Pure Mathematics 理论数学, 2021, 11(6), 1166-1180
Published Online June 2021 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2021.116131

Hans汉斯

Karcher均值算法在动作识别上的应用

赵　倩[1]，叶　震[2]，周朝政[2]

[1]上海大学，上海
[2]上海电气集团股份有限公司中央研究院，上海
Email: qianzhao_v@163.com, zhye1985@aliyun.com, zhouchzh2018@aliyun.com

收稿日期：2021年5月12日；录用日期：2021年6月14日；发布日期：2021年6月21日

摘　要

本文提出Karcher均值算法与测地距离相结合的平均轨迹计算方法，并将其应用到动作识别任务中。具体地说，本文通过基于李群的骨架表示方法，将动作骨架序列表示为流形上的轨迹。为了解决轨迹的时间错位问题，本文将Karcher均值算法与李群上测地距离的定义相结合，计算所有动作轨迹的平均轨迹，然后采用传输平方根向量场表示方法，将所有动作轨迹与平均轨迹进行时间对齐。此外，本文在特征提取阶段，提出对特征进行加权融合，实验结果验证了融合特征的有效性。

关键词

动作识别，Karcher均值算法，骨架表示，轨迹时间对齐

The Application of Karcher Mean Algorithm in Action Recognition

Qian Zhao[1], Zhen Ye[2], Chaozheng Zhou[2]

[1]Shanghai University, Shanghai
[2]Shanghai Electric Central Research Institute, Shanghai
Email: qianzhao_v@163.com, zhye1985@aliyun.com, zhouchzh2018@aliyun.com

Received: May 12th, 2021; accepted: Jun. 14th, 2021; published: Jun. 21st, 2021

Abstract

In this paper, we propose an average trajectory calculation method based on the combination of Karcher mean algorithm and geodesic distance, and apply it to the action recognition task. Specifically, the skeletal sequences of actions are represented as trajectories on manifolds by a skeleton

文章引用: 赵倩, 叶震, 周朝政. Karcher 均值算法在动作识别上的应用[J]. 理论数学, 2021, 11(6): 1166-1180.
DOI: 10.12677/pm.2021.116131

representation method based on Lie groups. In order to solve the temporal misalignment problem of trajectories, this paper combines the Karcher mean algorithm with the definition of geodesic distance on Lie group, calculates the average trajectories of all action trajectories, and then uses the Transported Square-Root Vector Field representation method to align all action trajectories with the average trajectories in time. In addition, the weighted fusion of features is proposed in the feature extraction stage, and the experimental results verify the effectiveness of the fusion features.

Keywords

Action Recognition, Karcher Mean Algorithm, Skeleton Representation, Trajectory Temporal Misalignment

1. 引言

动作识别算法作为计算机视觉领域热门的研究课题之一，被广泛地应用于人机交互[1] [2]、监控安全[3] [4]以及医疗健康[5]等领域。由于 RGB 图像数据存在遮挡、光线变化以及视角变换等挑战，基于骨架的动作识别算法为人们打开了新的思路。

骨架数据的一个重要优势在于它显式地记录了关节点的坐标信息，避免背景信息的干扰。但挑战也是显而易见的，动作的执行速率不同，导致采样的骨架序列存在时间错位的问题。尤其是，当以不同的演变速率执行相同的动作时，数据对应的采样帧并不完全相似。解决这个问题的常用方法是找到一个合适的模板动作，将所有动作序列样本与模板对齐。形状分析都在对齐后的骨架序列上进行，从而避免由时间错位引起的度量扭曲。

在动作识别的黎曼框架下，骨架序列被表示成流形上的轨迹，而流形上模板轨迹的计算方法，目前大致分为两种。第一种是在轨迹的切空间寻找模板。Vemulapalli 等人[1]将动作骨架序列表示为李群上的轨迹，随后将轨迹上的点映射到李代数上。由于李代数为向量空间，因此，作者首先随机选择一条轨迹的李代数来初始化模板，并将其他轨迹的李代数与之对齐。然后直接求取对齐后的所有轨迹李代数的线性平均，将该平均向量作为新的对齐模板。在切空间寻找对齐模板，方法简单，计算高效，但是一旦将轨迹映射到了切空间，它就不再具有轨迹的非欧结构。这在对齐轨迹时，可能会损失一些结构信息。第二种常用方法就是基于 Karcher 均值的概念，计算所有样本轨迹的平均轨迹，并将该平均轨迹作为模板轨迹。平均轨迹包含所有样本轨迹的统计信息，是合适的轨迹对齐模板。Su 等人[7]提出将传输平方根向量场(Transported Square-Root Vector Field, TSRVF)与 Karcher 均值概念相结合的平均轨迹计算方法，将平均轨迹定义为一条积分曲线，积分起点固定，沿着平均传输平方根向量场方向进行积分。Anirudh 等人[8]将 Su 等人[7]提出的平均轨迹计算方法与基于李群的骨架表示方法进行结合，用于解决李群上轨迹的时间错位问题。Amor 等人[9]将骨架序列表示成 Kendall 形状空间上的轨迹，然后结合轨迹在形状空间上的测地距离以及 Karcher 均值的概念来计算平均轨迹。

本文基于 Karcher 均值算法的概念，提出将 Karcher 均值算法与李群上测地距离相结合的平均轨迹计算方法。该方法不同于上述工作中的平均轨迹计算方法，其基本思想是平均轨迹到所有样本轨迹的测地

距离平方和最小。根据测地距离，将平均轨迹逐步向样本轨迹中心平行移动，因此，计算所得的平均轨迹更靠近样本轨迹中心。本文将该方法应用于动作识别中，构建一个新的动作识别框架。本文首先采用基于李群的骨架表示方法[6]，对骨架上身体部位之间的相对几何位置(旋转和平移)进行编码。随后，本文运用提出的平均轨迹计算方法，计算所有样本轨迹的平均轨迹，并利用传输平方根向量场表示方法[7]将所有样本轨迹与平均轨迹对齐。在特征提取阶段，本文提取两种分类特征，即平均轨迹到样本轨迹的射击向量以及样本轨迹的李代数。为了解决数据噪声问题，我们采用傅里叶时间金字塔表示方法[10]来处理两类特征，滤掉高频系数，并提出对两类特征的傅里叶系数进行加权融合，以获得更优的动作识别结果。最后，我们训练一个 one-vs-all 线性支持向量机分类器来实现动作分类。本文的主要贡献可以总结为以下两点：

1) 提出了 Karcher 均值与测地距离相结合的平均轨迹计算方法，并将其应用于动作识别任务，构建出一个新的动作识别框架。

2) 提出对两种动作特征的傅里叶系数进行加权融合。本文实验结果验证了融合特征的有效性。

2. 预备知识

2.1. 特殊欧氏群 $SE(3)$的基础知识介绍

特殊欧氏群 $SE(3)$是 4×4 矩阵的集合，矩阵形式为

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

其中 $R\in\mathbb{R}^{3\times 3}$ 表示旋转矩阵，它是特殊正交群 $SO(3)$上的元素。$\vec{d}\in\mathbb{R}^3$ 表示平移向量。从几何的角度看，特殊欧氏群 $SE(3)$中的元素形成了一个弯曲的 6 维流形，并且具有李群结构[11]。恒等矩阵 I_4 是 $SE(3)$中的元素，同时也是该李群的单位元。

$SE(3)$在单位元 I_4 处的切平面称为 $SE(3)$的李代数，记为 $\mathfrak{se}(3)$。对任意元素

$$\hat{\xi}=\begin{bmatrix} \hat{\omega} & \vec{v} \\ \mathbf{0} & 0 \end{bmatrix}=\begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & v_2 \\ -\omega_2 & \omega_1 & 0 & v_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}\in\mathfrak{se}(3).$$

其中 $\vec{v}\in\mathbb{R}^3$，$\hat{\omega}$ 是 3×3 的反对称矩阵。它的等价向量表示为 $\xi=\left[\omega^{\mathrm{T}},\vec{v}^{\mathrm{T}}\right]^{\mathrm{T}}=\left[\omega_1,\omega_2,\omega_3,v_1,v_2,v_3\right]^{\mathrm{T}}\in\mathbb{R}^6$，是一个 6 维的向量空间。

对于 $SE(3)$上的指数映射和对数映射，本文沿用[12]中给出的定义。对任意 $\hat{\xi}\in\mathfrak{se}(3)$，当 $\|\omega\|=0$ 时，指数映射定义为

$$\exp\hat{\xi}=\begin{bmatrix} \mathbf{1} & \vec{v} \\ \mathbf{0} & 1 \end{bmatrix}, \tag{1}$$

当 $\|\omega\|\neq 0$ 时，指数映射定义为

$$\exp\hat{\xi}=\begin{bmatrix} e^{\hat{\omega}} & A\vec{v} \\ \mathbf{0} & 1 \end{bmatrix}. \tag{2}$$

其中 $e^{\hat{\omega}}$ 和 A 可由 Rodrigue 公式显式给出，

$$e^{\hat{\omega}}=I+\frac{\hat{\omega}}{\|\omega\|}\sin\|\omega\|+\frac{\hat{\omega}^2}{\|\omega\|^2}\left(1-\cos\|\omega\|\right),$$

$$A = I + \frac{\hat{\omega}}{\|\omega\|^2}\left(1-\cos\|\omega\|\right) + \frac{\hat{\omega}^2}{\|\omega\|^3}\left(\|\omega\| - \sin\|\omega\|\right).$$

对任意 $P \in SE(3)$，对数映射的定义为

$$\hat{\xi} = \log\begin{bmatrix} R & \vec{d} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \hat{\omega} & A^{-1}\vec{d} \\ \mathbf{0} & 0 \end{bmatrix}, \tag{3}$$

其中 $\hat{\omega} = \log R$，当 $\|\omega\| \neq 0$ 时，

$$A^{-1} = I - \frac{1}{2}\hat{\omega} + \frac{2\sin\|\omega\| - \|\omega\|\left(1+\cos\|\omega\|\right)}{2\|\omega\|^2 \sin\|\omega\|}\hat{\omega}^2,$$

当 $\|\omega\| = 0$ 时，$A = I$。

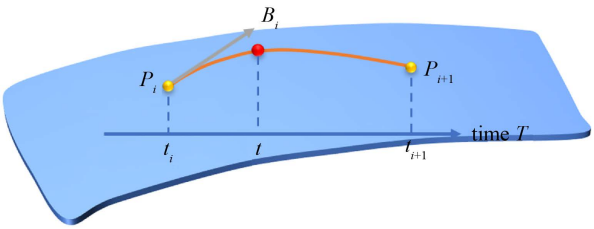


Figure 1. Pairwise interpolation method. The red point is the point to be interpolated, and the orange solid line is the Lie group *SE*(3) geodesic line
图 1. 成对插值法。红点为待插值点，橘色实线为李群 SE(3) 测地线

对 *SE*(3)上的数据进行插值有多种方法[13]。本文主要采用基于螺旋运动的成对插值法[14]。如图 1 所示，给定 $P(t_1), P(t_2), \cdots, P(t_n) \in SE(3)$，分别对应时刻 $t_1, t_2, \cdots, t_n$，然后定义以下曲线进行插值：

$$\beta(t) = P(t_i)\exp_{SE(3)}\left(\frac{t-t_i}{t_{i+1}-t_i}B_i\right),$$

其中 $t \in [t_i, t_{i+1}]$，$B_i = \log_{SE(3)}\left(P(t_i)^{-1}P(t_{i+1})\right)$，$i = 1, 2, \cdots, n-1$。指数映射 exp 和对数映射 log 通过(1)、(2)、(3)式进行估计。

2.2. 骨架表示的具体方法

令 $S = (L, E)$ 表示一副骨架，其中 $L = \{l_1, l_2, \cdots, l_N\}$ 表示关节点集合，$E = \{e_1, e_2, \cdots, e_M\}$ 表示带有方向的身体骨骼位置集合。如图 2 展示了一副包含 15 个关节点和 14 块身体骨骼的骨架。

给定一对身体骨骼 e_m 和 e_n，如图 3 所示，$e_{n1}, e_{n2} \in \mathbb{R}^3$ 分别表示身体骨骼的 e_n 起点和终点。为了表述它们之间的相对几何关系，我们首先对其中一块骨骼构建局部坐标系，然后描述将其移动到另一块骨骼位置处的旋转和平移。在图 3 中，骨骼 e_n 的局部坐标系是通过旋转一平移全局坐标系所得。令 e_n 为 x 轴，e_{n1} 是坐标原点，$e_{m1}^n(t), e_{m2}^n(t) \in \mathbb{R}^3$ 分别为时刻 t 时，骨骼 e_m 在 e_n 局部坐标系下的起始关节点。于是，骨骼 e_m 和 e_n 在时刻 t 下的相对几何可以描述为

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

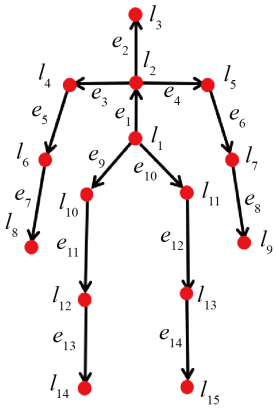


Figure 2. Schematic diagram of human skeleton
图 2. 人体骨架示意图

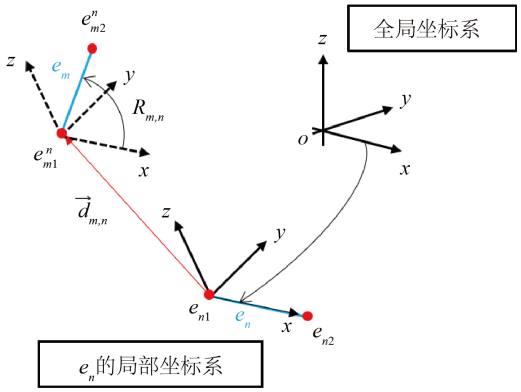


Figure 3. The representation of e_m in the local coordinate system of e_n
图 3. e_m 在 e_n 的局部坐标系下的表示

考虑到单个 $P_{m,n}(t)$ 在一些特定的相对运动下会保持不变，因此，本文同时采用 $P_{m,n}(t)$ 和 $P_{n,m}(t)$ 来表示骨骼 e_m 和 e_n 之间的相对几何。只有当骨骼间不存在相对运动时，$P_{m,n}(t)$ 和 $P_{n,m}(t)$ 才会同时保持不变[6]。

通过上述身体骨骼之间的相对几何，我们可以把时刻 t 时的一副完整骨架 S 表示为

$$\beta(t)=\left(P_{1,2}(t),P_{2,1}(t),\cdots,P_{M-1,M}(t),P_{M,M-1}(t)\right)\in SE(3)\times\cdots\times SE(3),$$

其中 M 为身体骨骼的数量。一副完整的骨架共有 $M(M-1)$ 个骨骼对，本文为了记号简洁，将骨骼对 $P_{m,n}(t)$ 下标 (m,n) 用骨骼对数目的序号表示，记为 $P_k(t)$，$k=1,\cdots,M(M-1)$，即完整的骨架表示又可以写为

$$\beta(t)=\left(P_1(t),P_2(t),\cdots,P_{M(M-1)}(t)\right)\in SE(3)\times\cdots\times SE(3).$$

依据这种骨架表示方法，一个完整的动作骨架序列就可以表示为 $SE(3)\times\cdots\times SE(3)$ 上的轨迹 $\{\beta(t),t\in[1,T]\}$.

3. Karcher 均值与测地距离相结合的平均轨迹计算方法

本文采用基于李群的骨架表示方法，将动作骨架序列表示为特殊欧氏群积空间 $SE(3)\times\cdots\times SE(3)$ 上

的轨迹，记为 $\beta(t)$，$t=1,\cdots,T$ 。T 为骨架序列的帧数。假设每副骨架包含 M 块身体骨骼，则每条轨迹的尺寸为 $\left[M(M-1),4\times4,T\right]$，其中 4×4 是相对几何的尺寸。

Karcher 均值这个概念首先由 Grove 和 Karcher [15]提出，之后被推广到黎曼流形上表示样本中心。本文提出 Karcher 均值与测地距离相结合的平均轨迹计算方法，其基本思想是样本轨迹中心(平均轨迹)到所有样本轨迹的测地距离平方和最小。给定 $SE(3)\times\cdots\times SE(3)$ 上 n 条样本轨迹 $\beta_1,\cdots,\beta_n$，根据 Karcher 均值的概念，平均轨迹 μ 的计算公式可以描述为

$$\mu=\underset{\beta\in SE(3)\times\cdots\times SE(3)}{\arg\min}\sum_{i=1}^{n}d(\beta_i,\beta)^2, \tag{4}$$

其中 $d(\beta_i,\beta)$ 为动作轨迹 β_i 和 β 之间的测地距离。如图 4 所示，红点代表流形上的轨迹，黄点表示潜在的轨迹中心。红色实线表示轨迹，黄色虚线表示两点之间的测地线。

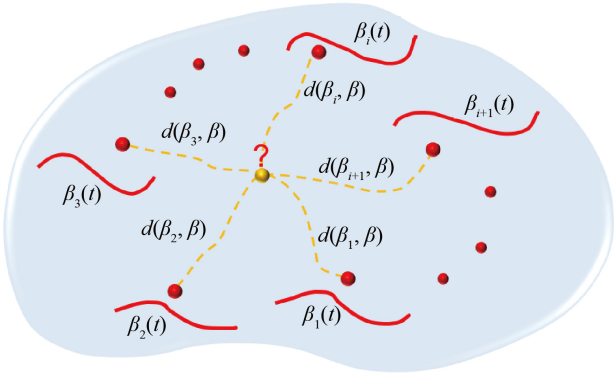


Figure 4. The average trajectory calculation method combining Karcher average value and geodetic distance
图 4. Karcher 均值与测地距离相结合的平均轨迹计算方法

令 $M'=M(M-1)$，那么每条动作轨迹应当包含 $M'\times T$ 个相对几何位置(旋转和平移)，每个相对几何位置记为 $P_k^i(t)\in SE(3)$，$i=1,\cdots,n$ 是样本轨迹的序号，$k=1,\cdots,M'$ 表示每副骨架上的相对几何位置的序号，$t=1,\cdots,T$ 为骨架帧数的序号。因此，动作轨迹之间的测地距离平方和可以写为两条动作轨迹上，所有对应相对几何位置之间的测地距离平方和，即

$$\mu=\underset{P_k^{\mu}\in SE(3)}{\arg\min}\sum_{i=1}^{n}\sum_{k=1}^{M'}\sum_{t=1}^{T}\delta\left(P_k^i(t),P_k^{\mu}(t)\right)^2, \tag{5}$$

其中 $\delta\left(P_k^i(t),P_k^{\mu}(t)\right)$ 表示的是 t 时刻，第 i 条动作轨迹的第 k 个相对几何到平均轨迹 μ 的第 k 个相对几何的测地距离。由于相对几何位置位于李群 $SE(3)$ 上，因此，距离 δ 定义在李群 $SE(3)$ 上。令 $P_k^i(t),P_k^j(t)\in SE(3)$，则第 i 条动作轨迹的第 k 个相对几何到第 j 条动作轨迹的第 k 个相对几何的测地距离可以描述为

$$\delta\left(P_k^i(t),P_k^j(t)\right)=\int_0^1\dot{\gamma}(s)\mathrm{d}s=\left\|v_{ij,k}(t)\right\|,$$

其中 $\mathrm{v}_{ij,k}(t)=\log_{P_k^i(t)}\left(P_k^j(t)\right)$。对数映射 log 的具体计算形式可以通过(3)式所得。

在求解目标函数(5)式的过程中，要求平均轨迹到所有动作轨迹的测地距离平方和最小，即要求轨迹上每个对应相对几何之间的测地距离最小。因此，(5)式的具体求解步骤可总结如下算法 1 所示。

算法 1 Karcher 均值与测地距离相结合的平均轨迹计算方法

输入： 样本轨迹 $\beta_0(t),\cdots,\beta_n(t)$， $t=1,\cdots,T$，每个轨迹点包含 M' 个相对几何位置 $\left\{P_k^i(t)\right\}_{k=1}^{M'}\in SE(3)$；超参数 $\epsilon>0$

输出： 平均轨迹 $\mu(t)$， $t=1,\cdots,T$

初始化平均轨迹 $\mu_0\leftarrow\beta_0$， $\tau\leftarrow 0$
for $t\leftarrow[1,\cdots,T]$ do
  for $k\leftarrow[1,\cdots,M']$ do
    repeat
      通过(3)式计算 $P_k^{\mu_\tau}(t)$ 和 $P_k^j(t)$ 之间的切向量 $v_{j,k}(t)$， $j=1,\cdots,n$
      计算 $P_k^{\mu_\tau}(t)$ 的平行移动方向 $\overline{v}_{j,k}(t)\leftarrow\frac{1}{n}\sum v_{j,k}(t)$
      通过(1)式和(2)式更新平均相对几何 $P_k^{\mu_{\tau+1}}(t)\leftarrow\exp_{P_k^{\mu_\tau}(t)}\left(\overline{v}_{j,k}(t)\right)$
      $\tau\leftarrow\tau+1$
    until $\left\|\overline{v}_{j,k}(t)\right\|<\epsilon$
  end for
  样本平均轨迹 $\mu(t)=\left(P_1^\mu(t),\cdots,P_{M'}^\mu(t)\right)$， $t=1,\cdots,T$
end for
return 平均轨迹 $\mu(t)$， $t=1,\cdots,T$

4. 平均轨迹计算方法在动作识别方向上的应用

4.1. 特征的提取和融合

由第二章可知，本文首先将骨架序列表示为李群 $SE(3)\times\cdots\times SE(3)$ 上的轨迹。为了解决轨迹时间错位问题，本文采用 Karcher 均值与测地距离相结合的平均轨迹计算方法，计算所有样本的平均轨迹，并将其作为轨迹时间对齐的模板。随后，本文采用 TSRVF 表示方法[7]，用于对齐平均轨迹和样本轨迹。然而，对齐后的轨迹仍位于流形空间，在这个空间上进行轨迹分类并不容易。为了解决这个问题，我们针对时间对齐后的轨迹提取两类动作特征：一类是平均轨迹到样本轨迹的射击向量(Shooting Vector，SV)，另一类是样本轨迹的李代数(Lie Algebra, LA)。射击向量可以理解为单位时间内，序列空间上一点到另一点的切向[8]。在本文中，轨迹时间对齐之后，我们在 $\tau=0$ 时刻，从平均轨迹 $\mu(t)$ 的某一点出发，在 $\tau=1$ 时刻，到达另一动作轨迹的对应点，计算两点之间的射击向量，如图 5 所示。具体计算过程如算法 2 所示。

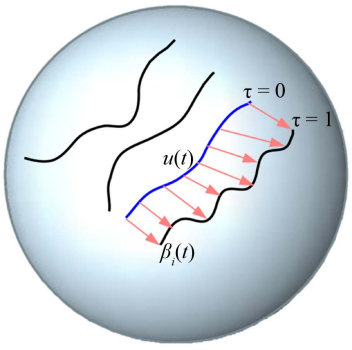


Figure 5. The schematic diagram of the characteristic shooting vector
图 5. 特征射击向量的示意图

算法 2 平均轨迹到样本轨迹的射击向量

输入：样本轨迹 $\beta_1(t),\cdots,\beta_n(t)$ 以及平均轨迹 $\mu(t)$， $t=1,\cdots,T$

输出：平均轨迹到样本轨迹的射击向量 V_S

1: 采用 TSRVF 表示方法[8]将样本轨迹与平均轨迹对齐，计算对齐后的样本轨迹 $\tilde{\beta}_1(t),\cdots,\tilde{\beta}_n(t)$

2: for $i\leftarrow[1,\cdots,n]$ do

3: for $t\leftarrow[1,\cdots,T]$ do

4: 通过(3)式计算射击向量 $v_S(i,t)\leftarrow\log_{\mu(t)}\left(\tilde{\beta}_i(t)\right)$

5: end for

6: 定义 $V_S(i)=\left[v_S(i,1)^{\mathrm{T}},\cdots,v_S(i,T)^{\mathrm{T}}\right]^{\mathrm{T}}$

7: end for

8: return 射击向量 V_S

由于轨迹位于李群 $SE(3)\times\cdots\times SE(3)$ 上，同时轨迹的李代数为向量空间，因此，本文提取轨迹的李代数进行动作表征。李代数的提取方法总结在算法 3 中

算法 3 样本轨迹的李代数

输入：样本轨迹 $\beta_1(t),\cdots,\beta_n(t)$ 以及平均轨迹 $\mu(t)$， $t=1,\cdots,T$

输出：平均轨迹到样本轨迹的射击向量 V_L

1: 采用 TSRVF 表示方法[8]将样本轨迹与平均轨迹对齐，计算对齐后的样本轨迹 $\tilde{\beta}_1(t),\cdots,\tilde{\beta}_n(t)$

2: for $i\leftarrow[1,\cdots,n]$ do

3: for $t\leftarrow[1,\cdots,T]$ do

4: 通过(3)式计算得到每条样本轨迹的李代数 $v_L(i,t)$

5: end for

6: 定义 $V_L(i)=\left[v_L(i,1)^{\mathrm{T}},\cdots,v_L(i,T)^{\mathrm{T}}\right]^{\mathrm{T}}$

7: end for

8: return 样本轨迹的李代数 V_L

为了克服数据噪声问题，获得更加有效、鲁棒的动作表征，本文采用傅里叶时间金字塔方法[10]处理两类特征，分别获得射击向量的傅里叶时间金字塔系数 V_{SC} 和李代数的傅里叶时间金字塔系数 V_{LC}。此外，为了获得更优的动作识别结果，本文提出对两类特征的傅里叶时间金字塔系数进行加权融合，融合特征 $V_{FC}=\alpha V_{SC}+(1-\alpha)V_{LC}$，其中 α 为特征的权重系数。融合特征既包含轨迹之间的距离信息，也包含单个样本轨迹的几何信息。

4.2. 动作识别框架

本文将 Karcher 均值和测地距离相结合的平均轨迹计算方法应用于动作识别中，构建一个新的动作识别框架，如图 6 所示。

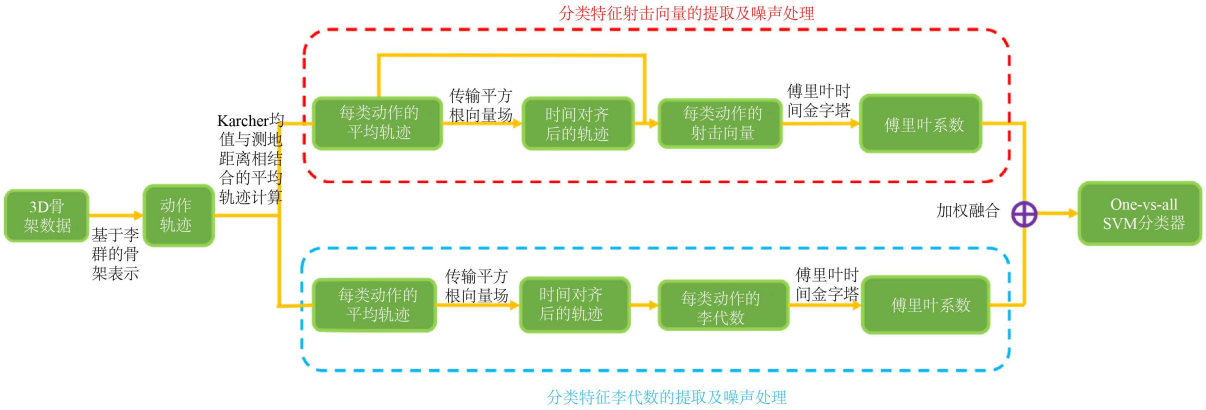


Figure 6. The action recognition framework of this article
图 6. 本文的动作识别框架

第一步，采用基于李群的骨架表示方法，对 3D 骨架形状的相对几何进行编码，从而得到位于特殊欧氏群积空间 $SE(3)\times\cdots\times SE(3)$ 上的动作轨迹。

第二步，针对每一个动作类别，运用本文提出的平均轨迹计算方法，计算每个动作的平均轨迹。该平均轨迹将用作轨迹时间对齐的对齐模板。

第三步，采用 TSRVF 表示方法将每条样本轨迹与平均轨迹进行对齐。

第四步，得到时间对齐后的轨迹，可以进一步实现特征提取。图 6 中第一行为特征射击向量的提取过程，计算的是平均轨迹到所有对齐后的动作轨迹的切向量。第二行为特征李代数的提取过程，计算的是每条轨迹在单位元 I_4 处的切向量。

第五步，为了解决数据噪声的问题，本文采用傅里叶时间金字塔表示方法处理两类特征，滤去高频系数。由于我们对特征的每一维都用傅里叶时间金字塔进行处理，因此最终获得的特征是所有傅里叶系数的级联。

第六步，在分类之前，我们对两种分类特征的傅里叶系数进行加权融合，然后训练一个 one-vs-all 线性支持向量机来分类特征。

5. 实验评估

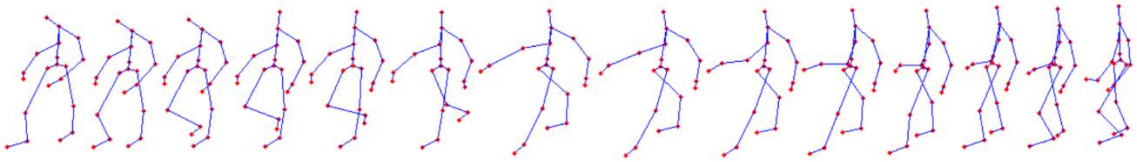
5.1. 数据集

本文采用了两个动作识别数据集，分别是UTKinect Action数据集[16]和MSR-Action3D数据集[17]。

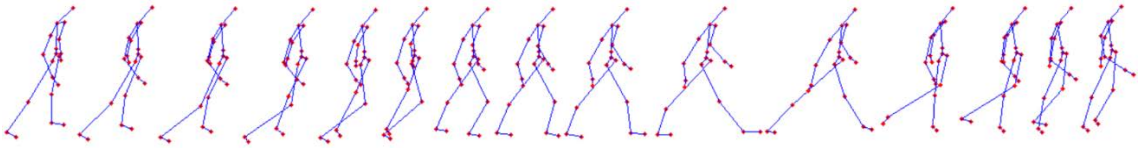
UTKinect Action 数据集：UTKinect Action 是 2012 年使用单个静态 Kinect 相机采集的数据集。它由 199 个动作序列组成，分别由 10 位演员执行 10 个动作：“walk”，“sit down”，“stand up”，“pick up”，“carry”，“throw”，“push”，“pull”，“wave hands”以及“clap hands”。每个动作重复执行两次。每帧骨架上包含 20 个关节点。这个数据集的挑战性在于，所有的动作序列并不是在同一视角下捕捉到的。如图 7 所示，(a)行和(b)行虽然都是动作“walk”的骨架序列，但是视角完全相反。因此在处理该数据集时，必须克服视角变换的差异。

MSR-Action3D 数据集：MSR-Action3D 是 2010 年由一个与 Kinect 类似的深度传感器采集的数据集。它包含 557 个动作序列，分别由 10 位演员执行 20 个动作：“high arm wave”，“horizontal arm wave”，“hammer”，“hand catch”，“forward punch”，“high throw”，“draw X”，“draw tick”，“draw circle”，“hand clap”，“two hand wave”，“side boxing”，“bend”，“forward kick”，“side kick”，“jogging”，

“tennis swing”，“tennis serve”，“golf swing”以及“pick up and throw”。每个动作重复执行两到三次。每帧骨架上包含 20 个关节点。如图 8 所示，骨架序列从左到右，展示了动作“high arm wave”的部分骨架序列。



(a) UTKinect 数据集中，动作“walk”的部分骨架序列，骨架顺序从右到左



(b) UTKinect 数据集中，动作“walk”的部分骨架序列，骨架顺序从左到右

Figure 7. Part of the skeleton sequence of the action “walk” in the two perspectives in the UTKinect data set
图 7. UTKinect 数据集中，动作“walk”在两个视角下的部分骨架序列

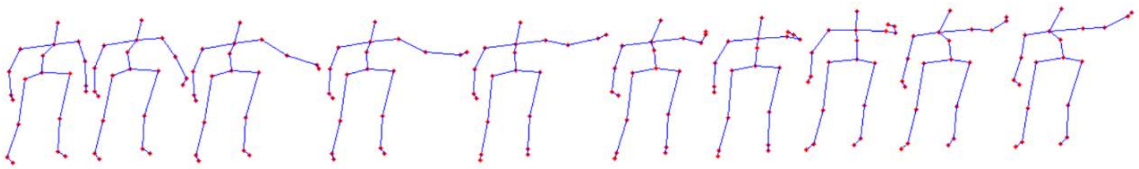


Figure 8. Part of the skeleton sequence of the action “high arm wave” in the MSR-Action3D data set
图 8. MSR-Action3D 数据集中，动作“high arm wave”的部分骨架序列

5.2. 实验设计

为了保证实验的丰富性，本文对两种分类特征以及它们的融合特征分别求取动作识别结果。假设每个动作有 T 帧骨架，每副骨架有 N 个关节点和 M 个身体部位。由于基于李群的骨架表示是对骨架骨骼之间的相对几何进行编码，因此每条动作轨迹的尺寸为 $\left[M\left(M-1\right),4\times 4,T\right]$，其中 4×4 是刚体变换矩阵的维度。令 $j=1,2,\cdots,J$ 为不同动作类别的下标，μ_j 是根据本文提出的平均轨迹计算方法得到的第 j 类动作的平均轨迹。对任意动作轨迹 β_i （i 是样本轨迹的下标），令 β_i^j 表示样本轨迹 β_i 到平均轨迹 μ_j 的最优对齐结果。接下来，我们计算射击向量 V_S 及其傅里叶系数 V_{SC}，李代数 V_L 及其傅里叶系数 V_{LC}，并进行加权融合。由于有 J 类动作，因此，可以计算得到 J 个特征射击向量 V_S 和特征李代数 V_L。

5.3. 实验参数设置

针对这两个动作数据集，本文采用[18]中提出的目标交叉实验设置(cross-subject test setting)。一半演员所做的动作用于训练分类器，剩下一半演员所做的动作用于测试。我们选择五种不同的训练和测试组合，最终的分类结果为五种组合结果的平均值。我们将算法 1 中的超参数 ϵ 设置为 0.1。在对两类分类特征进行加权融合时，我们分别测试了从 0.1 到 0.9 共 9 组参数，然后选择其中最佳的参数组合。

5.4. 实验结果及讨论

UTKinect Action 数据集上的动作分类结果如表 1 所示。特征射击向量和李代数的分类结果分别为

96.80%和 97.20%，李代数分类结果比射击向量高出 0.4%。特征融合的结果展示在图 9 中。由图可以看出，共有四组融合参数[0.7, 0.3]，[0.6, 0.4]，[0.5, 0.5]以及[0.4, 0.6]，使得融合特征取得最高的分类结果 98.20%。参数组合的前者为特征射击向量的权重系数，后者为李代数的权重系数。该融合特征的结果比射击向量和李代数的分类结果分别高出 1.4%和 1%。可以看出，在 UTKinect Action 数据集上，特征融合的操作能够有效提高动作识别的正确率。图 10 展示了 UTKinect Action 数据集上各类动作的分类正确率。10 类动作中，有 8 类动作达到了 100%的识别正确率，动作“throw”识别正确率最低，为 86%，动作“pick up”识别正确率为 96%。本文提出的动作识别框架在 UTKinect Action 数据集上取得了良好表现。

Table 1. Action classification results on UTKinect Action dataset. SV: shooting vector, LA: Lie algebra
表 1. UTKinect Action 数据集上的动作分类结果。SV：射击向量，LA：李代数

平均轨迹计算方法	分类特征(%)		
	SV	LA	SV + LA
Karcher 均值与测地距离相结合的算法 1	96.80	97.20	98.20

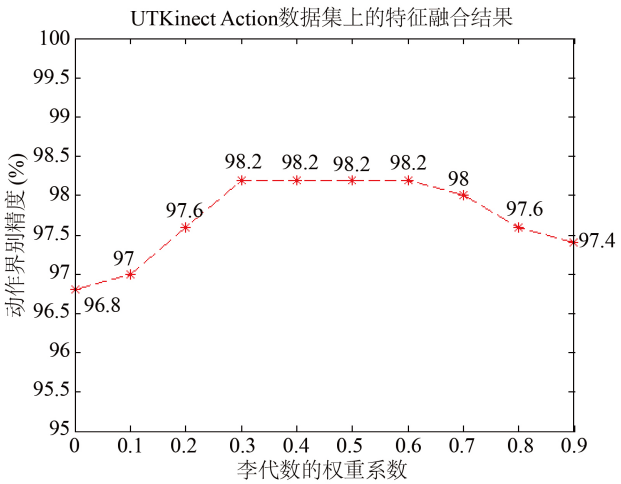


Figure 9. Feature weighted fusion on UTKinect Action dataset
图 9. UTKinect Action 数据集上的特征加权融合

MSR-Action3D 数据集上的动作分类结果如表 2 所示。特征射击向量和特征李代数的动作分类结果分别为 89.05%和 91.40%。李代数的分类正确率比射击向量高出 2.35%。最佳特征融合参数如图 11 所示，在参数组合为[0.2, 0.8]或[0.1, 0.9]时，融合特征取得最高的动作分类结果 91.48%。该融合特征结果比两个单类特征的分类结果分别高出 2.43%和 0.08%。图 12 展示了 MSR-Action3D 数据集中各项动作的识别正确率。大部分动作类别都能达到 80%以上的识别正确率。由于动作“hand catch”和“draw circle”相似度较高，因此误分类情况较多，两类动作的识别正确率仅有 31.67%和 69.33%。

通过对比三类特征的实验结果，可以发现，特征李代数的实验结果明显优于特征射击向量。这是因为在李群 SE(3)上，当两点之间的距离较远时，对数映射容易发生扭曲，从而使得特征射击向量的有效性低于特征李代数。其次，在两个数据集上，融合特征的分类正确率也要高于两个单类特征，验证了特征融合操作的有效性。

本文的动作识别实验结果也与采用了相同实验设置[18]的现有方法进行了比较。我们将现有方法的分类结果展示在了表格第一栏，表格第二栏为本文的最优分类结果。

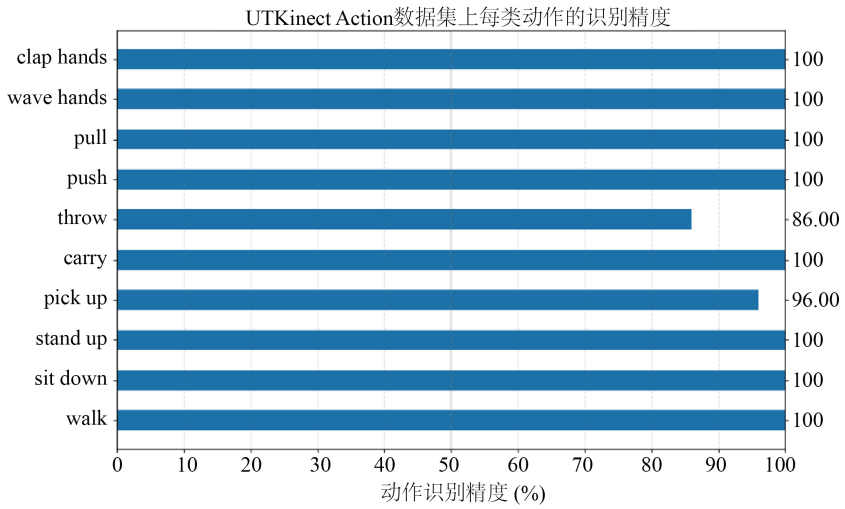


Figure 10. The classification accuracy rate of each type of action on the UTKinect Action dataset
图 10. UTKinect Action 数据集上的每类动作的分类正确率

Table 2. Action classification results on the MSR-Action3D dataset. SV: shooting vector, LA: Lie algebra
表 2. MSR-Action3D 数据集上的动作分类结果。SV：射击向量，LA：李代数

平均轨迹计算方法	分类特征(%)		
	SV	LA	SV + LA
Karcher 均值与测地距离相结合的算法 1	89.05	91.40	91.48

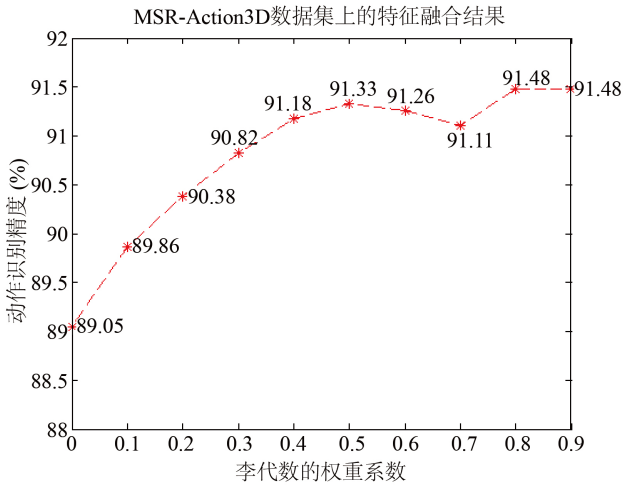


Figure 11. Feature weighted fusion on MSR-Action3D dataset
图 11. MSR-Action3D 数据集上的特征加权融合

本文方法与现有其他方法在 UTKinect Action 数据集上的动作识别正确率展示在表 3 中。由表格可知，在目标交叉实验设置下，本文实验取得了最优的动作识别结果 98.2%，比采用了相同骨架表示方法的黎曼方法 Lie Group [1]和 TSRVF on Lie Group [8]分别高出 1.12%和 3.33%。与非黎曼方法 JL-distance LSTM [19]的实验结果相比，也要高出 2.24%。与最近的 SCDL [20]方法相比，本文实验结果略高出

0.81%。本文的动作识别框架在 UTKinect Action 数据集上取得了最优的实验结果。

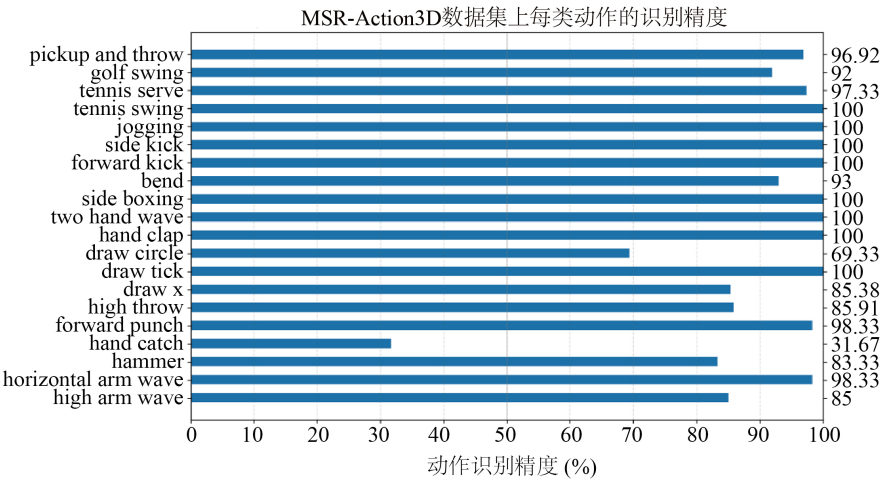


Figure 12. The classification accuracy rate of each type of action on the MSR-Action3D data set
图 12. MSR-Action3D 数据集上的每类动作的分类正确率

Table 3. Comparison of the results of the action recognition method in this paper and other methods on the UTKinect Action dataset
表 3. 本文动作识别方法与其他方法在 UTKinect Action 数据集上的结果比较

动作识别方法	UTKinect Action 数据集(%)
Lie Group [2014] [1]	97.08
TSRVF on Lie Group [2016] [8]	94.87
JL-distance LSTM [2017] [19]	95.96
SCDL [2020] [20]	97.39
Karcher 均值与测地距离相结合的算法 1 (SV + LA)	**98.20**

表 4 展示了本文动作识别方法与其他动作识别算法在 MSR-Action3D 数据集上的动作分类结果。可以看出，在目标交叉实验设置下，本文实验的动作识别正确率比采用了相同骨架表示方法的 Lie Group [1]和 TSRVF on Lie Group [8]分别高出 2%和 6.32%。比 TSRVF on *S* [9]工作高出 2.48%，该方法将骨架序列表示为 Kendall 形状空间上的轨迹，并且同样采用 TSRVF 表示方法进行时间对齐。与 SCDL [20]相比，本文实验结果仍高出 1.47%。本文的动作识别框架在 MSR-Action3D 数据集上仍取得最优的实验结果。

Table 4. The comparison of the results of the action recognition method in this paper with other methods on the MSR-Action3D dataset
表 4. 本文动作识别方法与其他方法在 MSR-Action3D 数据集上的结果比较

动作识别方法	MSR-Action3D 数据集(%)
Lie Group [2014] [1]	89.48
TSRVF on Lie Group [2016] [8]	85.16
TSRVF on *S* [2016] [9]	89.00
SCDL [2020] [20]	90.01
Karcher 均值与测地距离相结合的算法 1 (SV + LA)	91.48

6. 结论

本文主要针对轨迹时间对齐过程中，模板轨迹的求取问题，提出 Karcher 均值与测地距离相结合的平均轨迹计算方法。该方法计算所有样本轨迹的平均轨迹，然后将平均轨迹当作轨迹对齐的模板。随后，本文将该平均轨迹计算方法应用于动作识别框架中。本文动作识别框架流程清晰简洁，同时，通过实验结果可以看出，本文根据提出算法求得的平均轨迹是合适的对齐模板。然而，Karcher 均值算法存在初始值依赖的问题，可能会为结果引入偏差。因此，无偏的平均轨迹计算方法仍是一个值得研究的方向。

本文还针对动作识别框架中的特征提取，提出将特征射击向量和李代数的傅里叶系数进行加权融合。融合特征既包含了轨迹间的距离信息，也包含了轨迹自身的几何信息。实验在两个动作识别数据集上进行。通过对比三类特征的实验结果，验证了特征融合的有效性。同时，与现有的动作识别方法相比，本文提出的动作识别框架也取得了最优的实验结果。

基金项目

上海市“科技创新行动计划”(18441909000)。

参考文献

[1] Polat, E., Yeasin, M. and Sharma, R. (2003) Robust Tracking of Human Body Parts for Collaborative Human Computer Interaction. *Computer Vision & Image Understanding*, **89**, 44-69. https://doi.org/10.1016/S1077-3142(02)00031-0

[2] Mokhber, A., Achard, C. and Milgram, M. (2008) Recognition of Human Behavior by Space-Time Silhouette Characterization. *Pattern Recognition Letters*, **29**, 81-89. https://doi.org/10.1016/j.patrec.2007.08.016

[3] Chang, E. and Wang, Y.F. (2004) Introduction to the Special Issue on Video Surveillance. *Multimedia Systems*, **10**, 116-117. https://doi.org/10.1007/s00530-004-0144-5

[4] 房菲. 基于核方法的支持向量机在人体动作识别中的应用研究[D]: [硕士学位论文]. 青岛: 中国海洋大学, 2013.

[5] Chen, Y., Duff, M., Lehrer, N., *et al*. (2011) A Computational Framework for Quantitative Evaluation of Movement during Rehabilitation. *AIP Conference Proceedings*, **1371**, 317-326.

[6] Vemulapalli, R., Arrate, F. and Chellappa, R. (2014) Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Columbus, 588-595. https://doi.org/10.1109/CVPR.2014.82

[7] Su, J., Kurtek, S., Klassen, E., *et al*. (2014) Statistical Analysis of Trajectories on Riemannian Manifolds: Bird Migration, Hurricane Tracking and Video Surveillance. *Annals of Applied Statistics*, **8**, 530-552. https://doi.org/10.1214/13-AOAS701

[8] Anirudh, R., Turaga, P., Su, J., *et al*. (2016) Elastic Functional Coding of Riemannian Trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 922-936. https://doi.org/10.1109/TPAMI.2016.2564409

[9] Amor, B.B., Su, J. and Srivastava, A. (2015) Action Recognition Using Rate-Invariant Analysis of Skeletal Shape Trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 1-13. https://doi.org/10.1109/TPAMI.2015.2439257

[10] Wang, J., Liu, Z., Wu, Y., *et al*. (2012) Mining Actionlet Ensemble for Action Recognition with Depth Cameras. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, 16-21 June 2012, 1290-1297.

[11] Hall, B. (2015) Lie Groups, Lie Algebras, and Representations: An Elementary Introduction. Springer, Berlin. https://doi.org/10.1007/978-3-319-13467-3

[12] Murray, R.M., Li, Z., Sastry, S.S., *et al*. (1994) A Mathematical Introduction to Robotic Manipulation. CRC Press, Boca Raton.

[13] Zefran, M. and Kumar, V. (1998) Two Methods for Interpolating Rigid Body Motions. *Proceedings* 1998 *IEEE International Conference on Robotics and Automation*, Vol. 4, 2922-2927.

[14] Zefran, M., Kumar, V. and Croke, C. (1996) Choice of Riemannian Metrics for Rigid Body Kinematics. *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, V02BT02A030.

[15] Grove, K. and Karcher, H. (1973) How to Conjugate C^1-Close Group Actions. *Mathematische Zeitschrift*, **132**, 11-20. https://doi.org/10.1007/BF01214029

[16] Xia, L., Chen, C.C. and Aggarwal, J.K. (2012) View Invariant Human Action Recognition Using Histograms of 3D Joints. 2012 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Providence, 16-21 June 2012, 20-27. https://doi.org/10.1109/CVPRW.2012.6239233

[17] Li, W., Zhang, Z. and Liu, Z. (2010) Action Recognition Based on a Bag of 3d Points. 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 9-14. https://doi.org/10.1109/CVPRW.2010.5543273

[18] Zhu, Y., Chen, W. and Guo, G. (2013) Fusing Spatiotemporal Features and Joints for 3d Action Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Portland, 23-28 June 2013, 486-491. https://doi.org/10.1109/CVPRW.2013.78

[19] Zhang, S., Liu, X. and Xiao, J. (2017) On Geometric Features for Skeleton-Based Action Recognition Using Multilayer LSTM Networks. 2017 *IEEE Winter Conference on Applications of Computer Vision*, Santa Rosa, 24-31 March 2017, 148-157. https://doi.org/10.1109/WACV.2017.24

[20] Tanfous, A.B., Drira, H. and Amor, B.B. (2019) Sparse Coding of Shape Trajectories for Facial Expression and Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 2594-2607. https://doi.org/10.1109/TPAMI.2019.2932979