机制的效果。

### 4.2 MSR-Action3D 的结果与分析

MSR-Action3D 数据集<sup>[16]</sup> 由类似于 Kinect 的深度传感器捕获, 包含 20 类共 557 个动作序列, 每类动作分别由 10 个对象执行 2~3 次。每帧骨架包含 20 个关节, 并提供了每个关节的三维坐标。数据集的主要挑战性在于许多行为是高度相似的。

MSR-Action3D 数据集上的实验采用文献 [16] 中提供的标准评价方案, 将数据集分成 3 组: Action Set1 (AS1)、Action Set2 (AS2) 和 Action

Set3 (AS3)。对象 1、3、5、7、9 的动作视频用于训练, 而对对象 2、4、6、8、10 的动作视频用于测试。

表 1 MSR-Action3D 数据集结果对比

Tab. 1 Results comparison on MSR-Action3D dataset (%)

方法	AS1	AS2	AS3	Ave.
Ref. [17]	87.98	85.48	63.46	78.97
Ref. [18]	95.29	83.87	98.22	92.46
Ref. [9]	93.33	94.64	95.50	94.49
Ref. [11]	—	—	—	94.8
LSTM (JC+JJD)	75.87	76.82	77.46	76.72
LSTM (JC+JJD+RGBP)	91.38	89.18	91.79	90.78
TA-LSTM (JC+JJD+RGBP)	97.62	94.63	98.54	96.93

表 1 为 MSR-Action3D 数据集的实验结果对比, 本文算法的平均识别准确率为 96.93%, 较对比算法的最好结果 Ref. [11] 提高了 2.13%。由表 1 可知, 加入部件相对几何关系 (RGBP) 特征后平均准确率提高了 14.06%, 验证了部件相对几何关系特征能够有效描述人体骨架的空间结构; 在 LSTM 网络中加入本文的时序注意机制后, 平均准确率进一步提高了 6.15%, 说明本文提出的时序注意方法能够引导网络有效利用骨架序列中的强辨识度信息并排除冗余信息的干扰, 提升行为识别准确率。

此外, 本文算法在 3 组动作集上均取得了最优 (AS1、AS3) 或与最优相近的结果 (AS2), 验证了算法的适应性。

图 5 为本文算法在 AS1、AS2 和 AS3 上识别结果的混淆矩阵。错误分类主要发生在几个非常相似的动作之间。如图 5(a) 所示, 在 AS1 上 “Pickup & throw” 和 “Bend” 两个动作序列之间重叠较多, 误分类情况最严重, 类似的动作还有 “Forward-punch” 和 “Tennis serve”、“Hammer” 和 “High throw” 等; 如图 5(b) 所示, 在 AS2 上存在误分类情况的动作主要是 “Draw X” 和 “Draw tick”、“Hand catch” 和 “High arm wave”; 如图 5(c) 所示, 在 AS3 上 “High throw” 动作会被误分类为 “Tennis serve” 动作, 这两个动作的骨架空间结构高度相似, 即便对于人类自身也较难区分。



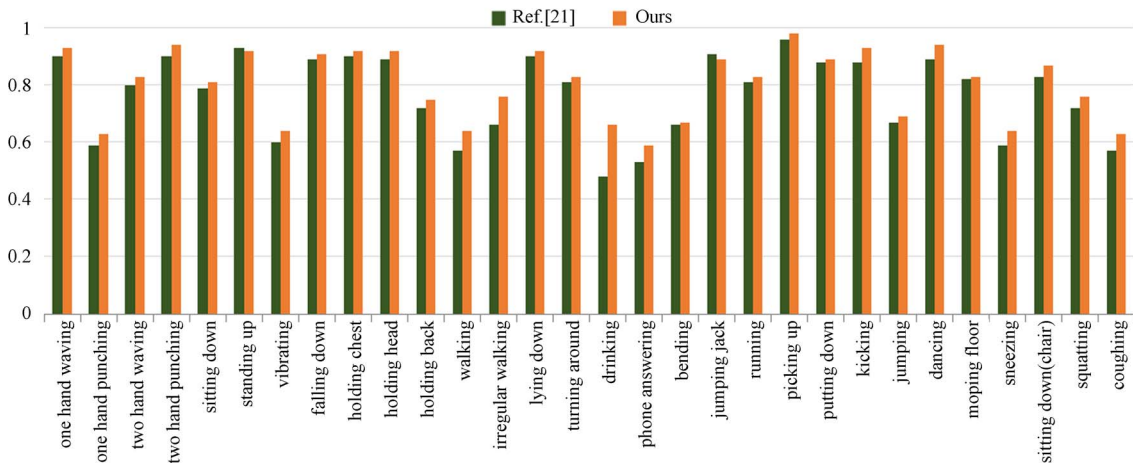


图 6 本文算法与 Ref. [22] 算法的类识别准确率

Fig. 6 Per class recognition accuracy of our method and Ref. [22] on the UWA3D Multiview ActivityII dataset

## 5 结 论

本文提出一种基于几何特征与时序注意递归网络的动作识别方法。算法首先利用旋转矩阵的向量化形式描述身体部件对之间的相对几何关系,并与骨架关节坐标、关节距离两种特征融合作为骨架的特征表示;然后引入计算当前帧与之前

帧加权平均的差值,作为一个多层感知机的输入来学习当前帧的权值;最后将每帧骨架的特征向量乘以对应权值后输入一个 LSTM 网络进行动作识别。本文算法在 MSR-Action3D 和 UWA3D Multiview Activity II 数据集上分别取得了 96.93% 和 80.50% 的识别结果,并在多数单类动作识别上超过现有算法,验证了算法的有效性和适应性。

### 参考文献:

- [1] 王世刚,鲁奉军,赵文婷,等. 应用在线随机森林投票的动作识别[J]. 光学 精密工程, 2016, 24(8): 2010-2017.  
WANG SH G, LU F J, ZHAO W T, *et al.*. Action recognition based on on-line random forest voting [J]. *Opt. Precision Eng.*, 2016, 24(8):2010-2017. (in Chinese)
- [2] 刘智,黄江涛,冯欣. 构建多尺度深度卷积神经网络行为识别模型[J]. 光学 精密工程, 2017, 25(3): 799-805.  
LIU ZH, HUANG J T, FENG X. action recognition model construction based on multi-scale deep convolution neural network [J]. *Opt. Precision Eng.*, 2017, 25(3):799-805. (in Chinese)
- [3] HAN F, REILY B, HOFF W, *et al.*. Space-time representation of people based on 3D skeletal data [J]. *Computer Vision & Image Understanding*, 2017,158(C):85-105.
- [4] 李庆武,席淑雅,王恬,等. 结合位姿约束与轨迹寻优的人体姿态估计[J]. 光学 精密工程, 2017, 25(4):528-537.  
LI Q W, XI SH Y, WANG T, *et al.*. Human pose estimation based on configuration constraints and trajectory optimization [J]. *Opt. Precision Eng.*, 2017, 25(4): 528-537. (in Chinese)
- [5] RAHMANI H, MAHMOOD A, HUYNH D Q, *et al.*. Real time action recognition using histograms of depth gradients and random decision forests [C]. *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2014:626-633.
- [6] YANG X, TIAN Y L. Effective 3D action recognition using EigenJoints [J]. *Journal of Visual Communication & Image Representation*, 2014, 25(1): 2-11.
- [7] BOUBOU S, SUZUKI E. Classifying actions based on histogram of oriented velocity vectors [J]. *Journal of Intelligent Information Systems*, 2015, 44(1): 49-65.
- [8] ZHANG S, LIU X, XIAO J. On geometric features for skeleton-based action recognition using multilayer lstm networks [C]. *Proceedings of IEEE Win-*

- ter Conference on Applications of Computer Vision, Los Alamitos: IEEE Computer Society Press, 2017: 148-157.
- [9] DU Y, WANG W, WANG L. Hierarchical recurrent neural network for skeleton based action recognition [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, USA: IEEE Press, 2015: 1110-1118.
- [10] ZHU W, LAN C, XING J, et al. Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks [C]. *AAAI Conference on Artificial Intelligence*, Palo Alto, USA: AAAI Press, 2016: 3697-3703.
- [11] LIU J, SHAHROUDY A, XU D, et al.. Spatio-temporal LSTM with trust gates for 3D human action recognition [C]. *Proceedings of the European Conference on Computer Vision*, Heidelberg: Springer, 2016:816-833.
- [12] SHARMA S, KIROS R, SALAKHU R. Action Recognition using Visual Attention [C]. *Proceedings of the International Conference on Learning Representations*, 2016:1-11.
- [13] LEE I, KIM D, KANG S, et al.. Ensemble deep learning for skeleton-based action recognition using temporal sliding LSTM networks [C]. *Proceedings of the IEEE International Conference on Computer Vision*, Los Alamitos: IEEE Computer Society Press, 2017:1012-1020.
- [14] KAR A, RAI N, SIKKA K, et al.. Adascan: Adaptive scan pooling in deep convolutional neural networks for human action recognition in videos [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2017: 3376-3385.
- [15] SONG S, LAN C, XING J, et al.. An end-to-end spatio-temporal attention model for human action Recognition from Skeleton Data [C]. *AAAI Conference on Artificial Intelligence*, Palo Alto, USA: AAAI Press, 2017: 4263-4270.
- [16] LI W, ZHANG Z, LIU Z. Action recognition based on a bag of 3d points [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2010: 9-14.
- [17] XIA L, CHEN C C, AGGARWAL J K. View invariant human action recognition using histograms of 3D joints [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2012: 20-27.
- [18] VEMULAPALLI R, ARRATE F, CHELLAPPA R. Human action recognition by representing 3D skeletons as points in a Lie group. [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2014:588-595.
- [19] RAHMANI H, MAHMOOD A, HUYNH D, et al.. Histogram of oriented principal components for cross-view action recognition [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2016, 38(12):2430-2443.
- [20] RAHMANI H, MAHMOOD A, HUYNH D, et al.. HOPC: Histogram of oriented principal components of 3D point clouds for action recognition [C]. *European Conference on Computer Vision*, 2014: 742-757.
- [21] WANG J, LIU Z, WU Y, et al.. Learning actionlet ensemble for 3D human action recognition [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2014, 36(5): 914-927.
- [22] RAHMANI H, MIAN A. 3d action recognition from novel viewpoints [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2016: 1506-1515.

#### 作者简介:



李庆辉(1989—),男,山东临沂人,博士研究生,2011年、2013年于第二炮兵工程大学分别获得学士、硕士学位,主要从事机器视觉及模式识别方面的研究。  
E-mail: mailto: brightlishi @ gmail. comlqhui1212@126. com

#### 导师简介:



李艾华(1966—),男,河北沧州人,博士,教授,博士生导师,1987年、1990年于第二炮兵工程学院分别获得学士、硕士学位,1998年于西安交通大学获得博士学位,主要从事机电设备故障诊断、自动控制、机器视觉等方面的研究。  
E-mail: ahli66@126. com