基于3D人体骨架的动作识别

张友梅,常发亮,刘洪彬
(山东大学控制科学与工程学院,山东济南 250061)

摘 要: 本文提出了一种基于3D人体骨架的动作识别方法.该方法以3D人体骨架为基础,将骨架中关节点的位置重新定义,形成简化的立体骨架模型,进而采用改进的动态时间规整算法(Reformative Dynamic Time Warping,R-DTW)对齐动作序列并进行识别.由于人体大小、形状、动作方式等差异,任意两个人表达同一动作都不尽相同,简化的立体骨架模型能有效缓解这种类内差异性.传统的DTW算法存在计算复杂性高,效率低的问题,本文在传统算法的基础上设计了"一次规划,二次细化"的方法,有效降低计算量,提高计算效率.该算法在MSR 3D Action数据库上的实验验证了其有效性.

关键词: 人体骨架;动态时间规整;动作识别
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2017)04-0906-06
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2017.04.020

Action Recognition Based on 3D Skeleton

ZHANG You-mei,CHANG Fa-liang,LIU Hong-bin
(*School of Control Science and Engineering,Shandong University,Jinan,Shandong* 250061,*China*)

Abstract: This paper presents an action recognition method based on 3D skeleton. This method redefines the coordinates of the articulations which belong to the skeleton to form a simplistic skeleton model firstly. Then a reformative dynamic time warping (R-DTW) algorithm is applied to implement action recognition. There are no two persons identical in an action owing to the difference of body size,shape and action expression. The simplistic skeleton model could decrease this intra-class variability effectively. The drawbacks of conventional DTW algorithm lie in high computational complexity and low recognition efficiency. To solve this problem,we design a method named "Planning & Refining". We conduct this algorithm on MSR Action3D dataset and the results demonstrate its effectiveness.

Key words: skeleton;reformative dynamic time warping(R-DTW);action recognition

1 引言

近年来,人体动作识别在人机交互、智能监控等领域应用广泛,已经成为机器视觉和模式识别领域的一个研究热点.动作识别问题主要围绕图像和视频展开,而当今的一些研究开始在图像的基础上结合深度信息以提高动作识别率[1,2]或直接使用深度信息以加快识别速度[3,4].随着深度相机、Kinect等技术的发展,这些工作变得切实可行.

结合深度信息后的动作数据包含更丰富的信息,同时也不可避免的导致数据冗余的问题,如何从这些数据中获取表达动作的关键信息也成为了值得研究的课题.早在1975年,Johansson[5]进行了一项实验:在黑暗的屋子里将几束光照在人体的主要关节点并根据这些光点进行人体动作识别,实验结果验证了采用骨架节点进动作识别的可行性.Ziaeefard[6]在二维图像的人体轮廓剪影中提取骨架点,并将整个视频序列缩略成骨架点累积图模型(Cumulative Skeletonized Images,CSI),进而根据CSI得出骨架点分布的直方图并进行识别,该方法在KTH行为数据库上获得了迄今为止的最高识别率.Li[3]提出了一种简单有效的从深度图中提取3D点云姿态模型的方法,并验证了该方法对于遮挡的处理能力.Xia[4]将Ziaeefard的方法扩展到三维空间中,采用隐马尔科夫模型(Hidden Markov Model,HMM)进行动作识别.

本文针对3D骨架模型中关节点的空间位置设计

收稿日期:2016-01-06;修回日期:2016-06-24;责任编辑:李勇锋
基金项目:国家自然科学基金(No.61273277,No.61673244);高等学校博士学科点专项科研基金(No.20130131110038)

了简化的立体骨架模型,这种模型能有效降低因不同人体的大小、形状及动作方式所产生的类内差异性. 为解决不同样本的时变匹配问题本文采用了动态时间规整算法对样本序列进行对齐,为缓解该算法计算量大的问题对其进行了改进:在相似距离计算过程中设计了"一次规划,二次细化"的算法,在保证识别精度的同时提高了计算效率.

2　基于骨架的动作识别方法

基于3D骨架模型的动作识别方法整体流程如图1所示,图中每个人体骨架代表一个动作序列. 该方法首先对数据进行预处理,将所有骨架归一化到同一位置下,保证了样本数据的统一性;在此基础上,建立了简化的立体骨架模型,有效降低动作的类内差异性;最后,采用改进的DTW算法对训练和测试样本进行相似距离计算,从而实现动作序列的快速匹配分析与识别.

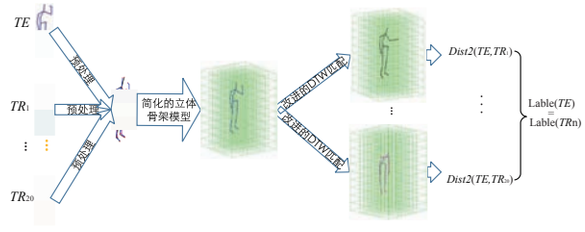


图1　动作识别方法流程

2.1　数据预处理

本文选取了MSR 3DAction数据库作为实验数据. 该数据库广泛应用于国内外基于骨架的行为识别算法的评估,共557个行为序列,包含20种人体行为.

参与者在表达某些动作(如慢跑)时并非保持在某个位置,而是会向不同方向及位置变换,因此不同参与者及同一个参与者在不同时刻的整体位置都会发生较大变化,为保证样本数据的统一性,将每个样本序列中的所有骨架以图2所展示的7号关节点为基准归一化到同一位置下.

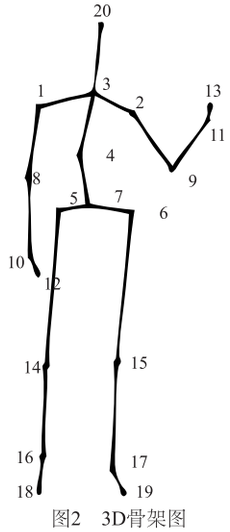


图2　3D骨架图

假设基准点坐标为 $Base=(x,y,z)$,各样本的关节点坐标为 $Samp_i^k$,其中 i 代表样本序列号,k 代表骨架节点序列号,经过预处理后样本新坐标为 $NSamp_i^k$. 预处理方法如式(1)所示.

$$\begin{aligned} d &= Samp_i^7 - Base \\ NSamp_i^7 &= Samp_i^k - d \end{aligned} \tag{1}$$

2.2　简化的立体骨架模型

图3(*a*)展示的是空间坐标系中两个人体骨架 S 和 $\hat{S}$ 在正常站立时的状态,其相似距离的计算公式如下:

$$\begin{aligned} d(S,\hat{S}) = \sum_{k=1}^{20} [&(x_k-\hat{x}_k)^2+(y_k-\hat{y}_k)^2 \\ &+(z_k-\hat{z}_k)^2] \end{aligned} \tag{2}$$

结合图3(*a*)和式(2)可知,在空间坐标系中两个骨架相应点对间存在空间位置偏移,采用原始空间坐标计算两骨架的相似距离会出现较大偏差.

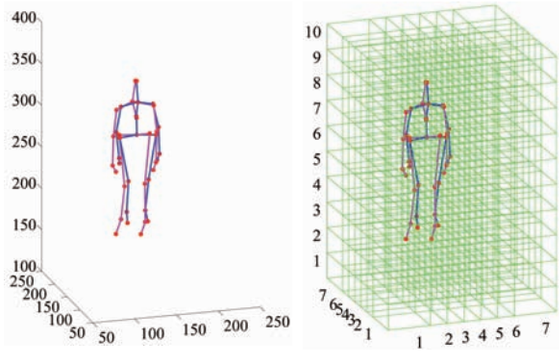
针对以上问题,本文设计了简化的立体骨架模型,如图3(*b*)所示:首先在3D坐标系中将三维空间分成若干网格,根据实验数据中骨架点分布的紧密度在 Z 方向上将坐标100~400以步长30划分成了10个区间;在 X、Y 方向上均将坐标100~200以步长20划分了5个区间,并将边缘上的50~100和200~250分别划分成单独的区间,然后将骨架点的空间坐标映射到这些网格中,用该网格区间坐标替换关节点的原始空间坐标. 网格划分的步长为定值,若步长较小则接近于原始坐标,步长过大则无法区分待匹配的骨架节点对,步长最终根据人体骨架相连关节点的长度以及选取不同步长的实验结果进行设定.

在简化的立体骨架模型中,如果两个相应节点投影在同一网格,则其网格坐标相同. 图3(*c*),3(*d*)分别展示了在 XZ 和 YZ 方向上的模型坐标表示. 图中方框所包含的两个相应骨架点在同一个网格中,则其欧氏距离为0. 椭圆框中两个相应骨架节点在网格坐标中仍然存在差距,因此简化的立体骨架模型并未完全消除这些偏差,但由于模型匹配累加了所有相应点对的欧式距离,只要能消除部分偏差,依然能有效降低因人体骨架大小、形状不同所产生的类内差异性.

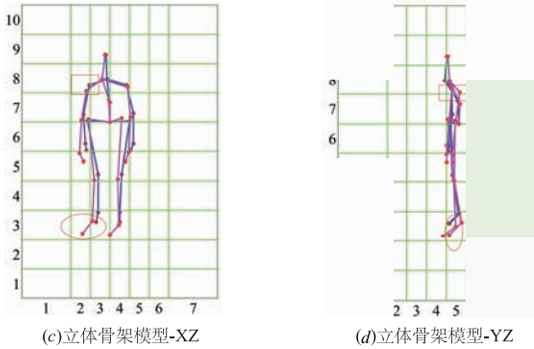
2.3　改进的DTW算法

DTW算法由日本学者Itakura提出[7],广泛应用在非等长时间序列匹配中,如语音识别[8]和动作识别[9]领域.

采用DTW算法进行行为识别本质上是进行模板匹配,假定训练样本(标准模板)*TR* 和测试样本(待匹配样本)*TE* 分别为:



(*a*)空间坐标系 (*b*)立体骨架模型-3D



(*c*)立体骨架模型-XZ (*d*)立体骨架模型-YZ

图3 简化的立体骨架模型

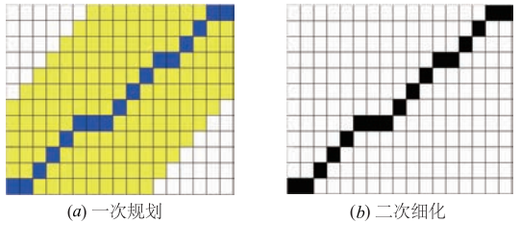
$$TR=(r_1,r_2\cdots,r_n)$$

$$TE=(e_1,e_2\cdots,e_m)$$

为实现两个不等长样本序列的匹配,首先计算两个样本中任意两帧间的相似距离 $d(r_n,e_m)$,得出相似距离矩阵 d^{N*M},进而获得两个样本的最终相似距离 $Dist[TR,TE]$.

DTW 算法简洁,但计算量大,运算效率低,因此出现了很多针对其计算量的改进算法[10~12]. DTW 算法匹配过程的低运算效率主要体现在相似距离的计算和最优路径的匹配,本文针对这两个方面设计了“一次规划,二次细化”的算法,如图 4 所示.

输入数据的维度是影响计算量的一大因素,因此本文首先采用头部和四肢,共 5 个节点进行路径规划,即“一次规划”过程,如图 4(*a*). 这 5 个关节点既能保证人体姿态的完整性,又降低了数据维度. 图 4(*a*)中每个方格代表了两个动作序列中两帧间的相似距离,白色方格是通过路径约束所避免的计算,其它为基于 5 个关节点所进行的匹配计算,蓝色方格为最终的匹配路径. “一次规划”过程的目标是获取两个样本时变匹配过程中的最佳路径 *Path*,其流程如下:



(*a*) 一次规划 (*b*) 二次细化

图4 改进的DTW

一次规划过程
输入:$\{TR_i^5,TE_i^5\}$
输出:$\{Path^{NR*NE}\}$
for $i=1:NR$
  for $j=1:NE$
    $Dist_1(i,j)=Dist[TR_i^5,TE_i^5]$;
  end
end

其中 TR_i^5 和 TE_i^5 表示仅采 5 个节点的训练和测试样本,*NR* 和 *NE* 分别为训练样本和测试样本的数量.

为更全面地体现动作细节,“二次细化”过程采用所有关节点,根据所规划的路径计算两样本序列的相似距离,如图 4(*b*)所示. 白色方格部分无需再进行计算,仅根据黑色方格所示路径采用 20 个骨架节点计算两个动作序列间的相似距离,其流程如下:

二次细化过程
输入:$\{TR_i^{20},TE_j^{20},Path^{NR*NE}\}$
输出:$\{Dist_2^{NR*NE}\}$
for $i=1:NR$
  for $j=1:NE$
    $Dist_2(i,j)=\sum_{Path\{i,j\}}d(tr,te)$
  end
end

$Dist_2(i,j)$ 为 TR_i 和 TE_j 的最终相似距离.

算法目标是判定样本序列行为类别,测试样本与所有训练样本之间最小相似距离 $Dist[TE,(TR_1,TR_2,\ldots TR_N)]$都已计算出,则 *TE* 的类别跟与其有着最近相似距离的训练样本 TR_n 的类别相同,即 $Label(TE)=Label(TR_n)$.

3 实验结果

3.1 与其他动作识别方法的比较

本文的实验设置参照文献[2],将数据库中的行为分为三组. AS1 和 AS2 旨在识别较为相似的动作,AS3 则集合了相对复杂的动作,如表 1 所示. 对于每组行为都进行了 3 次实验,Test1 和 Test2 分别随机抽取了所有数据中的 1/3 和 2/3 作为训练数据,为验证方法的泛化能力,实验还进行了交叉验证. 对比实验结果见图 5.

根据图 5 所展示的实验结果,文献[13]利用了新

的坐标系统得到了关键节点间的角度信息,实验效果最好。文献[4]效果次优,但需要利用标准数据提前训练得到较好的模型。文献[3]采用的数据除骨架节点外还有从深度图中得到的人体边缘关键点,数据信息更为丰富,在进行交叉验证时的结果与本文算法相差不大。因此本文算法具有一定的泛化能力。

3.2　简化的立体骨架模型有效性验证

在2.2节中提到简化的立体骨架模型能有效减小因人体骨架大小、形状不同所产生的匹配偏差,为验证该理论,本文采用预处理后的骨架坐标,应用改进的DTW算法进行了交叉验证,结果如图6所示.

表1　实验数据分组

AS1	AS2	AS3
挥手(高处)	挥手(高处)	高掷
挥手(胸前)	用手接物	前向踢腿
前向击拳	画叉	侧踢
高掷	画对号	慢跑
鼓掌	画圆	打网球
弯腰	双手挥舞	过网发球
过网发球	前向踢腿	打高尔夫
捡 & 扔	侧击拳	捡 & 扔

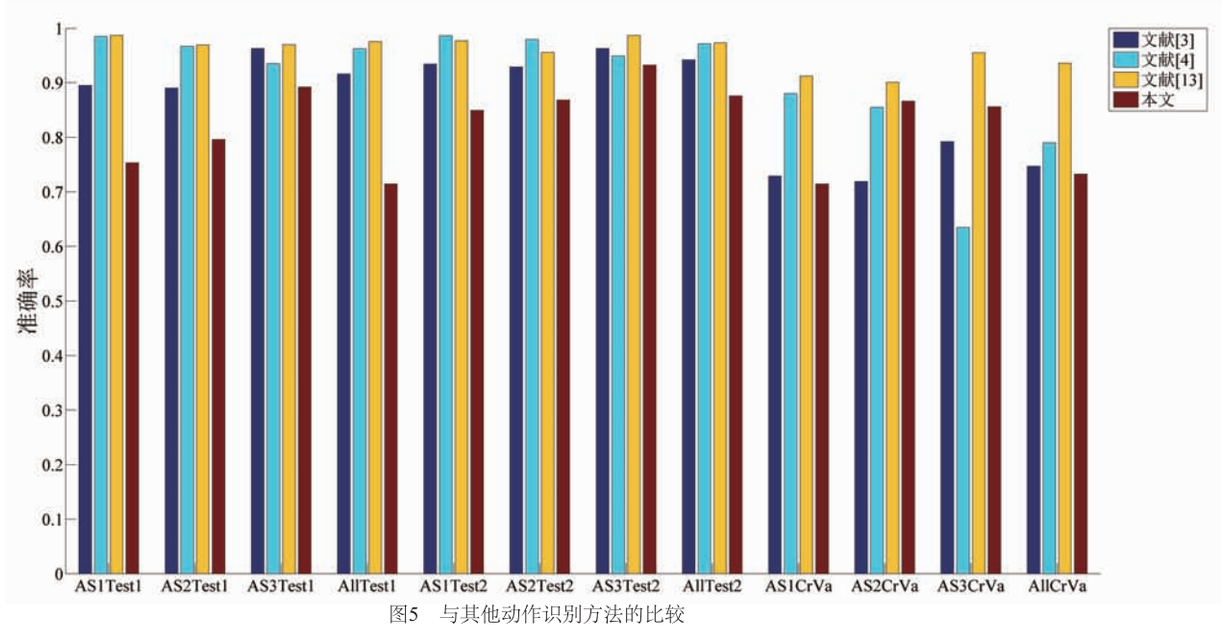


图5　与其他动作识别方法的比较

图6(*a*)和6(*b*)分别表示了以原始坐标和简化的立体骨架模型为输入数据所得识别结果的混淆矩阵,由该图可以看出,应用简化的立体骨架模型提高了识别准确率,同时减小了相似行为的误判率.

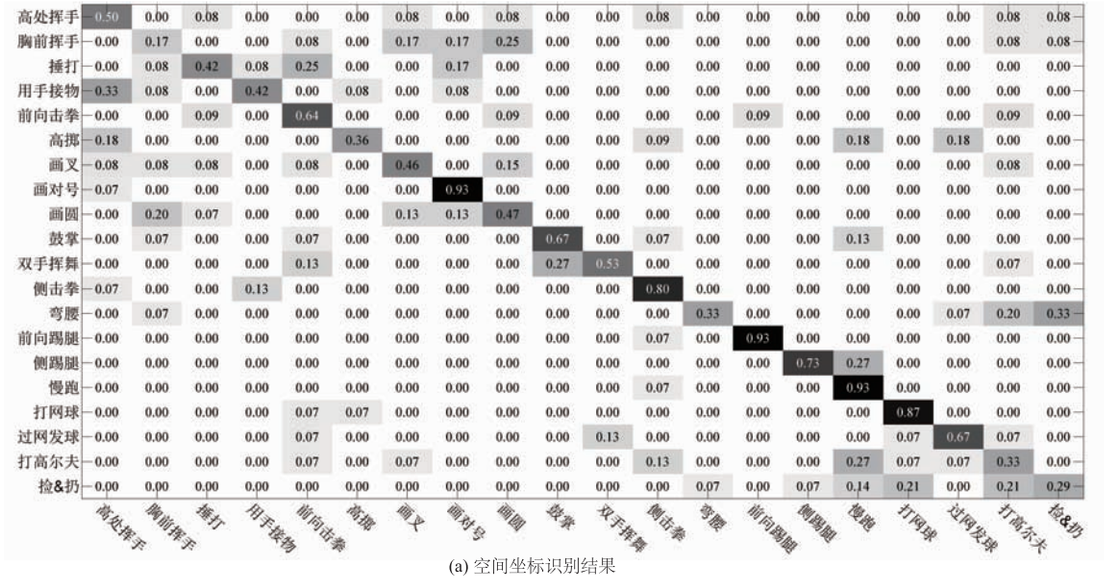
3.3　与其他DTW算法的比较

本文的第二个创新点在于对传统的DTW算上法进行了改进,目的在于在提高动作识别率的同时提高识别效率. 本文采用传统的和改进后的DTW以及其他文献中所提出的对DTW的改进算法分别对行为数据进行匹配与识别,图7展示了对比结果,分别从识别率和两个动作序列的平均匹配时间进行对比. 本次实验采用了交叉验证.

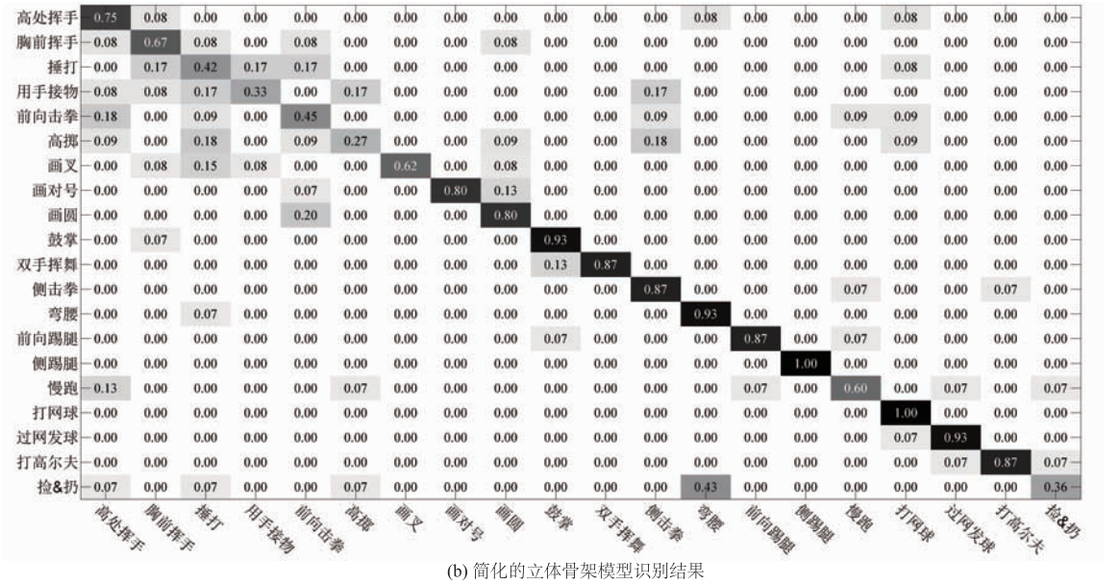
由图7可以看出与其他改进的DTW算法以及传统DTW算法相比,本文对DTW算法的改进在降低计算量,提高动作序列的匹配效率的同时,也提高了动作识别准确率.

4　结论

本文提出了一种基于3D人体骨架的动作识别方法. 简化的立体骨架模型减小了因人体骨架大小、形状不同所产生的匹配偏差. 为改进传统DTW算法计算量大的弊端,设计了“一次规划,二次细化”的匹配算法,在提高匹配和识别精度的同时提高了识别效率. 本文所使用的每个数据样本中仅包含了一种动作,因此下一步的工作重点是进一步完善算法,实现单个序列中多种动作的识别.



(a) 空间坐标识别结果



(b) 简化的立体骨架模型识别结果

图6 与原始坐标输入的比较

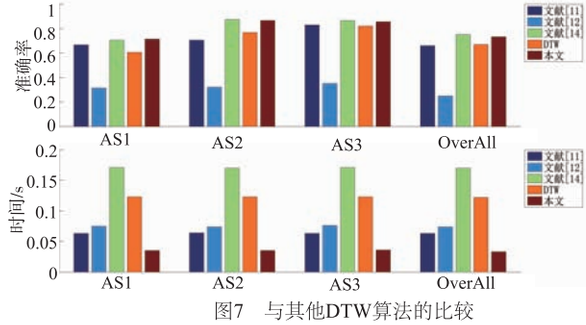


图7 与其他DTW算法的比较

参考文献

[1] Wang J, Liu Z, Wu Y. Mining actionlet ensemble for action recognition with depth cameras[A]. IEEE Conference on Computer Vision and Pattern Recognition[C]. Providence, Rhode Island, USA: IEEE, 2012. 1290 – 1297.

[2] Yu G, Liu Z, Yuan J. Discriminative orderlet mining for real-time recognition of human-object interaction[A]. Asian Conference on Computer Vision[C]. Singapore: Springer, 2014. 50 – 65.

[3] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3d points[A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops[C]. San Francisco, California, USA: IEEE, 2010. 9 – 14.

[4] Xia L, Chen C C, Aggarwal J K. View invariant human action recognition using histograms of 3D joints[A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops[C]. Providence, Rhode Island, USA: IEEE, 2012. 20 – 27.

[5] Johansson G. Visual motion perception[J]. Scientific American, 1975: 76 – 88.

[6] Ziaeefard M, Ebrahimnezhad H. Hierarchical human action recognition by normalized-polar histogram[A]. International Conference on Pattern Recognition[C]. Istanbul, Turkey: IEEE 2010. 3720 – 3723.

[7] Itakura F. Minimum prediction residual principle applied to speech recognition[J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1975, 23(1): 67 – 72.

[8] Morales-Cordovilla J A, Cabanas-Molero P, Peinado A M, et al. A robust pitch extractor based on DTW lines and CASA with application in noisy speech recognition[A]. Advances in Speech and Language Technologies for Iberian Languages[M]. Heidelberg, Berlin: Springer, 2012. 197 – 206.

[9] Ikizler N, Duygulu P. Histogram of oriented rectangles: A new pose descriptor for human action recognition[J]. Image and Vision Computing, 2009, 27(10): 1515 – 1526.

[10] Candan K S, Rossini R, Wang X, et al. sDTW: computing DTW distances using locally relevant constraints based on salient feature alignments[J]. Proceedings of the VLDB Endowment, 2012, 5(11): 1519 – 1530.

[11] Zhang W, Zhang Y, Gao C, et al. Action recognition by joint spatial-temporal motion feature[J]. Journal of Applied Mathematics, 2013, 2013(1): 467 – 471.

[12] Ruan X, Tian C. Dynamic gesture recognition based on improved DTW algorithm[A]. IEEE International Conference on Mechatronics and Automation[C]. Beijing, China: IEEE, 2015. 2134 – 2138.

[13] Theodorakopoulos I, Kastaniotis D, Economou G, et al. Pose-based human action recognition via sparse representation in dissimilarity space[J]. Journal of Visual Communication and Image Representation, 2014, 25(1): 12 – 23.

[14] 徐贵力,赵妍,姜斌,等. 基于局部尺度特征描述和改进 DTW 技术的局部轮廓匹配算法[J]. 电子学报, 2016, 44(1): 135 – 142.
Xu GuiLi, Zhao Yan, Jiang Bin, et al. Partial contour matching algorithm based on local scale description and improved DTW[J]. Acta Electronica Sinica, 2016, 44(1): 135 – 142. (in Chinese)

作者简介

张友梅　女,1990 年 3 月出生,山东济南人. 2013 年毕业于山东大学控制科学与工程学院,同年入读本校硕士研究生. 现为硕博连读生,从事异常行为分析方面的有关研究.
E-mail: zhanghyoumei@163.com

常发亮　男,1965 年 10 月出生,山东潍坊人,博士,教授,博士生导师,分别于 1989 年和 2005 年获得山东大学硕士和博士学位. 主要研究方向为:模式识别、机器视觉、智能系统控制等.
E-mail: flchang@sdu.edu.cn