ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.9, September 2008, pp.2285−2292
DOI: 10.3724/SP.J.1001.2008.02285

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于 Segmental-DTW 的无监督行为序列分割*

吴晓婕+, 胡占义, 吴毅红

(中国科学院 自动化研究所 模式识别国家重点实验室,北京 100190)

Unsupervised Behavior Sequence Segmentation Based on Segmental-DTW

WU Xiao-Jie+, HU Zhan-Yi, WU Yi-Hong

(National Laboratory of Pattern Recognition, Institute of Automation, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: xjwu@nlpr.ia.ac.cn

Wu XJ, Hu ZY, Wu YH. Unsupervised behavior sequence segmentation based on segmental-DTW. *Journal of Software*, 2008,19(9):2285−2292. http://www.jos.org.cn/1000-9825/19/2285.htm

Abstract: Behavior sequence segmentation is the first and most fundamental step of behavior analysis and recognition. In this paper, a novel unsupervised algorithm for behavior sequence segmentation is proposed. The algorithm consists of the following steps: (1) The video sequence is coarsely segmented into equal length subsequences with overlapping time window; (2) Segmental-DTW is used to find out matching behavior clips between pairs of video subsequences; (3) The similarity between behavior clips is represented by an adjacency graph, and an efficient graph clustering algorithm is used to generate behavior clusters. The algorithm, based on a coarse-to-fine strategy, is able to satisfactorily segment behavior sequences and cluster typical behavior patterns. The segmentation results can be used for further behavior modeling and recognition. Experimental results show the behavior clips segmented by this algorithm are prototypical and meaningful.

Key words: behavior sequence segmentation; unsupervised method; Segmental-DTW; graph clustering

摘 要: 行为序列分割是行为分析与识别中最初始、最基础的一个步骤.提出了一种无监督的行为序列分割算法,主要步骤包括:(1) 采用等长有重叠的时间窗口对视频序列进行粗分割;(2) 将粗分割的视频段两两作比较,通过Segmental-DTW 算法分割出两个视频段中最相似的行为片断;(3) 将行为片断的相似性转化为邻接图表示,通过图聚类方法对分割出的行为片断进行聚类.该算法采用了从粗到细的分割思想,能够准确地分割出视频序列中大量出现的行为的片断,并将相同行为的片断聚为一类.分割结果可以直接用于行为建模和识别.实验结果也表明了分割出的行为片断具有较好的代表性和有效性.

关键词: 行为序列分割;无监督方法;Segmental-DTW;图聚类

中图法分类号: TP181 **文献标识码**: A

人的行为分析与识别在视觉监控、视频检索、医疗诊断、运动视频分析以及人机交互等方面具有广泛的应用前景,是当前计算机视觉领域的一个研究热点[1].行为序列分割是行为分析与识别中最初始、最基础的一个步骤.能否准确地将一个有意义的行为序列从整个视频序列中分割出来,直接影响到后续的行为建模和识别.传

* Supported by the National Natural Science Foundation of China under Grant Nos.60633070, 60475009 (国家自然科学基金)
Received 2007-01-15; Accepted 2007-04-25

统的基于有监督的行为分析方法[2−5]依靠人来分割和标注行为序列,这项工作非常繁琐、耗时,且不够鲁棒[6].因此,人们又提出了基于无监督的行为分析方法[6−9].这类方法不需要手工分割和标注,能够自动或半自动地建立行为模型,具有较强的适用性.在无监督的分析方法中,根据不同的应用场合和视频数据,大体有以下两种不同的分割方法[6]:(1) 根据视频序列的间断点、突变点进行分割;(2) 按照等长、有重叠的时间窗口进行分割.如果视频序列中含有明显的、易检测的间断点或突变点,那么,按照第 1 种分割方法就可以较准确地将每个行为从视频序列中分割出来[6,7];然而在很多应用场合,从所获得的视频数据中往往无法准确检测出间断点、突变点,此时,大部分的无监督方法采用了上述第 2 种方法来分割视频数据[8,9].Zelnik-Manor 和 Irani[8]通过等长有重叠的时间窗口分割视频序列,对每个视频段提取多时间尺度特征,形成相似性矩阵,最后通过谱聚类的方法自动地建立了行为事件的模型.李和平等人[9]采用同样的方法分割视频序列,然后提取每个视频段的时空特征,通过动态时间归整(dynamic time warping,简称 DTW)度量每两个视频段之间的距离,形成相似性矩阵,最后通过谱聚类建立正常行为的模型.上述采用等长、有重叠时间窗口分割视频数据的方法虽然简单、易行,但也存在以下两个主要问题:

(1) 没有考虑到视频序列中行为的分布情况,行为序列分割得不够准确,必然会造成所建行为模型的不准确[8];

(2) 鉴于实际获得的视频数据,不同的行为在持续时间上可能会有比较大的差异,而不同的人完成的相同行为也会有比较显著的快慢差异,因此,为了不漏掉长序列行为,分割视频序列的时间窗口的长度应能够保证将平均持续时间最长的行为比较完整地分割出来.此时,采用等长有重叠时间窗口分割出来的视频段中可能只有一种行为(持续时间较长的行为),也可能同时包含有不同的几种行为(持续时间较短的行为).在这种情况下,文献[8,9]提出的简单度量两个视频段之间整体距离的方法,无论是在特征空间上的直接度量还是基于 DTW 的度量都将不再适用.

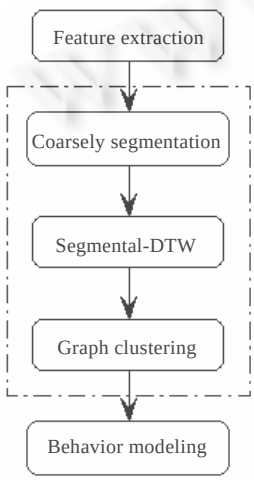


Fig.1 The flowchart of the proposed method

图 1 本文算法流程图

针对上述问题,我们提出了一种新的无监督的行为序列分割方法.本文方法首先用等长、有重叠的时间窗口对视频序列进行粗略的分割,然后采用 Segmental-DTW 算法分割出每两个视频段中最相似行为的片断,最后将行为片断的相似性转化为邻接图表示,通过图聚类方法对分割出的行为片断进行聚类.本文提出的行为序列分割方法的特点是:采用了从粗到细的分割思想,可以无监督地分割出视频序列中大量出现的行为的片断,并将相同行为的片断聚为一类.本文方法有效地解决了等长有重叠时间窗口分割视频数据时存在的两个问题.本文的重点是行为序列的分割,但为了验证行为序列的分割效果,在实验部分,我们对聚类后的行为片断用 HMM 进行了行为建模和识别.实验结果表明了本文方法的有效性.

图 1 给出了本文算法的流程图,虚线框内是分割算法的主要步骤.后面的几节将对算法的各个步骤进行详细介绍.

1 特征提取

本文中,对每帧图像提取简单的时空特征,该特征同时刻画了运动目标的形状信息以及运动目标在微小时间段上的运动信息.具体方法如下:

用背景减除的方法获得运动信息的二值化图像,如图 2(b)所示.可以假设第 t 帧图像中的运动目标与它的前 Δt 帧图像和后 Δt 帧图像中的运动目标在空间尺度上的变化很小(当 Δt 很小时),此时,这种变化可以不考虑.因此,我们将第 t 帧图像中的运动目标与它的前 Δt 帧图像和后 Δt 帧图像中的运动目标作区域对齐.区域对齐的方法是:截取出前后两帧图像中运动目标的外接矩形区域,将第 1 个矩形区域固定,将第 2 个矩形区域的质心与第 1

个重合,然后在质心周围一个较小的范围内(本文取 5×5 像素大小)移动后一帧图像矩形区域的质心位置,搜索使得两个运动目标的相关性取到最大值的位置,此时,两帧图像的运动目标就对齐了.对多帧情况而言,其他帧用类似的步骤与第 1 帧对齐.这样,我们就可以得到运动目标的时空矩阵,记作 $ST_{x,y,z}, z\in[t-\Delta t, t+\Delta t]$,它将运动目标在空间上按上述方法对齐,然后按时间顺序排列起来.本文中取 $\Delta t=1$,也就是将连续的 3 帧图像中的运动目标空间对齐,并按时间排列起来形成时空矩阵.图 2(c)是将时空矩阵 $ST_{x,y,z}$ 沿时间轴叠加到一幅图像上显示的结果.可以看出,区域对齐后,运动目标前后帧之间的差异反映了由行为本身引起的变化.将时空矩阵 $ST_{x,y,z}$ 在空间上划分成 $m\times n$ 个小立方块,如图 2(d)所示.在本文中取 $m=9, n=5$.统计每个小立方块 $ST^i_{x,y,z}, i\in[1,45]$ 内时空矩阵每前后两帧变化的总和,记为 μ^i_t:

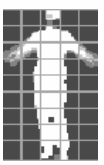
$$\mu^i_t=\sum_{x,y}\sum_{z=t-\Delta t+1}^{t+\Delta t}|I_z(x,y)-I_{z-1}(x,y)|\forall x,y\in ST^i_{x,y,z} \tag{1}$$

并用立方块的体积将之归一化,记为 $d^i_t, i\in[1,45]$.将上述统计量按行排列就得到了第 t 帧图像的特征向量:$D_t=[d^1_t,d^2_t,...,d^J_t], J=45$.该特征刻画了运动目标的形状以及在微小时间段内形状的变化信息.

(a) An original image
(a) 原始图像

(b) The binary image of motion information
(b) 运动信息二值图像

(c) The cumulated spatial-temporal matrix
(c) 时空矩阵叠加显示结果



(d) The extraction of spatial-temporal features
(d) 时空运动特征提取示意图

Fig.2 Examples of feature extraction
图 2 特征提取示例

2 无监督行为序列分割算法

准确、有效地将行为序列从视频序列中分割出来,是行为分析中最初始也是非常关键的一个步骤,因为行为序列分割结果的好坏将直接影响到后续的行为建模和识别.本文采用了从粗到细的分割思想:首先用等长、有重叠的时间窗口对视频序列进行粗分割,然后将分割出的视频段两两作比较,通过 Segmental-DTW 算法将每两个视频段中相同的行为片断分割出来.值得指出的是,Segmental-DTW 不是简单地整体比较两个视频段,而是通过比较两个视频段的局部片断,进而得到对两个视频段中行为片断之间相似性的评价.所以,当视频数据中的行为在持续时间上有较大差异时,引言部分中指出的采用等长、有重叠时间窗口分割视频数据时存在的问题,在这里可以得到有效的解决.

2.1 视频序列粗分割

给定一个连续的视频序列 V,我们不知道其中包含有哪些行为,也无法通过检测间断点或突变点确定其中每个行为的确切起始和终止位置,在这种情况下,首先用等长、有重叠的时间窗口将 V 粗分割为 N 段,即 $V=\{v_1, v_2,\ldots,v_n,\ldots,v_N\}$.对每帧图像提取一个特征向量 $D_t=[d^1_t,d^2_t,...,d^J_t]$,则每个视频段可以表示为 $v_n=\{D_{n1},D_{n2},\ldots,D_{nM}\}$.时间窗口的长度,也即分割出的每个视频段的长度为 M.

2.2 基于Segmental-DTW的细分割

当视频数据中行为与行为之间在持续时间上差异较大时,上节粗分割后得到的视频段,有可能只有一种行为,也有可能同时包含几种不同的行为,需要进一步将视频段中同一行为的片断分割出来.本文采用了 DTW 的一种变形算法——Segmental-DTW[10].

假定两个视频段提取出的特征向量序列分别是 $\{u_i\}_{i=1}^{M}$,$\{v_j\}_{j=1}^{N}$,利用 DTW 通常的作法[11]是首先计算距离矩阵 D:

$$D(i,j)=\|u_i-v_j\| \tag{2}$$

然后,通过动态规划算法以 $D(1,1)$为起点、$D(M,N)$为终点在距离矩阵上搜寻一条最佳的归整路径,使得累加距离达到最小值.DTW 得到的最小累加距离可以用来度量两个视频段之间的整体相似性.

然而,现在我们面临的问题是,每个视频段中可能包含不止一种行为,DTW 必须以 $D(1,1)$为起点、$D(M,N)$为终点的首尾对齐性质只能得到两个视频段的整体最小距离和整体最佳归整路径.这样的整体信息并不能反映出要作比较的两个视频段中是否含有相同的行为片断,以及这样的行为片断在视频段中的具体位置信息.我们希望获得反映两个视频段中相同的行为片断的局部归整路径和局部距离.

为了解决上述问题,本文采用了 Segmental-DTW.这个算法最初被用于找出两段语音信号中是否含有相同的单词[10],在这里,我们用它来分割出两个视频段中最相似行为的片断.该算法的具体过程为:将距离矩阵 D 沿对角线方向划分成 l 个宽为 W 相互重叠的带状区域 $b_1,b_2,\ldots,b_l$.在每个带状区域内通过 DTW 搜寻一条最佳路径,这样我们就得到了可能的路径 $p_1,p_2,\ldots,p_l$.然后,在这 l 条可能的路径上搜寻所有长度不小于 L 的子路径,其中,平均距离 A_p 最小的那条子路径记做 $p_{\min}$,最小平均距离记作 $A_{\min p}$.子路径的平均距离 A_p 由子路径的累加距离 W_p 和子路径的长度 G_p 确定:

$$A_p=W_p/G_p \tag{3}$$

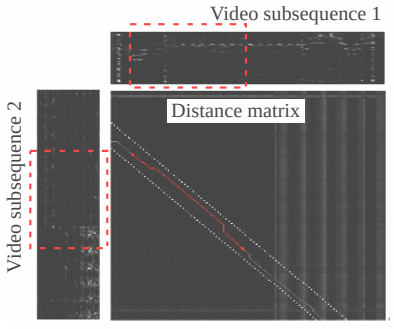


Fig.3 The segmental-DTW algorithm

图 3 Segmental-DTW 方法示意图

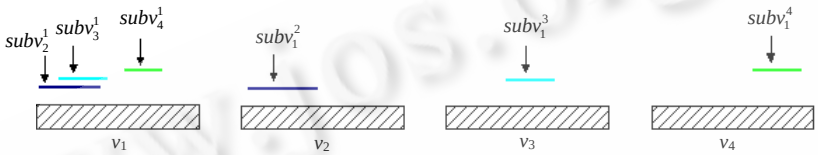
$p_{\min}$ 在距离矩阵 D 中的位置反映了相比较的两个视频段中最相似的行为片断在相应视频段中的位置,最小平均距离 $A_{\min p}$ 是两个行为片断之间的距离,算法如图 3 所示.图中 Distance Matrix 中的线段就是通过 Segmental-DTW 算法找到的平均距离最小的子路径 $p_{\min}$,根据 $p_{\min}$ 分别在 video subsequence 1 和 video subsequence 2 中分割出的一组最相似的行为片断,图中用虚线框表示.要求子路径的长度不小于 L,可以防止找到的行为片断过短而导致错误匹配.实验中,取 $L=\sqrt{2}M/5$ (粗分割后每个视频段的长度为 M).我们用下面最小平均距离 $A_{\min p}$ 的函数来评价两个行为片断间的相似性:

$$s=\exp[-A_{\min p}/\sigma] \tag{4}$$

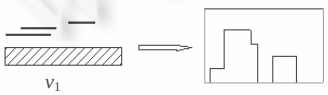
σ为常数因子.

将 $V=\{v_1,v_2,\ldots,v_N\}$中的视频段两两比较,通过 Segmental-DTW 找出每两个视频段中最相似的一组行为片断.这样,对每个视频段 $v_i,i\in[1,N]$可以与其他 $N-1$ 个视频段作比较,在 v_i 中分割出 $N-1$ 个行为片断,记做 $\{subv_1^i,\ldots,subv_{i-1}^i,subv_{i+1}^i,\ldots,subv_N^i\}$.$v_i$ 与 $v_j,j\neq i$ 作比较,在 v_i 和 v_j 中分别分割出的一组最相似的行为片断可记做 $[subv_j^i,subv_i^j]$.视频段 v_i 上找到的这 $N-1$ 个行为片断,在度上一般并不相同,在位置上有的相互重叠,有的相隔较远;有的同属于视频段中的一种行为,有的同属于视频段中的另一种行为(同属于一种行为的行为片断在位置上可能是相互重叠的,属于视频段中不同行为的行为片断在位置上可能就相隔较远),如图 4(a)所示.该图示意了视频段 v_1 与视频段 v_2,v_3,v_4 作比较,在 v_1 上找到的 3 个行为片断 $\{subv_2^1,subv_3^1,subv_4^1\}$ 的分布情况.图中在相比较的两个视频段上找到一组最相似的行为片断.比如,$[subv_2^1,subv_1^2]$ 是视频段 v_1 与视频段 v_2 作比较分割出的一组行为片断.通过观察发现,视频段中行为片断密集地在某个位置附近出现表明在该位置处某种行为的存在.因此,对视频段 v_i,假定 $\{subv_1^i,\ldots,subv_{i-1}^i,subv_{i+1}^i,\ldots,subv_N^i\}$ 是我们找到的行为片断,每个片断赋予视频段 v_i 与视频段 $v_j(j\neq i)$

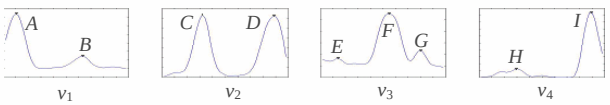
作比较,根据式(4)得到的相似性值,将这些值在时间轴上投影累加,可以得到一条相似性曲线.图 4(b)示意了图 4(a)中在 v_1 上找到的 3 个行为片断按上述规则累加得到的相似性曲线.将视频段 v_i 上找到的 N–1 个行为片断按上述规则累加起来得到相似性曲线,经平滑后求出该曲线上的峰值点,如图 4(c)所示.对每个视频段用上述方法取到若干峰值点,以这些峰值点为节点,构造邻接图(adjacency graph).两个视频段上行为片断的匹配意味着行为片断对应的两个节点之间存在一条加权边,权重为两个行为片断的相似性 s,邻接图的构造如图 5 所示.该图示意了图 4(a)中 v_1 与 v_2,v_3,v_4 作比较分割出的 3 组行为片断的邻接图构造方法.[$subv_2^1$,$subv_1^2$]是 v_1 与 v_2 作比较分割出的一组行为片断,其中,$subv_2^1$ 在 v_1 中的位置距 v_1 形成的相似性曲线中找到的峰值点 A 最近(如图 4(c)所示),$subv_1^2$ 在 v_2 中的位置距 v_2 中形成的相似性曲线中找到的峰值点 C 最近.所以,邻接图中节点 A 和 C 之间有一条加权边,它表明 $subv_2^1$ 和 $subv_1^2$ 是一组匹配的行为片断.该加权边的权重是 $subv_2^1$ 和 $subv_1^2$ 的相似性 s.按照同样的方法,图 5 所示的邻接图用 5 个节点、3 条加边示意了图 4(a)中 v_1 分别与 v_2,v_3,v_4 作比较找到的 3 组行为片断的邻接图构造.



(a) Behavior clips segmented in video subsequences
(a) 视频段中行为片断分布示意



(b) Production of similarity curve
(b) 相似性曲线生成方法示意



(c) Nodes extracted from similarity curves
(c) 相似性曲线、图节点示意

Fig.4
图 4

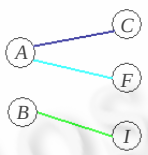


Fig.5 Production of the adjacency graph
图 5 邻接图的构造示意

2.3 图聚类

上节中得到的邻接图,每个节点对应于 Segmental-DTW 找到的行为片断,节点间的加权边对应于一组行为片断之间的相似性,可以通过图聚类的方法将相同行为的片断聚为一类,并获得视频序列中大量出现的行为的类别.本文中采用了 Newman[12,13]提出的图聚类方法.这是一种自底向上聚类的方法,它较文献[7–9]中普遍采用的谱聚类方法的优势在于能够自动地决定聚类的终止条件,且计算量低[13].通过定义一个描述聚类结果好坏(modularity)的 Q 值,当 Q 值取最大值时,理论上讲就达到了最优的聚类结果.具体步骤如下:

(1) 归一化邻接图,假定连接节点 i,j 之间的加权边权重为 e_{ij},则对于上节得到的无向邻接图赋值 e_{ij}=e_{ji}.归一化邻接图,使得 $\sum_{i,j} e_{ij} = 1$.然后定义[12]:

$$a_i = \sum_j e_{ij},\ b_j = \sum_i e_{ij} \tag{5}$$

对无向图有 $a_i=b_i$.

(2) 定义描述聚类结果好坏(modularity)的 Q 值[13]:

$$Q=\sum_i(e_{ii}-a_i^2) \tag{6}$$

Q 值落在[0,1]之间.聚类前每一个节点就是一类,此时,Q 值很低,接近于 0.

(3) 将节点 i,j 聚为一类引起 Q 值的变化为

$$\Delta Q=e_{ij}+e_{ji}-2a_ia_j=2(e_{ij}-a_ia_j) \tag{7}$$

将使 Q 值增长最快的两个节点聚为一类,然后更新邻接图,回到第 1 步,直到 Q 值不再增长,则聚类停止.实验发现,最大的 Q 值一般落在[0.3,0.7]之间,更高的 Q 值比较罕见[13].

3 实验结果

3.1 实验数据

实验中,我们从 Schuldt 等人[14]的数据库中获取了如图 6 所示的 4 种行为(“拳击”2 651 帧、“挥手”2 585 帧、“鼓掌”2 663 帧、“行走”2 624 帧)进行实验.Schuldt 等人的数据库中,每个单独的视频序列仅包含一个人连续的一种行为,所以,在实验中使用的视频序列是这 4 种行为的时间序列合成的数据.具体合成过程为:首先对每种行为的时间序列进行无重叠的分割,每次取大于 25、小于 45 帧的随机整数作为每次分割的片断的长度值,然后对所有的行为片断进行随机排序,获得混合行为的时间序列.将前 6 523 帧取出作为训练集,后面的 4 000 帧作为测试集.在训练集上,首先采用等长、有重叠的时间窗口粗分割视频序列.本文实验中,采用了宽为 50 帧、步长为 15 帧的滑动窗口.

Fig.6 The images of four behaviors' samples from the dataset of Schuldt, *et al*[14]

图 6 Schuldt 等人[14]的数据库中 4 种行为的样本图像

考虑到实际中获得的视频数据行为与行为之间,无论是不同的行为之间还是不同的人完成的相同行为之间,在持续时间上可能有较大的差异,所以在合成数据时,我们取每个行为的长度为大于 25、小于 45 帧的随机整数.视频粗分割中滑动窗口的长度取为 50 帧,它保证了粗分割后视频段中至少包含有一个完整的行为,同时也因为行为与行为之间在持续时间上的差异比较大,所以,粗分割的视频段中也可能同时包含有两个完整的行为,需要进一步细分割.

3.2 图聚类结果

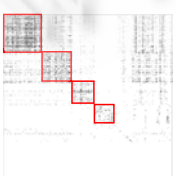


Fig.7 The affinity matrix ordered according to the result of clustering

图 7 根据图聚类结果将节点重排后的亲和力矩阵

图 7 是邻接图聚类后将节点重排的结果.可以看出,训练集被自动聚成了 4 大类,分别对应于实验数据中的 4 种行为.训练集视频数据总共有 6 523 帧,通过本文算法分割并经图聚类后得到的行为片断总长度为 4 612 帧,其中,类别 1“鼓掌”分割出 83 个行为片断(1 500 帧),类别 2“挥手”分割出 78 个行为片断(1 030 帧),类别 3“拳击”分割出 44 个行为片断(838 帧),类别 4“行走”分割出 68 个行为片断(828 帧).本文的特点是不仅能够无监督地分割视频序列,同时还能将视频序列中大量出现的行为聚为一类.如果通过 Segmental-DTW 分割出的行为片断不够准确,则行为片断间的相似性就会比较低,在邻接图中加权边的权重就会

比较小,图聚类过程中就不容易被聚类到大量出现的行为类中;而如果分割出的行为片断比较准确,则行为片断间的相似性就比较高,在邻接图中加权边的权重就会比较大,图聚类过程中就比较容易被聚类到相应的大量出现的行为类中.因此,最终图聚类的正确率可以用来评价分割结果的好坏.表 1 中统计了聚类结果的正确率.

表 1 给出的聚类结果正确率是评价序列分割结果好坏的一项重要指标.然而,分割出的行为片断将用于后续的行为建模,直接影响行为识别的结果.因此,我们希望本文算法不仅要有较高的分割和聚类正确率,还希望分割出的行为片断确实能够代表某类行为.也就是说,希望分割出的行为片断具有代表性和有效性.为了进一步验证序列分割的结果,下面我们对聚类后的行为片断用 HMM 建模,然后进行行为识别,与采用手工标注的有监督方法达到的识别率作了比较.

Table 1 The clustering results

表 1 聚类结果的正确率

	Handclapping	Handwaving	Boxing	Walking
Cluster 1 (handclapping)	**0.787**	0.211	0	0.002
Cluster 2 (handwaving)	0.103	**0.771**	0.126	0
Cluster 3 (boxing)	0.054	0.039	**0.907**	0
Cluster 4 (walking)	0	0	0.006	**0.994**

3.3 建立行为模型与识别

由于图聚类后每类抽取出的行为片断个数比较少,因此首先将抽取出的行为片断全体作为一个样本集,建立一个如图 8 所示的隐马尔可夫模型(HMM).HMM 包含 3 个隐藏节点,并采用高斯混合模型(GMM)拟合每个隐藏节点的输出概率密度函数.本文中,GMM 的成份个数取为 3.

以上述 HMM 参数作为先验知识,每一类中抽取出的行为片断作为新的样本,通过最大后验自适应(MAP adaptation)得到新的 HMM[15],这样就建立起 4 种行为的 HMM.

将后面的 4 000 帧作为测试集用于识别.对观察到的时间序列 O,用下式判断它的类别

$$\hat{c} = \arg\max_{c}\{P(O \mid B_c)\} \tag{8}$$

其中,$c \in [1,4]$,B_c 为第 c 种行为的 HMM 参数.

为了进行比较,我们对上述训练集进行了手工标注,然后对每一类行为建立起如图 8 所示的 HMM,同样,采用式(8)进行识别.分别采用上述两种方法得到的识别率见表 2.从表 2 可以看出,直接在本文序列分割结果上训练 HMM 达到的识别率只比手工标注的有监督方法获得的识别率略有降低.表 2 的结果说明了本文方法分割出的行为片断能够代表相应的行为,具有代表性和有效性.

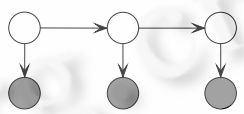


Fig.8 An example of hidden Markov model

图 8 HMM 模型示意图

Table 2 Comparison of our unsupervised method with the supervised method

表 2 本文采用的无监督方法和采用手工标注的有监督方法在识别率上的比较

	Average recognition rate	"Boxing"	"Handclapping"	"Handwaving"	"Walking"
Unsupervised method	0.91	0.86	0.82	0.96	1.00
Supervised method	0.945	0.98	0.90	0.90	1.00

4 结 论

无监督的行为分析是近年来被广泛关注的研究热点,行为序列的分割是其中最初始也是非常关键的一个步骤.本文提出的无监督行为序列分割方法由于使用了 Segmental-DTW,可以无监督地分割出视频序列中大量出现的行为的片断,这样就可以放宽对粗分割时间窗口长度选择的限制.而对大多数行为序列分割方法而言,选择长度合适的时间窗口是一个关键问题.另外,本文将行为片断的相似性转化为邻接图表示,然后利用图切割的全局优化性质进行分割,可以容忍局部错误或不匹配对整体分割的影响.

本文方法最大的不足是 Segmental-DTW 算法在每个带状区域内都要使用 DTW 搜寻最佳路径,运算量比较

大.如何提高计算效率是我们下一步要进行的工作.

References:

[1] Hu WM, Tan TN, Wang L, Maybank S. A survey on visual surveillance of object motion and behaviors. IEEE Trans. on Systems, Man, and Cybernetics—Part C, 2004,34(3):334−352.

[2] Haritaoglu I, Harwood D, Davis L. W4: Real-Time surveillance of people and their activities. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2000,22(8):809−830.

[3] Bobick A, Davis J. The recognition of human movement using temporal templates. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(3):257−267.

[4] Gong S, Xiang T. Recognition of group activities using dynamic probabilistic networks. In: Werner B, ed. Proc. of the IEEE Int'l Conf. on Computer Vision. Washington: IEEE Computer Society, 2003. 742−749.

[5] Laptev I, Linderberg T. Space-Time interest points. In: Werner B, ed. Proc. of the IEEE Int'l Conf. on Computer Vision. Washington: IEEE Computer Society, 2003. 432−439.

[6] Xiang T, Gong SG. Video behaviour profiling and abnormality detection without manual labeling. In: Werner B, ed. Proc. of the IEEE Int'l Conf. on Computer Vision. Beijing: IEEE Computer Society, 2005. 1238−1245.

[7] Rui Y, Anandan P. Segmenting visual actions based on spatio-temporal motion patterns. In: Jacobs A, Baldwin T, eds. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society Press, 2000. 111−118.

[8] Zelnik-Manor L, Irani M. Event-Based analysis of video. In: Jacobs A, Baldwin T, eds. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society, 2001. 123−130.

[9] Li H, Hu Z, Wu Y, Wu F. Behavior modeling and recognition based on space-time image features. In: Werner B, ed. Proc. of the IEEE Int'l Conf. on Pattern Recognition, Vol.1. Hong Kong: IEEE Computer Society, 2006. 243−246.

[10] Park A, Glass J. Unsupervised word acquisition from speech using pattern discovery. In: Jacobs A, Baldwin T, eds. Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Toulouse: IEEE Computer Society, 2006. 1−1.

[11] Chu S, Keogh EJ, Hart D, Pazzani MJ. Iterative deepening dynamic time warping for time series. In: Proc. of the 2nd SIAM Int'l Conf. on Data Mining. 2002. http://citeseer.ist.psu.edu/chu02iterative.html

[12] Newman MEJ. Mixing patterns in networks. Physical Review E, 2003,67:26−126.

[13] Newman MEJ. Fast algorithm for detecting community structure in networks. Physical Review E, 2004,69:66−133.

[14] Schuldt C, Laptev I, Caputo B. Recognizing human actions: A local SVM approach. In: Werner B, ed. Proc. of the IEEE Int'l Conf. on Pattern Recognition. Cambridge: IEEE Computer Society, 2004. 32−36.

[15] Li H, Hu Z, Wu Y, Wu F. Behavior modeling and abnormality detection based on semi-supervised learning method. Journal of Software, 2007,18(3):527−537 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/18/527.htm

附中文参考文献:

[15] 李和平,胡占义,吴毅红,吴福朝.基于半监督学习的行为建模与异常检测.软件学报,2007,18(3):527−537. http://www.jos.org.cn/1000-9825/18/527.htm

吴晓婕(1979－),女,北京人,博士生,主要研究领域为模式识别,计算机视觉,行为分析.

胡占义(1961－),男,博士,研究员,博士生导师,主要研究领域为摄像机标定,三维重建,主动视觉,机器人导航,基于图像的建模和绘制.

吴毅红(1973－),女,博士,副研究员,博士生导师,主要研究领域为多项式消元理论及应用,几何不变量的计算及应用,几何定理机器证明,摄像机标定,三维重建,视觉几何.