

文章编号: 2095-4980(2017)05-0841-08

基于动作标准序列的 3D 视频人体动作识别

聂 勇, 张 鹏, 冯 辉, 杨 涛, 胡 波

(复旦大学 信息科学与工程学院 智慧网络与系统研究中心, 上海 200433)

摘 要: 基于 3D 视频的人体动作识别近年来受到越来越广泛的关注。基于动态时间规整的算法考虑了动作的时序信息, 并能较好地解决人体运动在时间上的不确定性, 但是随着训练样本增加, 效率会变得较低。本文提出了一种基于动作标准序列的动作识别方法。通过特征提取将 3D 动作视频样本构建为动作序列, 在动态时间规整度量下将动作标准序列学习建模成一个序列平均的优化问题, 并使用动态时间规整重心平均算法(DBA)求解。对于动作类别类中存在显著差异的场景, 研究了多重动作标准序列学习, 并针对无监督学习的情况, 提出了 DBA-K-means 聚类算法。实验结果表明, 该方法可进一步提高动作识别的效率和准确率。

关键词: 人体动作识别; 3D 视频; 动态时间规整; 序列平均; 动作标准序列

中图分类号: TN911.73; TP391.4 **文献标志码:** A **doi:** 10.11805/TKYDA201705.0841

Human action recognition for 3D video based on action standard sequence

NIE Yong, ZHANG Peng, FENG Hui, YANG Tao, HU Bo

(Research Center of Smart Networks and Systems, School of Information Science and Engineering, Fudan University, Shanghai 200433, China)

Abstract: Human action recognition for 3D videos has taken more and more attention in recent years. Approaches based on Dynamic Time Warping(DTW) method consider the information of sequential order and can deal with the temporal uncertainty of action. But with the increase of training action samples, the efficiency of action recognition decreases. In this work, a new framework is designed for action recognition based on the action standard sequence. First, the action sequences is constructed from 3D action video samples by feature extraction. Then, the learning of action standard sequence is modeled as an optimization problem of sequence averaging under DTW measure, and the problem is solved by DTW Barycenter Averaging(DBA) algorithm. Furthermore, the learning of multiple action standard sequences is studied for the situation where there is large intra-class variation within one action category, and DBA-K-means algorithm is proposed for the unsupervised learning of multiple standard sequences. The experiment results show that both accuracy and efficiency can be improved by the proposed approach.

Keywords: human action recognition; 3D videos; dynamic time warping; sequence averaging; action standard sequence

人体的动作识别是计算机视觉领域一个十分活跃的课题^[1], 在智能视觉监控、人机交互、体感游戏、运动视频分析和视频检索等方面有着广泛的应用前景。然而, 人体的动作具有高度的灵活性, 相同的动作可以从不同视角观察, 运动速度有快有慢, 尺度大小变化、光照变化、复杂背景等等, 而且计算效率也是需要考虑的问题。一方面, 存在类似体感游戏等实时性要求的应用; 另一方面, 随着笔记本电脑、智能手机的普及, 以后动作识别将更多地应用于这些便携式设备上。这些因素给动作识别带来了很大的挑战^[2-3]。3D 视频拥有深度信息, 能够极大地提高计算机系统感知真实的三维世界和提取低层视觉信息的能力, 可以还原三维世界场景, 提取人体骨架, 不受光照条件的影响, 不依赖颜色、纹理等, 具有无可比拟的优势。而且近些年来, 随着微软公司的 Kinect, 华硕的 Xtion 以及 Lytro 光场相机等大众消费级的深度传感器的出现, 大量捕捉 3D 视频数据(RGB-D)成为可能, 近来 3D 视频的应用和研究受到越来越广泛的关注^[4]。因此, 本文将 3D 视频作为研究对

象。根据对时间维度处理方式的不同,动作识别可以分为时空类算法和序贯类算法^[1]。时空类算法,分类时忽略时序关系,主流的框架是词袋(Bag of Words, BoW)模型^[5]。序贯类算法将动作的时序关系考虑其中,所以序贯类算法通常比对应的时空类算法有更好的结果。序贯类算法,又可细分为基于范例的分类方法和基于状态模型的方法。基于范例的分类方法将人的动作作用单个或一个集合的模板的时间序列描述,然后将新的输入与定义好的序列模板比较进行分类。动态时间规整(DTW)算法被广泛应用到这类方法中^[6-8]。基于动态时间规整的范例分类方法,对人动作的快慢或者变化鲁棒性较好^[6],而且相比基于状态模型的方法,该方法需要的训练样本较少^[9]。但基于动态时间规整的动作识别方法,一般使用最近邻分类器,在动作训练样本较多的情况下,由于序列匹配的计算量大,存在分类效率低^[10]的问题。本文提出了一种基于动作标准序列的动作识别方法,实验结果表明,动作识别的效率和准确率在本文所提方法中得到了明显提高。

1 动作标准序列

动作标准序列的主要思想是,对于每个动作类别,将同类别的所有训练样本合并,求得代表这个动作类别的一个平均序列样本,那么以后对于新的动作视频输入,只需要与每个动作类别的平均序列样本匹配即可,这样可以极大提高动作识别的效率。这些平均序列样本,本文称为动作标准序列。

1.1 问题建模

设由某个动作类别的 N 个 3D 动作视频训练样本构建的动作序列集合为 M :

$$\begin{cases} M = (M_1, M_2, \dots, M_i, \dots, M_N) \\ M_i = (m_1, m_2, \dots, m_j, \dots, m_{T_i}) \end{cases} \quad (1)$$

式中: M_i 表示第 i 个样本动作序列; T_i 表示第 i 个样本序列的长度; m_j 表示样本序列中第 j 个元素。

那么,对于如何将同类别的多个动作样本序列进行合并,从而求得动作标准序列的问题,本文定义该动作类别的动作标准序列 $M_s = (m_1^s, m_2^s, \dots, m_{T_s}^s)$ 为:与该动作类别的样本序列集合中所有样本序列的距离平方和最小的序列,即满足如下表达式:

$$\forall t \in [1, +\infty], \left(\forall X \in \mathcal{E}^t, \sum_{i=1}^N \Delta^2(M_s, M_i) \leq \sum_{i=1}^N \Delta^2(X, M_i) \right) \quad (2)$$

式中 \mathcal{E}^t 表示所有长度为 t 的序列的集合。

1.2 距离度量选择

序列间距离最直观简单的方法是计算欧氏距离:

$$\left. \begin{aligned} S_A &= (a_1, a_2, \dots, a_T) \\ S_B &= (b_1, b_2, \dots, b_T) \end{aligned} \right\} \Rightarrow D(S_A, S_B) = \sqrt{\delta(a_1, b_1)^2 + \delta(a_2, b_2)^2 + \dots + \delta(a_T, b_T)^2} \quad (3)$$

但是这种欧氏距离的度量方法存在 2 个问题:一方面,它不能比较 2 个不同长度的序列,这是显而易见的;另一方面,它不能有效地表征序列间的相似度,例如:

$$S_A = (a, b, c, d, e, f, g), S_B = (b, c, d, e, f, g, h) \quad (4)$$

S_A, S_B 2 个序列的长度一致,它们大部分元素都是一致的,只是位置上偏差了 1 个元素,相似度很大,但如果用欧氏距离计算,则 2 个序列每个对应位置上都有差异,累积下来,会认为这 2 个序列差别很大。而 DTW 度量则是近年来颇受关注的一种度量方法^[11],可以较好地度量这 2 个序列间的相似度。DTW 度量是基于 Levenshtein 距离,在给定规则下,尝试寻找序列间最佳的映射关系来对齐 2 个序列,使得序列间差异最小,最大限度捕捉序列间相似性,而且对于不同长度的序列也可以很好地对齐处理。因此,本文选择 DTW 度量作为序列间的距离度量,则式(2)可表述为:

$$\forall t \in [1, +\infty], \left(\forall X \in \mathcal{E}^t, \sum_{i=1}^N DTW^2(M_s, M_i) \leq \sum_{i=1}^N DTW^2(X, M_i) \right) \quad (5)$$

式中 $DTW(\cdot, \cdot)$ 表示 2 个序列间的 DTW 度量距离。

1.3 优化问题求解

对于式(2)这类数学问题,当距离空间是线性空间时,可利用代数均值的性质求解,但是本文使用的 DTW 度量,见式(5),是半伪度量(semi-pseudometric)空间,这样的平均序列求解问题,难以精确求解,一般通过迭代近似求解^[11]。F.Petitjean 等^[11-12]提出的动态时间规整重心平均算法(DBA)就是一种通过迭代计算近似求解这个序列平均问题的方法。

DBA 算法总的来说是一种启发式算式,它通过逐步迭代计算优化初始的平均序列求解。具体来说,问题求解的原目标是寻找最佳的平均序列,使得它与集合中所有序列的 DTW 距离平方和最小。DBA 算法将横向上每个序列的距离求和,转换为纵向的序列中每个元素映射后的距离求和,如图 1 所示,然后不断优化平均序列中的元素值,使得其与映射到这个位置上的所有元素距离平方和最小化。

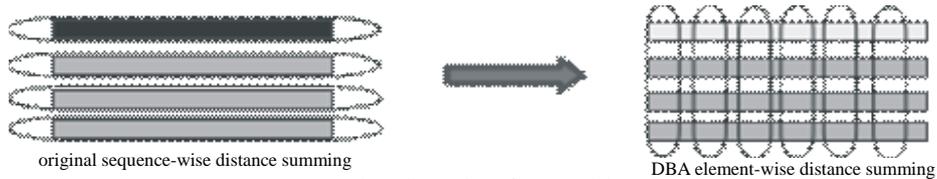


Fig.1 Illustration of DBA solving
图 1 DBA 算法求解

DBA 算法是不断迭代更新优化平均序列,那就要求给定初始的平均序列,理论上这个初始序列可以任意取值,但为了更快地收敛,一般从序列集合选取一个序列作为初始序列。设初始序列为 $C = (c_1, c_2, \dots, c_i, \dots, c_{T_c})$, 具体迭代计算更新过程分为 2 步:

1) 分别计算集合中所有序列和等更新平均序列的 DTW 映射,并以平均序列中每个元素为索引,将与之存在映射关系的集合中序列的元素关联在一起形成一个组;

2) 计算每个组的重心,并用这个值更新平均序列中对应元素的值,即 $c'_i = \text{barycenter}(\text{group}(c_i))$,其中 $\text{group}(c_i)$ 指与元素 c_i 相关联那组元素, $\text{barycenter}()$ 表示是重心计算操作, $\text{barycenter}(x_1, x_2, \dots, x_k) = (x_1 + x_2 + \dots + x_k) / k$ 。

通过 DBA 算法求解平均序列的操作记为:

$$M_s = \text{DBA}(M) \tag{6}$$

式中 M 是训练样本序列的集合,见式(1)。

2 多重动作标准序列学习

动作标准序列的学习,是用一个平均序列样本代表一个动作类别的众多训练序列样本,但是当某个动作类别情况较复杂时,类内差异导致仅用一个动作标准序列可能无法有效地描述出这个动作类别。比如,同样的一个“挥手”的动作,有人可能习惯用右手,有人可能习惯用左手;“拾起”这个动作,有人可能单手,也有人可能用双手。针对这类问题,解决的方法是在单一动作标准序列和所有训练样本序列之间折衷,为存在较大类内差异的动作类别,学习多个动作标准序列,即一个动作类别有多个“模板”,但仍相对于原始众多的训练序列样本要少得多。本文将为一个动作类别学习多个动作标准序列的过程,称为“多重动作标准序列学习”,其中每一个动作标准序列代表该动作类别的一个“子动作类”。根据条件不同,可分为有监督学习和无监督学习 2 种情况。

2.1 有监督的多重动作标准序列学习

有监督的多重动作标准序列学习,指的是已知具有 N_Q 个子动作类的动作类别 Q ,且该动作类别的所有训练样本序列根据类内差异可以划分到 N_Q 个子集合:

$$\begin{cases} Q = Q_1 \cup Q_2 \cup \dots \cup Q_{N_Q} \\ Q_i \cap Q_j = \emptyset, 1 \leq i < j \leq N_Q \end{cases} \tag{7}$$

目标是学习得到 N_Q 个能够代表该动作类别的子动作类的动作标准序列 $\{M_{s_i}\}, i = 1, 2, \dots, N_Q$ 。对于这类有监督的情况,可以直接根据式(6)求解:

$$M_{s_i} = DBA(Q_i), \quad i = 1, 2, \dots, N_Q \quad (8)$$

2.2 无监督的多重动作标准序列学习

无监督的多重动作标准序列学习,指的是对于一个类动作类别,希望学习多个动作标准序列以便更好地表征该动作类别的类内差异,但并不知道该动作类别中的训练样本具体应该划分到哪个子动作类。这可以认为是一个聚类问题:给定一个动作样本序列的集合 Q ,要求将 Q 划分成 N_Q 个子序列集合,并且每个子序列集合内部的序列元素之间相异度尽可能低,而不同子序列集合的序列元素尽可能高,每个类簇的中心就是待求的动作标准序列。**K-means**是聚类算法的典型代表,但这个聚类过程并不能直接用**K-means**方法完成,因为**K-means**算法是以欧式距离作为相似度测度设计的,采用误差平方和为聚类准则函数,要求均值是类簇中心的最佳估计,而动作标准序列的学习是基于DTW度量,并不满足**K-means**的算法前提。为解决这个问题,本文提出了DTW度量下的“DBA-K-means聚类算法”。

2.3 DBA-K-means 算法

K-means是最常用的聚类方法之一,其核心思想比较简单:把给定的 n 个点划分到 k 个聚类中,使得每个点都属于距离最近的均值(即聚类中心点)对应的聚类,以之作为聚类的标准。**K-means**聚类不能直接用于动作标准序列的学习,是因为应用场景中距离度量从欧氏度量变成了DTW度量,需要对优化目标、类簇中心的计算等做相应调整。本文称DTW度量下的**K-means**聚类问题为DBA-K-means聚类。

2.3.1 算法描述

DTW度量下,对应的DBA-K-means聚类问题描述为:已知序列集合 $Q = \{M_1, M_2, \dots, M_i, \dots, M_N\}$,要求把这 N 个序列元素划分到 k 个子序列集合 $S_Q = \{Q_i\}, i = 1, 2, \dots, k$ 中,使得相应子序列集合的组内平方和最小,满足:

$$\arg \min_{S_Q} \sum_{i=1}^k \sum_{M \in Q_i} DTW^2(M, M_{s_i}) \quad (9)$$

式中 M_{s_i} 是集合 Q_i 中所有序列的均值中心,通过DBA算法计算,即: $M_{s_i} = DBA(Q_i), i = 1, 2, \dots, k$ (10)

求解这个问题,DBA-K-means聚类算法的具体描述如下:

DBA-K-means 聚类算法
参数: N_{it} ——最大迭代次数
输入: 序列集合: $Q = \{M_1, M_2, \dots, M_i, \dots, M_N\}$
类簇的数目: k
初始化:
步骤 1: 在 Q 中随机取 k 个序列元素, 作为 k 个簇各自的中心, $M_{s_i}^{(0)}, i = 1, 2, \dots, k$, $S_Q^{(0)} = \emptyset$
迭代计算:
for $t = 1, 2, \dots, N_{it}$ do
步骤 2: 分别计算 Q 中所有序列元素到 k 个簇中心 $M_{s_i}^{(t-1)}, i = 1, 2, \dots, k$ 的 DTW 度量, 将元素分别归属于 DTW 度量最小的簇,
得到 $S_Q^{(t)} = \{Q_i^{(t)}\}, i = 1, 2, \dots, k$
步骤 3: 根据聚类结果和式(10), 重新计算 k 个簇各自的中心, 得到 $M_{s_i}^{(0)}, i = 1, 2, \dots, k$;
if $S_Q^{(t)}$ 等于 $S_Q^{(t-1)}$ then
步骤 4: 类簇不再变化, 跳出迭代循环
end if
end for
输出: 类簇元素集合: $S_Q = \{Q_i\}, i = 1, 2, \dots, k$
类簇中心: $M_{s_i}, i = 1, 2, \dots, k$

2.3.2 收敛性证明

DBA-K-means 聚类算法的目标,就是最小化代价函数:

$$J = \sum_{i=1}^k \sum_{M \in Q_i} DTW(M, M_{s_i})^2 \quad (11)$$

从一次循环的迭代过程分析, 设迭代前代价为 J_0 :

1) 固定簇中心 $M_{s_i}, i=1,2,\dots,k$: 集合 Q 中的所有元素都分别归属于 DTW 度量最小的那个簇中心所在的簇, 这保证了更新归属后的代价 $J_1 \leq J_0$;

2) 固定元素的归属簇: 根据式(10), 重新计算 k 个簇各自的中心, 由于 DBA 算法本身是求解使得在 DTW 度量下距离平方和最小的平均序列, 因此, 更新簇中心后的代价 $J_2 \leq J_1 \leq J_0$ 。

从迭代过程的分析可知, 代价 J 是单调递减的, 且 $J \geq 0$; 当 J 不变时, 会跳出循环, 因此 DBA-K-means 聚类算法一定收敛。

3 动作识别框架

学习得到动作的标准序列后, 将其应用于动作识别框架中, 图 2 为基于动作标准序列的动作识别方法的框图。相对于一般序列匹配动作识别框架, 本框架多了动作标准序列学习模块, 并将学习得到的动作标准序列传递给最近邻分类器。这样在最近邻分类时, 对于每个动作类别, 都只有一个或少数几个动作标准序列的样本需要匹配计算。

设动作训练集合共有 N_{action} 种动作类别, 学习得到所有动作类别的动作标准序列集合为 $S_{standard}$

$$\begin{cases} S_{standard} = \{T_1, T_2, \dots, T_i, \dots, T_{N_{action}}\} \\ T_i = \{t_1^i, t_2^i, \dots, t_{N_i}^i\} \end{cases} \quad (12)$$

其中 T_i 表示第 i 个动作类别的动作标准序列集合, N_i 表示第 i 个动作类别的动作标准序列的个数。

那么, 对于新的待分类动作视频输入, 提取其动作特征序列 v , 与每个动作类别的动作标准序列进行匹配, 计算得到 DTW 距离最小的动作标准序列, 其所在的动作类别 T_v 即为分类结果。

$$\forall T \in S_{standard}, \forall y \in T, (\exists x \in T_v, T_v \in S_{standard}, DTW(x, v) \leq DTW(y, v)) \quad (13)$$

4 实验结果

本文分别在 MSR Action 3D^[13]和 UTKinect^[14]2 个公开的 3D 动作数据集上对所提方法进行测试。

MSR Action 3D 动作数据集是由微软雷德蒙德研究院和卧龙岗大学收集整理的公开测试集。这个数据集共有 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, pick up&throw。由 10 个不同的实验者执行动作, 每个动作分别重复 2 次~3 次, 共 567 个动作实例。所有的动作视频都是从固定角度拍摄, 实验者在执行动作时都是面向摄像头。

UTKinect 动作数据集是由德克萨斯大学奥斯汀分校收集整理的公开测试集。这个数据集共有 10 个动作: walk, sitDown, standup, pickup, carry, throw, push, pull, waveHands, clapHands。由 10 个不同的实验者执行动作, 每个动作分别重复 2 次, 共 200 个动作实例。实验者不是固定地面对摄像头, 不同的动作角度会有改变, 而且不同的实验者执行动作时的角度是由各自决定的。

在 MSR Action 3D 数据集上, 测试了如下 2 种配置下的实验结果。图 3 给出了各配置下不同算法用时的直观对比结果。

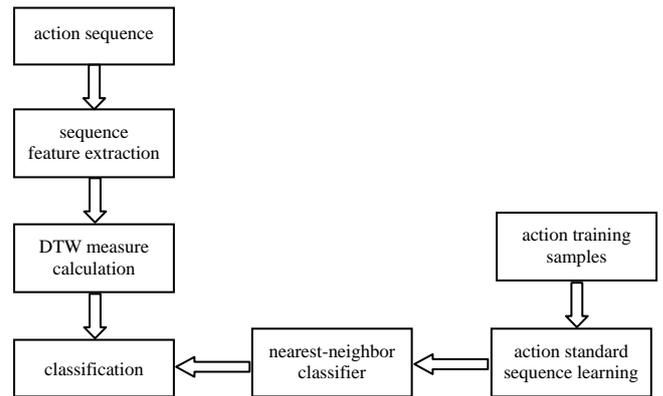


Fig.2 Diagram of action recognition based on action standard sequence
图 2 基于动作标准序列的动作识别框图

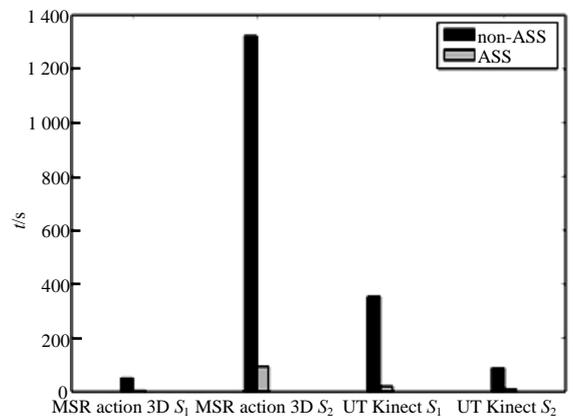


Fig.3 Time comparison of different algorithms and settings
图 3 不同算法计算时间对比

配置一：使用实验条件^[15]，选择 high arm wave, horizontal arm wave, hand clap, two-hand wave 等动作子集，其中 5 个实验者的动作视频作为训练集，余下用作测试。表 1 给出了本文算法此配置下的实验结果。

配置二：在整个数据集上进行测试，将数据集一半用于训练，一半用于测试数据，具体地，将序号为 (1,3,5,7,9) 的实验者视频用于训练，序号为 (2,4,6,8,10) 的实验者视频用于测试验证。表 2 给出了本文算法此配置下的实验结果。

在 UTKinect 数据集上，测试了 2 种配置。

配置一：使用留一交叉验证 (Leave One Out Cross Validation, LOOCV)，即 200 个动作序列中只有 1 个序列是用来交叉验证。表 3 和图 4 给出了本文方法在此配置下整个数据集上的实验结果和混淆矩阵。

配置二：将数据集一半用于训练，一半用于测试数据，具体地，将序号为 (1,3,5,7,9) 的实验者视频用于训练，序号为 (2,4,6,8,10) 的实验者视频用于测试验证。表 4 和图 5 分别给出了本文方法在此配置下整个数据集上的实验结果和混淆矩阵。

算法用时的测量环境：Intel® Core™ i7-2600 CPU @ 3.40GHz × 8 的台式电脑，8 G 内存，操作系统 ubuntu 14.04 LTS x86_64，基于 python 语言编写算法程序。

实验中，对比了“动作标准序列分类” (Action Standard Sequence Method, ASSM) 和“多训练样本序列分类” (Non-ASS) 2 种算法。“动作标准序列分类”指的是本文提出的基于动作标准序列的动作识别方法；“多训练样本序列分类”指不学习动作标准序列，直接用每个动作类别所有训练样本序列进行匹配的动作识别方法，这一方法作为参考基准 (baseline)。

表 1 MSR action 3D 数据集实验结果

Table 1 Results of MSR action 3D with setting 1

	accuracy/%	time/s
non-ASS	85.00	51.8
ASS	93.33	4.4

表 3 UT Kinect 数据集配置一实验结果

Table 3 Results of UT Kinect with setting 1

	accuracy/%	time/s
non-ASS	89.45	354.6
ASS	92.46	19.6

表 2 MSR action 3D 数据集实验结果

Table 2 Results of MSR action 3D with setting 2

	accuracy/%	time/s
non-ASS	63.01	1 322.2
ASS	65.73	94.4

表 4 UT Kinect 数据集配置二实验结果

Table 4 Results of UT Kinect with setting 2

	accuracy/%	time/s
non-ASS	85.00	88.2
ASS	92.00	9.62

从表 1~表 4 的多种配置下实现结果数据综合分析可以得出，“动作标准序列分类”相比“多训练样本序列分类”在识别的准确度和效率上都有提高。从图 4 和图 5 的混淆矩阵可以看出，本文算法在许多动作类别的识别上都达到了 100% 的正确率。而从配置二条件下的正确率为 92.00%，可以看出，本文算法在较少的训练样本的情况下也能取得较好结果。

表 5 给出了本文算法与已有研究的其他一些算法在 UTKinect 数据集上的正确率对比结果。从表 5 可以看出，本文算法在 UTKinect 数据集表现效果很好，从正确率来看比已有算法^[14,16-17]都有更好的识别效果。

walk	1	0	0	0	0	0	0	0	0	0
sit down	0.05	0.95	0	0	0	0	0	0	0	0
stand up	0	0	1	0	0	0	0	0	0	0
pick up	0.1	0.1	0.05	0.7	0.05	0	0	0	0	0
carry	0	0	0	0	1	0	0	0	0	0
throw	0	0	0	0	0	0.65	0.15	0.05	0	0.15
push	0	0	0	0	0	0	1	0	0	0
pull	0	0	0	0	0	0	0	1	0	0
wave hands	0	0	0	0	0	0	0	0	1	0
clap hands	0	0	0	0	0.05	0	0	0	0	0.95

Fig.4 Confusion matrix for UT Kinect dataset with setting 1

图 4 UT Kinect 数据集上配置一的混淆矩阵

walk	1	0	0	0	0	0	0	0	0	0
sit down	0.1	0.9	0	0	0	0	0	0	0	0
stand up	0	0	1	0	0	0	0	0	0	0
pick up	0	0	0	1	0	0	0	0	0	0
carry	0	0	0	0	0.8	0	0	0	0	0.2
throw	0	0	0	0	0.1	0.5	0.2	0.1	0	0.1
push	0	0	0	0	0	0	1	0	0	0
pull	0	0	0	0	0	0	0	1	0	0
wave hands	0	0	0	0	0	0	0	0	1	0
clap hands	0	0	0	0	0	0	0	0	0	1

Fig.5 Confusion matrix for UT Kinect dataset with setting 2

图 5 UT Kinect 数据集上配置二的混淆矩阵

表 5 UT Kinect 数据集上不同算法的正确率比较

Table 5 Accuracy comparison of different algorithms for UT Kinet dataset

	accuracy/%	year
Xia et al ^[14]	90.92	2012
Zhu et al ^[16]	91.90	2013
Slama et al ^[17]	91.21	2015
proposed	92.46	—

5 结论

本文提出了一种基于动作标准序列的动作识别方法,解决在训练序列样本较多的情况下,序列匹配动作识别方法存在的问题。在DTW度量下,将动作标准序列学习建模成一个序列平均的优化问题,并使用DBA算法求解;在多重动作标准序列学习的研究中,针对无监督学习的情况,提出了DBA-K-means聚类算法。最后得到在MSR Action 3D和UTKinect两个数据集上的实验结果,表明本文所提方法动作识别的准确率和效率都有明显提高。但也存在一些问题和有待研究的地方,比如,虽然本文提出的方法在UTKinect数据库上有着优秀的表现,许多动作的识别率都达到了100%,但分析可以发现,识别准确率较低的动作主要是throw, pickup等相对复杂动作,这一方面是由于方法中使用的肢体角度特征较简单,而另一方面,更重要的是要从模型角度出发,由单层模型向多层模型、组合模型转变。因为多层次多特征模型更符合实际情况,更有前景,将复杂的人的动作、行为分解,各个模块负责各自部分的简单处理,而不是企图用一层结构、一个特征就实现所有功能。而且基本动作的研究,最终是希望向复杂的高级的人体行为研究迈进,越往上层,就越需要多层次多特征模型来处理。

参考文献:

- [1] CHENG G, WAN Y, SAUDAGAR A N, et al. Advances in human action recognition: a survey[J]. Computer Science, 2015.
- [2] 陈万军, 张二虎. 基于深度信息的人体动作识别研究综述[J]. 西安理工大学学报, 2015(3):253-264. (CHEN W J, ZHANG E H. A review for human action recognition based on depth data[J]. Journal of Xi'an University of Technology, 2015(3):253-264.)
- [3] POPPE R. A survey on vision-based human action recognition[J]. Image and Vision Computing, 2010, 28(6):976-990.
- [4] HAN J, SHAO L, XU D, et al. Enhanced computer vision with microsoft kinect sensor: a review[J]. IEEE Transactions on Cybernetics, 2013, 43(5):1318-1334.
- [5] FARAKI M, PALHANG M, SANDERSON C. Log-euclidean bag of words for human action recognition[J]. Iet Computer Vision, 2014, 9(3):331-339.
- [6] SEMPENA S, MAULIDEVI N U, ARYAN P R. Human action recognition using dynamic time warping[C]// 2011 International Conference on Electrical Engineering and Informatics (ICEEI). [S.l.]: IEEE, 2011:1-5.
- [7] WANG J, ZHENG H. View-robust action recognition based on temporal self-similarities and dynamic time warping[C]// 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE). [S.l.]: IEEE, 2012:498-502.
- [8] 傅颖, 郭晶云. 基于动态时间规整的人体动作识别方法[J]. 电子测量技术, 2014, 37(3):69-72. (FU Y, GUO J Y. Dynamic time warping-based human action recognition[J]. Electronic Measurement Technology, 2014, 37(3):69-72.
- [9] KULKARNI K, EVANGELIDIS G, CECH J, et al. Continuous action recognition based on sequence alignment[J]. International Journal of Computer Vision, 2015, 112(1):90-114.
- [10] 李瑞峰, 王亮亮, 王珂. 人体动作行为识别研究综述[J]. 模式识别与人工智能, 2014, 27(1):35-48. (LI R F, WANG L L, WANG K. A survey of human body action recognition[J]. Pattern Recognition & Artificial Intelligence, 2014, 27(1):35-48.)
- [11] PETITJEAN F, KETTERLIN A, GANCARSKI P. A global averaging method for dynamic time warping, with applications to clustering[J]. Pattern Recognition, 2011, 44(3):678-693.
- [12] PETITJEAN F, GANCARSKI P. Summarizing a set of time series by averaging: from Steiner sequence to compact multiple alignment[J]. Theoretical Computer Science, 2012, 414(1):76-91.
- [13] LI W, ZHANG Z, LIU Z. Action recognition based on a bag of 3D points[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.]: IEEE, 2010:9-14.
- [14] XIA L, CHEN C C, AGGARWAL J K. View invariant human action recognition using histograms of 3D joints[C]// 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.]: IEEE, 2012:20-27.
- [15] PHAM C H, LE Q K, LE T H. Human action recognition using dynamic time warping and voting algorithm[J]. VNU Journal of Science: Computer Science and Communication Engineering, 2014, 30(3):22-30.
- [16] ZHU Y, CHEN W, GUO G. Fusing Spatiotemporal Features and Joints for 3D Action Recognition[C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.]: IEEE Computer Society, 2013:486-491.
- [17] SLAMA R, WANNOUS H, DAOUDI M, et al. Accurate 3D action recognition using learning on the Grassmann manifold[J]. Pattern Recognition, 2015, 48(2):556-567.

作者简介:



聂 勇(1990-), 男, 湖南省沅江市人, 在读硕士研究生, 主要研究方向为图像处理、模式识别. email: yongnie13@fudan.edu.cn.

杨 涛(1970-), 男, 陕西省汉中市人, 副教授, 主要研究方向为宽带无线通信理论与信号处理.

张 鹏(1990-), 男, 山东省济南市人, 在读硕士研究生, 主要研究方向为图像处理、模式识别.

冯 辉(1980-), 男, 江苏省常州市人, 讲师, 主要研究方向为分布式信号处理理论与应用.

胡 波(1968-), 男, 江苏省常州市人, 教授, 博士生导师, 主要研究方向为数字信号处理、数字通信和系统设计.

“第十四届卫星通信学术年会”征文通知

第十四届卫星通信学术年会拟定于 2018 年 3 月 22 日在北京召开。本届年会将由中国通信学会卫星通信委员会和中国宇航学会卫星应用专业委员会共同主办, 年会面向各界公开征集论文, 现将有关事项通知如下:

一、征文范围 (包括但不限于):

- 1、卫星通信和通信卫星新理论、新技术及发展前沿;
- 2、卫星通信、卫星广播融合卫星导航及卫星遥感技术在各领域中的应用;
- 3、卫星移动通信新技术发展及与陆地移动 4G、5G 系统的融合互联;
- 4、卫星通信在大数据、云计算、互联网+和移动互联网时代的技术应用及技术特点;
- 5、S 波段移动卫星通信和 Ka 波段 HTS 卫星通信技术与发展;
- 6、量子通信和激光通信卫星等新技术和新业务;
- 7、卫星通信在航空、航海、极地科考、深空探测和其它领域的应用;
- 8、小卫星及星座技术的新发展和应用;
- 9、软件定义的卫星通信系统及终端;
- 10、卫星通信系统中的天线、功放、变频、调制解调器新技术及应用;
- 11、卫星轨道频谱资源的研究和有效利用;
- 12、卫星通信的标准化及互联互通;
- 13、先进的卫星通信系统、地球站设备及技术 (包括全 IP 等各类协议及灵活的网络拓扑结构和管理形式);
- 14、卫星通信信号分析/处理新技术;
- 15、卫星通信的军民融合应用与发展;
- 16、卫星通信产业化发展及创新。

二、来稿要求:

- 1、内容新颖具体, 突出作者的创新与成果, 具有较高的学术价值与应用推广价值, 体现学术和应用上的前瞻性, 且未在国内公开发行的刊物或会议上发表或宣读过。文稿应通过保密安全审查, 文责自负。
- 2、应征论文应在 3000-6000 字, 要求主题明确、文字通顺; 图中符号与文章中的一致; 文章内的名词术语、公式符号统一; 文章字体字号: 题目为黑体小二, 正文为宋体五号。
- 3、论文投稿为中文全文投稿方式, 投稿稿件请用 Word 排版。
- 4、论文需按如下顺序撰写: 题目、作者姓名、作者所在单位、城市、邮编、摘要、关键词、正文、参考文献、作者简介(姓名、性别、职称、职务、研究领域); 中文论文需提交其对应英文题目、作者姓名、单位、摘要和关键词。
- 5、所有论文必须提供至少一个可靠的 E-mail 地址, 并在稿件最后注明联系人详细的联系方式, 包括详细的通信地址、邮编、联系电话、手机、电子信箱等。

三、重要日期: 论文截稿日期: 2018 年 2 月 23 日 出版日期: 2018 年 3 月 21 日 会议日期: 2018 年 3 月 22 日

四、论文出版:

- 1、编印“第十四届卫星通信学术年会论文集”, 所有收录文章将被 CNKI 重要学术会议论文数据库全文检索;
- 2、年会将评选优秀论文并向相关核心期刊推荐, 包括已获 SCI 检索的《China Communications》英文刊及其它出版机构 (需同时提交英文文稿)。

五、联系方式: 翟峰 手机: 13901128671 电子信箱: csatcc@126.com

中国通信学会卫星通信委员会
中国宇航学会卫星应用专业委员会