

指导教师： 杨涛

提交时间： 2015/3/28

The task of
Digital Image Processing

数字图像处理

School of Computer Science

No: 01

姓名： 王科

学号： 2012302536

班号： 10011207

基于李氏群点的三维骨架人体动作识别

Raviteja Vemulapalli, Felipe Arrate, Rama Chellappa

马里兰大学帕克分校

自动化研究中心, 高级计算机研究所

摘要

近来提出的具有成本效益的深度传感器及Shotton等人【16】早已提出的实时骨架估测算法使得基于骨架的人体行为检测再一次成为人们研究的热点。目前大多数基于骨架的处理方法或者采用关节位置, 或者采用关节的旋转角度去构造人体的骨架模型。在本文中, 我们提出了一种新的骨架表示方法, 该模型通过旋转和空间三维变换清楚地表示了人体不同部位的三维几何模型。既然三维刚体运动被认为是一种特殊的欧几里得群 $SE(3)$, 我们提出的流行面骨架表示法基于李氏群

$SE(3) \times \dots \times SE(3)$ 产生。使用这种表示方法, 人体动作可以基于李氏群的曲线建模。由于基于李氏群理论的曲线分类不是一项容易的工作, 所以我们将李氏群中的人体骨架曲线映射为李代数中的空间向量。然后采用动态时间规划、傅里叶时空金字塔表示及特征向量相结合的方法进行分类。三个动作数据集的实验结果表明我们提出的这种方法比以往的所有的骨架表示法都要好。这种方法甚至已经超过了目前最先进的基于骨架的人体动作识别方法。

1. 简介

由于人体行为识别在监控、电子游戏、机器人方面的广泛应用, 在过去的数十年间, 这一方面的研究一直很热。许多基于单眼的RGB视频序列【1】的人体动作识别方法在过去的是十年间相继被提出。不幸的是, 基于单眼RGB模型的数据对诸如光线变化、视角变化、遮挡和背景杂斑非常敏感。甚至单目视频传感器不能完全在三维空间中去拍摄人体运动。因此, 忽略人们在过去数十年间的研究工作, 人体动作识别仍然是一个非常具有挑战性的问题。

人体可以视为通过关节连接的刚性段组

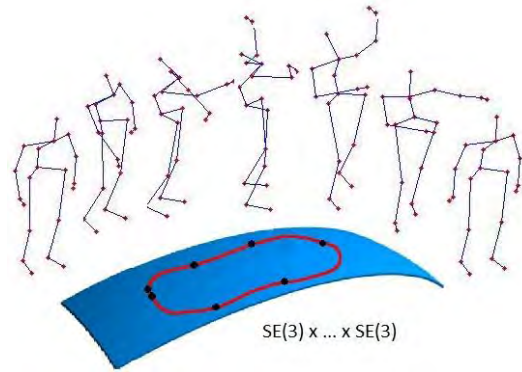


图 1: 用李氏群 $SE(3) \times \dots \times SE(3)$ 中的曲线表示一个人体骨架序列。

成的铰链系统, 并且人体的移动可以被视为是这些刚性段在空间中的连续变化【24】。因此, 如果我们可靠地捕获并提取出人体骨架, 那么通过对骨架在时间维的分类将会识别出人体动作。但是, 通过单目RGB视频可靠地提取出人体骨架是非常困难的一项工作【9】。

复杂的运动捕获系统可以捕获人体的三维标定点坐标。但是, 这种系统不仅非常昂贵, 而且需要使用者身着可以标定人体移动, 但严重影响行动的服装。随着具有成本效益的深度传感器的出现, 捕获人体骨架变得相对容易起来。这些对传感器可以提供对光照变化非常鲁棒的场景三维深度和其他一些对重建人体三维骨架模型非常有用的信息。最近, Shotton等人【16】提出了通过一幅三维深度图像可以快速并且准确地估计人体关节的三维位置的方法。这些最新的研究致使基于人体骨架的动作识别又一次热起来。

目前基于骨架的人体行为识别可以大致分为两类: 基于关节和基于身体部位。受到Johansson【6】经典的光线变化实验的启发, 基于关节的研究方法认为人体骨架就是一系列简单点的集合。这种方法试图去使用关节点的多种显著特征【5, 8】, 通过已知坐标轴确定关节方向【20】, 成对关节的相对位置【19, 22】等方法去建立个别的关节或一个关节组。

另一方面，基于身体部位的方法认为人体骨架是一系列相互联系的刚体段（身体部分）。这些方法要么通过个人的身体部位【21】随时间的变化建模，要么（直接）着眼于身体的连接部位并通过关节之间的夹角随时间的变化建模【12, 13】。

在本文中，为了实现动作识别，我们提出了全新基于身体部件的骨架表示法。通过观察人体的活动受到启发，我们发现不同身体部位（尽管不直接通过关节相连）之间的相对几何关系比它们之间的绝对位置（通过两个手掌之间的相对关系可以更直观地描述鼓掌）提供了一个更有意义的描述方法。通过不同身体部位之间的相对三维几何关系我们可以清楚地对骨架表示进行建模。给我们两个刚性身体部位，它们之间的相对几何关系就可以通过旋转和必要的平移来描述（图3）。在数学上，刚性部位在三维空间的旋转和平移是一种特殊的欧几里得群 $SE(3)$ 【11】，并且这种欧几里得群是一种特殊的李氏群矩阵。因此，我们用 $SE(3)$ 中的一个点来表示一对身体部位之间的相对几何关系，并且用李氏群 $SE(3) \times \dots \times SE(3)$ 中的一个点来表示整个人体的骨架， \times 表示李氏群之间的直接相乘关系。

在我们所提的方法中，人体动作可以用李氏群 $SE(3) \times \dots \times SE(3)$ 中的曲线（图1）建模，并且动作识别可以通过对这些曲线进行分类来完成。请注意，李氏群 $SE(3) \times \dots \times SE(3)$ 是一个流行面，并且在该空间中对其进行分类也是一项很重要的工作。而且，类似于SVM和傅里叶时域建模分析的标准分类方法对这种曲面空间并不是直接适用的。为了克服这种困难，我们将运动曲线从 $SE(3) \times \dots \times SE(3)$ 映射到它对应的李代数 $se(3) \times \dots \times se(3)$ 中，并且在这个群中的元素处于正交空间。

不考虑正在被使用的骨架表示方法，由于比率的不同、时间的不同步、噪声等问题，对一个时间序列上的动作进行分类是一个很困难的问题。对每一个动作分类，为了控制比率的不同我们通过动态时间规整（DTW）【10】计算出了一个虚曲线，并且将所有的曲线变化到这个虚曲线上。为了控制时间的不同步和噪声问题，我们采用先前提到过的傅里叶时空金

字塔（FTP）【19】表示方法去描绘这个翘曲曲线。最终的分类方法就是采用FTP和线性SVM分类器。图4显示了这种方法的一个综述。

优点：1) 我们采用李氏群 $SE(3) \times \dots \times SE(3)$ 中的点表示人体骨架。这种表示方法通过旋转和平移清楚地对人体不同部位之间的三维几何关系进行建模。2) $SE(3) \times \dots \times SE(3)$ 是一种流形曲线，我们将所有的运动曲线从李氏群映射到相应的李代数中，并且在李代数中完成了时态化建模和分类。3) 我们希望通过在三种不同的数据集上的评估向你们展示出这种方法比大多数已经存在的骨架表示方法都要优秀。这三种数据集包括：MSR-Action3D数据集【7】，UTKinect-Action数据集【20】和Florence3D-Action数据集【14】。我们同时展示了这种方法超过了各种目前最先进的基于骨架的人体行为识别方法。

组织结构：我们在第2部分对一些文献中的方法进行了一个简短的回顾，在第3部分讨论了特殊的欧几里得群 $SE(3)$ 。在第4部分中展示了我们提出的基于骨架的表示方法，在第5部分中描绘了时态化建模和分类方法。我们在第6部分中展示了我们的实验结论，在第7部分对文章进行了总结。

2. 相关工作

在这部分中，我们简短地回顾了不同的基于骨架的人体行为识别方法。我们建议读者去【1】中寻找一些最近的基于视频的RGB方法的评论，去【23】中寻找基于深度匹配方法的评论。

目前的基于骨架的人体行为检测方法概括来说可以分为两个主要的类别：基于关节和基于身体部位。基于关节的研究方法认为人体骨架就是一系列点的集合，然而基于身体部位的方法认为人体骨架是一系列相互联系的刚体段。使用关节之间夹角的方法可以被认为是基于部分身体部位的，因为关节之间的夹角可以（直接）估测出连接身体部位的骨架之间的几何关系。

基于关节的方法：在【5】中提到人体骨架可以采用关节在三维空间中的位置来表示，

并且可以采用协方差算符之间的时间层次关系来对关节的运动轨迹进行建模。在【8】中提到一种类似的表示是使用隐马尔可夫模型（HMMs）。一组13个关节在四维XYZT空间的运动轨迹在【15】中被用来表示人体的动作，并且使用基于角度的视点无关相似性测量子空间对它们投影之间的仿射关系进行了对比。在【19】中，使用成对关节之间的相对位置关系来表示人体骨架，并且使用傅里叶系数层来对这种表示的时间变化进行建模。此外，使用多核学习方法选择出具有显著特征的关节组合的基于运动的方法也被使用过。在【22】中，使用初始框架，即关节之间的相对位置、关节的位移和偏移量来表示人体骨架。在低维空间使用朴素临近贝叶斯法则并结合主成分分析（PCA）的方法对运动进行分类。在【27】中使用了一种相似的随机森林的方法去代表骨架。在臀部中心建立坐标，依据三维关节的方向将他们的位置量化到直方图，这种视点不变的表示可以在【20】中看到。随着时间的推移，使用HMMs对这种视角不变表示方法进行建模。

基于身体部位的方法：在【21】中人体被分为五个不同的区域，而且使用各部分诸如水平、竖直平移，面内旋转等运动参数去表示人体行为活动。使用主数值分析法去表示一系列基础动作的线性组合，并且结合主成分分析系数对其进行分类。在【2】中，将人体骨架分层划分为更小的块，并且每一小块使用仿生形态特征表示。使用线性动态系统对这些仿生形态特性随着时间的变化进行建模。在【3】中使用三维关节夹角来表示人体骨架，并且使用DTW对这些夹角随着时间的变化进行比较。

在【12】中，一些信息量大的关节在每个取样点依据一些高度可判断的测量信息被自动选择出来，例如：它们本身所代表的意义、关节角度的变化、关节的最大角速度等。这些采用编辑距离比较过的富含信息量的序列就可以表示人体动作。在【13】中，使用成对有密切关系的关节运动轨迹对来表示骨架系列，然后采用线性SVM对其进行分类。

3. 特殊的欧几里得群SE(3)

在这部分当中，我们对特殊的欧几里得群SE(3)进行简短的介绍。我们建议读者去【4】

中寻找李氏群的一般介绍，在【11】中寻找更多关于SE(3)的细节和刚体运动学的知识。

用SE(3)表示的特殊的欧几里得群可以采用如下形式的 4×4 矩阵表示

$$P(R, \vec{d}) = \begin{bmatrix} R & \vec{d} \\ 0 & 1 \end{bmatrix}, \quad (1)$$

$\vec{d} \in R^3$ ，并且 $R \in R^{3 \times 3}$ 是旋转矩阵，SE(3)中的成员通过旋转和平移作用于点 $z \in R^3$ ：

$$\begin{bmatrix} R & \vec{d} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} z \\ 1 \end{bmatrix} = \begin{bmatrix} Rz + \vec{d} \\ 1 \end{bmatrix}. \quad (2)$$

从几何学的观点来看，这组元素通过一般意义上的矩阵乘法可以被平滑地过渡到6维流行面，从而具有李氏群【4】的结构。 4×4 矩阵 I_4 是SE(3)的一员并且是该群的单位元。

SE(3)在单位元 I_4 上的切平面被称为SE(3)

上的李代数，用 $\mathfrak{se}(3)$ 表示。它是一个由所有

的形如 $\begin{bmatrix} U & \vec{w} \\ 0 & 0 \end{bmatrix}$ 的 4×4 矩阵构成6维空间向量，

$\vec{w} \in R^3$ 并且 U 是 3×3 斜对称矩阵。对每一个元素

$$B = \begin{bmatrix} U & \vec{w} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -u_3 & u_2 & w_1 \\ u_3 & 0 & -u_1 & w_2 \\ -u_2 & u_1 & 0 & w_3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in \mathfrak{se}(3), \quad (3)$$

它的向量表示vec(B)如下

$$\text{vec}(B) = [u_1, u_2, u_3, w_1, w_2, w_3]. \quad (4)$$

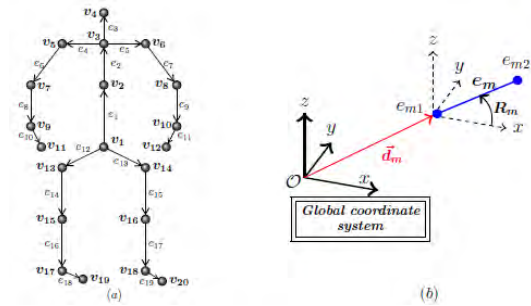


图 2: (a)一个由20个关节和19个身体部位组成的骨架示例, (b)在世界坐标系表示身体部位的 e_m 。

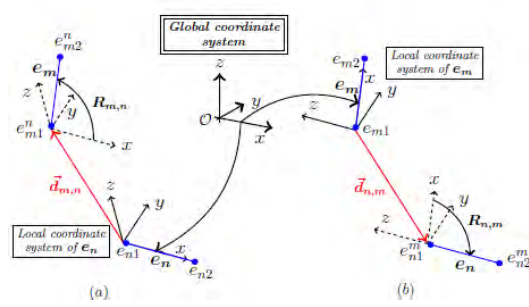


图 3: (a) 在世界坐标系表示身体部位的 e_m

到局部坐标系的 e_n 的转换, (b) 在局部坐标系表示身体部位的 e_n 到世界坐标系 e_m 的转换。

指数映射 $\exp_{SE(3)} : \mathfrak{se}(3) \rightarrow SE(3)$ 和对数映射 $\log_{SE(3)} : SE(3) \rightarrow \mathfrak{se}(3)$ 在李代数 $\mathfrak{se}(3)$ 和李氏群 $SE(3)$ 中的表示如下

$$\left. \begin{aligned} \exp_{SE(3)}(B) &= e^B \\ \log_{SE(3)}(P) &= \log(P) \end{aligned} \right\} (5)$$

e 和 \log 分别代表通常的指数的对数。由于 $\log(P)$ 并不是唯一的, 我们对该值使用最小标准。请参考【11】去实现 $SE(3)$ 非常高效的指数和对数之间的映射。

SE(3)的插值: $SE(3)$ 多种不同的插值方法已经被提出【25】。在本文中, 我们采用基于旋转运动的简单分段插值【26】。分别给定在时间序列 t_1, t_2, \dots, t_n 上的

$Q_1, Q_2, \dots, Q_n \in SE(3)$ 我们使用如下的插值曲线:

$$\gamma(t) = Q_i \exp_{SE(3)} \left(\frac{t-t_i}{t_{i+1}-t_i} B_i \right) \text{ 且 } t \in [t_i, t_{i+1}], (6)$$

$$B_i = \log_{SE(3)}(Q_i^{-1}Q_{i+1}) \text{ 且 } i = 1, 2, \dots, n-1.$$

SE(3) × ... × SE(3): 我们可以对单位元 (I_4, \dots, I_4) 和李代数 $\mathfrak{m} = \mathfrak{se}(3) \times \dots \times \mathfrak{se}(3)$ 结合多个 $SE(3)$ 采用直积 × 得到一个新的李氏群 $\mathcal{M} = SE(3) \times \dots \times SE(3)$ 。对

$(B_1, \dots, B_K) \in \mathfrak{m}$ 和 $(P_1, \dots, P_K) \in \mathcal{M}$, 它们之间指数和对数映射关系如下

$$\left. \begin{aligned} \exp_{\mathcal{M}}((B_1, \dots, B_K)) &= (e^{B_1}, \dots, e^{B_K}) \\ \log_{\mathcal{M}}((P_1, \dots, P_K)) &= (\log(P_1), \dots, \log(P_K)) \end{aligned} \right\} (7)$$

对 $SE(3) \times \dots \times SE(3)$ 的插值可以通过分别对 $SE(3)$ 同时插值实现。

4. 我们提出的骨架表示方法

设 $S=(V,E)$ 表示一个骨架, $V = \{v_1, \dots, v_N\}$

代表关节集合, $E = \{e_1, \dots, e_M\}$ 代表面向刚性身体部位的集合。图 2 展示了一个由20个关节和19个身体部位组成的骨架示例。设 $e_{n1} \in R^3$,

$e_{n2} \in R^3$ 分别表示身体部位 e_n 的开始和结束点。

给出一对身体部位坐标 e_m 和 e_n , 为了描述他们之间的相对几何关系, 我们在用局部坐标系表示每一个点时, 同时附上其他点的信息。图 3 解释了这种做法。通过旋转 (最小旋转) 和平移世界坐标系使 e_{n1} 为原点, e_n 和 x 轴 (参照图 3(a)) 相重合就可以得身体部位在局部坐标系的 e_n 。设 $e_{m1}^n(t), e_{m2}^n(t) \in R^3$ 分别代表开始和结束点, e_m 则与在局部坐标系中和时间序列 t 所在时刻的 e_n 相对应。从而

$$\begin{bmatrix} e_{m1}^n(t) & e_{m2}^n(t) \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} R_{m,n}(t) & \vec{d}_{m,n}(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & l_m \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad (8)$$

$R_{m,n}(t)$ 和 $\vec{d}_{m,n}(t)$ 是旋转和平移变化(在局部坐标系和 e_n 相一致), 二者都需要 e_n 提供坐标, e_m 提供方向。

类似的, 在局部坐标系我们可以用 e_n 结合 e_m (参考图 3(b))。设 $e_{n1}^m(t), e_{n2}^m(t) \in \mathcal{R}^3$ 分别代表开始和结束点, e_n 则与在局部坐标系中和时间序列 t 所在时刻的 e_m 相对应。从而

$$\begin{bmatrix} e_{n1}^m(t) & e_{n2}^m(t) \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} R_{n,m}(t) & \vec{d}_{n,m}(t) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & l_m \\ 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \quad (9)$$

$R_{n,m}(t)$ 和 $\vec{d}_{n,m}(t)$ 是旋转和平移变化(在局部坐标系和 e_n 相一致), 二者都需要 e_m 提供坐标, e_n 提供方向。

由于身体部位的长度不会随着时间而变化, 所以在时间序列 t 上 e_m 和 e_n 之间的关系可以用如下关系描述

$$\left. \begin{aligned} P_{m,n}(t) &= \begin{bmatrix} R_{m,n}(t) & \vec{d}_{m,n}(t) \\ 0 & 1 \end{bmatrix} \in SE(3) \\ P_{n,m}(t) &= \begin{bmatrix} R_{n,m}(t) & \vec{d}_{n,m}(t) \\ 0 & 1 \end{bmatrix} \in SE(3) \end{aligned} \right\}, \quad (10)$$

考虑到 e_n 是绕 e_m 某个坐标轴旋转得到的, 乍看上去似乎仅使用 $P_{m,n}(t)$ 或者 $P_{n,m}(t)$ 就足够了。虽然二者之间存在着某种关系, 但是 $P_{m,n}(t)$ 不会被改变。同样, 如果 e_m 是绕 e_n

某个坐标轴旋转得到的, $P_{n,m}(t)$ 也不会被改变。所以, 如果我们仅仅使用 $P_{m,n}(t)$ 或者

$P_{n,m}(t)$ 去表示相对几何关系, 在一些特定的相对运动下, 这种表示得到的结果将保持不变, 这是我们不愿意看到的。因此, 我们使用

$P_{m,n}(t)$ 和 $P_{n,m}(t)$ 去代表 e_m 和 e_n 之间的相对

几何关系。值得注意的是当仅仅对 e_m 和 e_n 作相同的旋转和平移, 也就是说二者之间没有相对位移时, $P_{m,n}(t)$ 和 $P_{n,m}(t)$ 都不会改变。

使用所有成对身体部位间的相对几何关系, 在时间序列 t 上我们使用

$$C(t) = (P_{1,2}(t), P_{2,1}(t), \dots, P_{M-1,M}(t), P_{M,M-1}(t)) \in$$

$SE(3) \times \dots \times SE(3)$, M 表示身体部位的数量。

使用我们提出的骨架模型, 一个骨架序列就描述了一个动作, 这个动作可以用

$SE(3) \times \dots \times SE(3)$ 中的曲线 $\{C(t), t \in [0, T]\}$ 代表 (图 1)。

在 $SE(3) \times \dots \times SE(3)$ 曲线空间中对动作曲线分类是一项十分重要的工作。此外, 像 SVM 这种标准的分类方法和像傅里叶分析

这种时变方法都无法直接适用于这个空间。为了克服这些难题, 我们将动作曲线从

$SE(3) \times \dots \times SE(3)$ 映射到 $se(3) \times \dots \times se(3)$, 对单位元来说, 李代数所对应的是切向空间。与李代数 (向量表示法) 相一致的 $C(t)$ 表示法如下

下

$$\mathcal{C}(t) = [\text{vec}(\log(P_{1,2}(t))), \text{vec}(\log(P_{2,1}(t))), \dots, \text{vec}(\log(P_{M-1,M}(t))), \text{vec}(\log(P_{M,M-1}(t)))]. \quad (11)$$

在任何取样时刻 t , $\mathcal{C}(t)$ 是一个 $6M(M-1)$ 维向量。因此, 我们用随着时间变化的 $6M(M-1)$ 维向量来表示动作。

请注意, 在我们的骨架模型中我们仅使用

$P_{m,n}(t)$ 的相对测量值。在完成实验时我们在

骨架表示法中还添加了身体部位的绝对三维

坐标。刚体段 e_m 的三维坐标可以通过它所对

应的关于 x 轴的旋转量 R_m 和始点关于原点

(参照图 2(b)) 平移量 \vec{d}_m 来描述。但是，采用身体部位的绝对坐标并不能得到多大的改进，这暗示我们身体部位的绝对坐标所提供的信息在我们的实验中对动作定位来说是多余的。因此，在本文中我们仅使用相对测量量。

5. 时变模型和分类

由于诸如比率不匹配、时间错位和噪声等问题对李代数中的曲线进行分类并不是那么简单的。根据【17】，我们使用DTW【10】去控制比率的变化。在训练过程中，对每一个动作分类，我们使用表 1 中的算法计算出了一个虚曲线，并且使用DTW将所有经过训练的曲线变形的这条虚曲线上。对DWT我们使用李代数中欧氏距离的平方。注意到为了算出虚曲线，所有的曲线都应该有相同数量的样本。为了达到这个目的，我们使用了在第3部分提出的插值算法，并且在把他们映射到李代数之前对 $SE(3) \times \dots \times SE(3)$ 中的曲线进行重新取样。为了控制时间错位和噪声问题，我们使用新近提出的傅里叶时间金字塔【19】方法表示流行面而去除高频系数。我们对每一个分离出来的维采用FTP方法，并且将所有的傅里叶系数联系起来得到了最终的特征向量。使用一对多模式的线性SVM方法对最终的特征向量进行分类后，完成了对动作的识别。图 4 对完整的方法进行了回顾。

表 1: 计算形式曲线的算法

<p>输入: 曲线 $\mathcal{C}_1(t), \dots, \mathcal{C}_j(t), t = 0, 1, \dots, T$.</p> <p>迭代次数的最大值 max 和极限 δ .</p>
<p>输出: 虚曲线 $\mathcal{C}(t), t = 0, 1, \dots, T$.</p>

初始化: $\mathcal{C}(t) = \mathcal{C}_1(t), iter = 0$.

while $iter < max$

 使用DTW和欧式距离的平方的到一个曲面 $\mathcal{C}_j^w(t)$ ，将每一个曲面 $\mathcal{C}_j(t)$ 变换到虚曲面 $\mathcal{C}(t)$ 。

 计算一个新的虚曲面 $\mathcal{C}'(t)$ 使用

$$\mathcal{C}'(t) = \frac{1}{J} \sum_{j=1}^J \mathcal{C}_j^w(t).$$

if $\sum_{t=0}^T \|\mathcal{C}'(t) - \mathcal{C}(t)\|_2^2 \leq \delta (\|\cdot\|_2 \text{ 表示 } l_2 \text{ 范式})$

break

end

$\mathcal{C}(t) = \mathcal{C}'(t); iter = iter + 1;$

end

6. 实验评价

在这一部分中，我们使用三个不同的数据集: MSR-Action3D【7】，UTKinect-Action【20】和 Florence3D-Action【14】对提出的骨架模型进行评价。我们使用的数据可以从下面的网站进行下载

<http://ravitejav.weebly.com/kbac.html>.

MSR-Action3D数据集【7】: 这个数据集使用类似于Kinect的深度传感器获得。它由10个不同目标的20个动作组成。每个目标通过2到3次完成这个动作。总而言之，我们有557个动作序列。在数据集中我们提供了20个关节的三维位置。因为大多数动作彼此之间高度相似，所以这是一个富有挑战的数据集。

UTKinect-Action数据集【20】: 使用固定的Kinect传感器得到了这个数据集。它由10

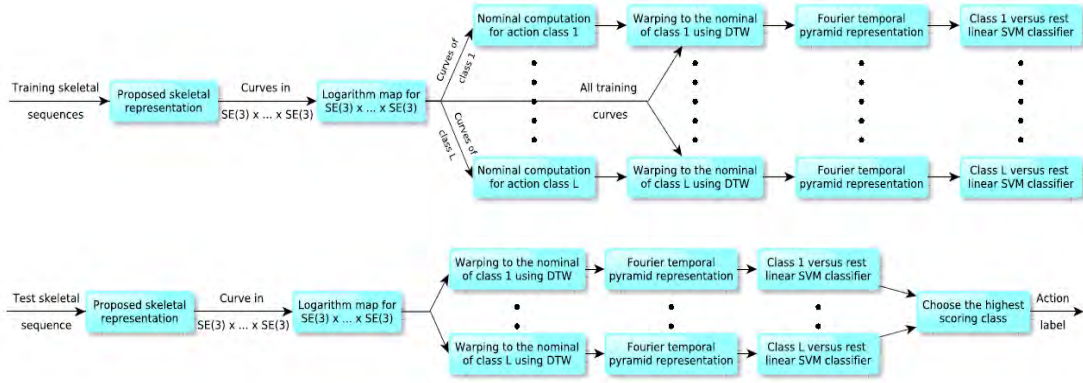


图 4：上层展示了训练过程中的步骤，下层展示了测试包含的所有步骤

个不同目标的10个动作组成。每个目标通过2次完成这个动作。总而言之，我们有199个动作序列。在数据集中我们提供了20个关节的三维位置。因为视点的不同和高度类内无关性，所以这是一个充满挑战的数据集。

Florence3D-Action数据集【14】：使用固定的Kinect传感器得到了这个数据集。它由10个不同目标的9个动作组成。每个目标通过2到3次完成这个动作。总而言之，我们有215个动作序列。在数据集中我们提供了15个关节的三维位置。因为高度类内无关性（在一些序列中使用左手而在另一些序列中使用右手完成动作）和一些动作姿态太过相似（比如从瓶中喝水和打电话的动作）的缘故，所以这是一个充满挑战的数据集。

基本预处理：在MSR-Action3D和UTKinect-Action数据集中，每一个骨架都由19个部位和20个关节组成，然而在Florence3D-Action数据集中，每一个骨架都由14个部位和15个关节组成。为了使骨架数据在绝对位置保持不变，通过将人体的臀部中心放到起始点，从而把三维骨架坐标从世界坐标系转换到了以人为中心的坐标系。对每一个数据集我们使用其中的一个作为参考，然后将其他所有的骨架标准化（不改变它们之间的夹角），从而使这些部位的长度和参考骨架的长度相一致。标准化工作使得骨架大小恒定。我们同时通过旋转骨架使得左髌骨到右髌骨地面投影向量和总体 x 轴相平行。旋转使得骨架与视点无关。

6.1. 其他的骨架表示法

为了展示出我们提出的骨架表示法的效果，我们对其和以下四种可供选择的方法进行了对比。

基于关节位置（JP）：连接所有关节的三维坐标 v_1, \dots, v_N 。

成对关节的相对位置（RJP）：连接所有的向量 $\overrightarrow{v_i v_j}, 1 \leq i < j \leq N$ 。

关节角度（JA）：连接所有关节夹角的一致四元数。我们曾尝试用欧拉角和欧拉角轴代表关节夹角，但是四元数法给出了最好的结果。

独立身体部位坐标（BPL）：使用关于总体 x 轴的旋转和平移，每一个独立的身体部位被表示为 $SE(3)$ 中的点。将点 $SE(3)$ 从映射到 $se(3)$ ，我们得到了一个6M维向量的表示方法，M在此处表示身体部位的数量。

为了公平起见，我们在第5部分中对所有的表示方法使用时变模型和和分类法进行描述。

6.2. 评价参数和设置

对于MSR-Action3D数据集，我们按照【7】中交叉学科测试里的设置，一半数据用来训练，另一半用来测试。按照【7】中的方法，我们将数据集划分成 AS_1, AS_2, AS_3 三个子集，每个子集由8个动作组成并且独立完成识别工作。

子集 AS_1 和子集 AS_2 对相似动作进行分组，

子集 AS_3 对复杂动作进行聚类。

对于UTKinect-Action和Florence3D-Action数据集，我们按照【27】中交叉学科测试里的设置，一半数据用来训练，另一半用来测试。

在所有的实验中，我们使用每个刚体段的1/4长度和三级傅里叶时空金字塔方法所得结果作为低频系数。SVM的参数C值在所有的实验中都设为1。正如在第5部分中所解释的那样，对所有的数据集， $SE(3) \times \dots \times SE(3)$ 中的曲线都被重新进行采样以得到相同长度。在重新采样前，选择数据集中所有采样曲线的最大值作为参考长度。本文中提到的所有结果，对训练和测试数据都做了超过十种不同组合的平均。

6.3. 结果

和其他骨架表示法的比较：

表 2汇报了MSR-Action3D数据集中不同骨架表示法的识别率。最后一行是三个子集

AS_1, AS_2, AS_3 的平均识别率。可以清楚地看到

到我们提出的表示方法在所有的子数据集上都要比其他一些常用的表示方法表现得好。我们所提方法的平均精度要比基于关节夹角的高10.6%，比独立身体部位位置法高8.9%，比关节位置法高5.2%，比相对关节位置法高4.2%。

在 AS_1 和 AS_2 子集上的优良表现表明我们提出的方法比其他区分相似动作的方法都要好。在 AS_3 子集上的优良表现表明我们提出的方法比其他对复杂动作进行建模的方法都要好。

表 3汇报了UTKinect-Action和Florence3D-Action数据集中不同骨架表示法的识别率。在UTKinect-Action数据集中，我们所提方法的平均精度要比基于关节夹角的高3%，比独立身体部位位置法高2.5%，比关

表 2：在MSR-Action3D数据集中使用【7】中约定得到的不同骨架表示法的识别率

数据集	JP	RJP	JA	BPL	所提法
AS_1	91.65	92.15	85.80	83.87	95.29
AS_2	75.36	79.24	65.47	75.23	83.87
AS_3	94.64	93.31	94.22	91.54	98.22
平均	87.22	88.23	81.83	83.54	92.46

表 3：在UTKinect-Action和Florence3D-Action数据集中得到的不同骨架表示法的识别率

数据集	JP	RJP	JA	BPL	所提法
UTKinect	94.68	95.58	94.07	94.57	97.08
Florence3D	85.26	85.20	81.36	80.80	90.88

表 4：与最先进方法所得结果的比较

MSR-Action3D数据集 (采用【7】中所用约定)	
三维关节直方图【20】	78.97
本征关节【22】	82.30
关节夹角相似性【13】	83.53
时空部位集合【18】	90.22
协方差描述符【5】	90.53
随机森林【27】	90.90
我们的方法	92.46
MSR-Action3D数据集(采用【19】中所用约定)	
动作【19】	88.20
我们的方法	89.48
UTKinect-Action 数据集	
三维关节直方图【20】	90.92
随机森林【27】	87.90
我们的方法	97.08
Florence3D-Action 数据集	
多部位姿势集【14】	82.00
我们的方法	90.88

节位置法高2.4%，比相对关节位置法高1.5%。在Florence3D-Action数据集中，我们所提方法的平均精度要比基于关节夹角的高9.5%，比独立身体部位位置法高10.1%，比关节位置法高5.6%，比相对关节位置法高5.7%。这些结果清楚地证明了我们所提方法对其他已有方法的优越性。

图 5展示了 MSR-Action3D AS_1 ，

MSR-Action3D AS_2 ，Florence3D Action数据集的混淆矩阵。我们跳过MSR-Action3D

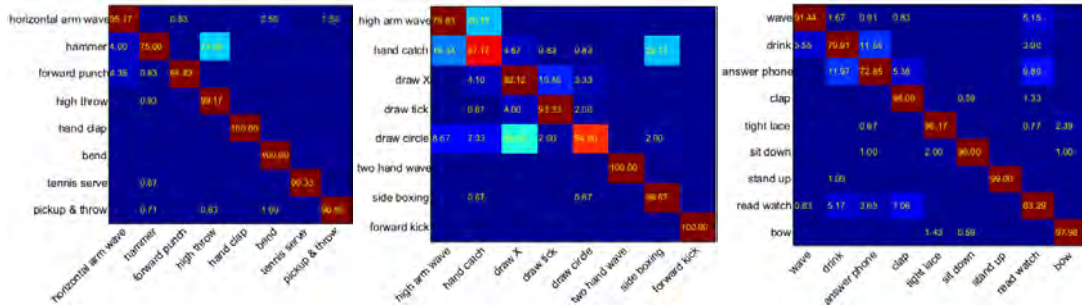


图 5: 混淆矩阵: 左—MSR-Action3D AS_1 ; 中—MSR-Action3D AS_2 ; 右—Florence3D-Action

AS_3 和UTKinect-Action数据集, 因为二者的一致识别率非常高。我们可以看到大部分混乱产生在高度相似的动作之间, 比如在

MSR-Action3D AS_1 子集中**捶打**和**高抛**动作;

在MSR-Action3D AS_2 子集中**画×**、**画√**和**画**

圈, **手抓**和**拳击**动作; 在数据集

Florence3D-Action中**喝水**、**打电话**和**看时间**动作。

和最先进方法结果的比较:

表 4对我们提出的方法和和目前最先进的基于骨架的行为检测方法进行了比较。可以看出在所有的数据集中我们的方法都得到了最优的结果。尤其是在UTKinect-Action数据集胜过目前最优秀的方法6.1%, 在Florence3D-Action数据集胜过目前最优秀的方法8.8%。

请注意, 在MSR-Action3D数据集上, 我们对所提方法公布了两个不同的识别比率。92.46%的识别率采用的是【7】中设置的相关参数, 89.48%的识别率采用的是【19】中设置的参数。在【19】中, 基于动作的方法采用的是包含20个动作的完整数据集, 而不是将数据集分成三个子集。这个实验设置比【7】中的还要困难。

一些最新的方法【13, 27】结合基于深度的骨架特性, 在MSR-Action3D数据集上, 其识别率可以达到94.5%左右。由于本文关注的不是多特征识别, 所以我们仅仅和【13, 27】中发布的基于骨架的数据进行比较。

很有趣地注意到, 在UTKinect-Action和Florence3D-Action数据集上基于关节位置(当使用时变模型和在第5部分中提出的分类方法

时)和基于相对关节位置方法产生的结果都要比目前已发表的最先进的方法得到的结果要好。这表明对于基于骨架的动作分类, DTW、FTP和线性SVM同样是适用的。

7. 结论和进一步的工作

在本文中, 通过使用旋转和平移建立不同身体部位间的三维几何关系, 我们提出了用李氏群 $SE(3) \times \dots \times SE(3)$ 中的点表示人体骨的方法。使用我们提出的方法, 采用李氏群中的曲线对人体动作建模。由于 $SE(3) \times \dots \times SE(3)$ 是流行面, 我们将所有的曲线映射到相对应的李代数中, 这些曲线是一个向量空间, 并且在李代数中完成即时建模和分类工作。我们的实验展示了我们提出的方法在三个不同的动作数据集上要优于很多已经提出的骨架表示法要优秀。我们同时展示了我们的方法胜过了各种目前最优秀的基于骨架的人体动作识别方法。

在我们的工作中, 我们使用所有身体部位对间的相对集几何关系。但是, 每个动作通常都用一些互相关联的特殊身体部位来描述。因此, 我们计划找到不同的策略去自动识别出和静止不同的身体部位序列。在本文中, 我们仅关注单个人的动作。我们计划发展这种表示方法到多目标协同作用的识别。

致谢: 这项研究得到了美国海军研究办公室支持的多学科研究创新计划的1141221258513号授权。

参考文献

- [1] J. K. Aggarwal and M. S. Ryoo. Human Activity Analysis: A Review. *ACM Computing Surveys*, 43(3):16:1–16:43, 2011. 1, 2

- [2] R. Chaudhry, F. Ofli, G. Kurillo, R. Bajcsy, and R. Vidal. Bio-inspired Dynamic 3D Discriminative Skeletal Features for Human Action Recognition. In *CVPRW*, 2013. 3
- [3] D. M. Gavrila and L. S. Davis. Towards 3-D Model-based Tracking and Recognition of Human Movement: A Multi-view Approach. In *International Workshop on Automatic Face and Gesture Recognition*, 1995. 3
- [4] B. Hall. Lie Groups, Lie Algebras, and Representations: An Elementary Introduction. *Springer*, 2003. 3
- [5] M. Hussein, M. Torki, M. Gowayed, and M. El-Saban. Human Action Recognition Using a Temporal Hierarchy of Co-variance Descriptors on 3D Joint Locations. In *IJCAI*, 2013. 2, 7
- [6] G. Johansson. Visual Perception of Biological Motion and a Model for its Analysis. *Perception & Psychophysics*, 14(2):201–211, 1973. 2
- [7] W. Li, Z. Zhang, and Z. Liu. Action Recognition Based on a Bag of 3D Points. In *CVPRW*, 2010. 2, 5, 6, 7
- [8] F. Lv and R. Nevatia. Recognition and Segmentation of 3D Human Action Using HMM and Multi-class Adaboost. In *ECCV*, 2006. 2
- [9] T. B. Moeslund, A. Hilton, and V. Krüger. A Survey of Advances in Vision-based Human Motion Capture and Analysis. *CVIU*, 104(2-3):90–126, 2006. 1
- [10] M. Müller. Information Retrieval for Music and Motion. *Springer-Verlag New York, Inc.*, 2007. 2, 5
- [11] R. M. Murray, Z. Li, and S. S. Sastry. A Mathematical Introduction to Robotic Manipulation. *CRC press*, 1994. 2, 3
- [12] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy. Sequence of the Most Informative Joints (SMIJ): A New Representation for Human Skeletal Action Recognition. In *CVPRW*, 2012. 2, 3
- [13] E. Ohn-bar and M. M. Trivedi. Joint Angles Similarities and HOG² for Action Recognition. In *In CVPRW*, 2013. 2, 3, 7
- [14] L. Seidenari, V. Varano, S. Berretti, A. D. Bimbo, and P. Pala. Recognizing Actions from Depth Cameras as Weakly Aligned Multi-part Bag-of-Poses. In *CVPRW*, 2013. 2, 5, 7
- [15] Y. Sheikh, M. Sheikh, and M. Shah. Exploring the Space of a Human Action. In *ICCV*, 2005. 2
- [16] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time Human Pose Recognition in Parts From a Single Depth Image. In *CVPR*, 2011. 1
- [17] A. Veeraraghavan, A. Srivastava, A. K. Roy-Chowdhury, and R. Chellappa. Rate-invariant Recognition of Humans and Their Activities. *IEEE Trans. on Image Processing*, 18(6):1326–1339, 2009. 5
- [18] C. Wang, Y. Wang, and A. L. Yuille. An Approach to Pose-based Action Recognition. In *CVPR*, 2013. 7
- [19] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining Actionlet Ensemble for Action Recognition with Depth Cameras. In *CVPR*, 2012. 2, 5, 7
- [20] L. Xia, C. C. Chen, and J. K. Aggarwal. View Invariant Human Action Recognition Using Histograms of 3D Joints. In *CVPRW*, 2012. 2, 3, 5, 7
- [21] Y. Yacoob and M. J. Black. Parameterized Modeling and Recognition of Activities. In *ICCV*, 1998. 2, 3
- [22] X. Yang and Y. Tian. EigenJoints-based Action Recognition Using Naïve-Bayes-Nearest-Neighbor. In *CVPRW*, 2012. 2, 3, 7
- [23] M. Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall. A Survey on Human Motion Analysis from Depth Data. In *CVPR Tutorial on RGBD Cameras*, 2013. 2
- [24] V. M. Zatsiorsky. Kinematics of Human Motion. *Human Kinetics Publishers*, 1997. 1
- [25] M. Zefran and V. Kumar. Two Methods for Interpolating Rigid Body Motions. In *ICRA*,

1998. 3

- [26] M. Zefran, V. Kumar, and C. Croke. Choice of Riemannian Metrics for Rigid Body Kinematics. In *ASME Design Engineering Technical Conference and Computers in Engineering Conference*, 1996. 3
- [27] Y. Zhu, W. Chen, and G. Guo. Fusing Spatiotemporal Features and Joints for 3D Action Recognition. In *CVPRW*, 2013. 3, 6, 7

