

视频监控中的行为序列分割与识别

钱惠敏 茅耀斌 王执铨 叶曙光

(南京理工大学自动化学院, 南京 210094)

摘要 智能监控系统中的行为分析与识别是当前计算机视觉领域的研究热点, 而行为序列分割则是行为分析与识别的基础。提出了一种无监督的行为序列分割算法, 并对分割结果进行识别。首先, 采用鲁棒的形状编码方案得到人体轮廓的紧凑表示, 提取轮廓点集特征描述运动人体; 然后, 基于奇异值分解(SVD)估计行为序列数据的本征维数, 确定数据对应的低维流形, 并通过检测特征数据在该流形上的投影误差的突变实现行为序列分割; 最后, 采用隐马尔可夫模型(HMM)对分割结果进行识别。在公共数据库上的实验结果表明了此分割和识别算法的有效性。

关键词 行为序列分割 行为识别 本征维数 奇异值分解 隐马尔可夫模型

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2009)11-2416-05

Time-sequential Activity Segmentation and Recognition for Video Surveillance

QIAN Hui-min, MAO Yao-bin, WANG Zhi-quan, YE Shu-guang

(School of Automation, Nanjing University of Science and Technology, Nanjing 210094)

Abstract Human motion analysis in an intelligence surveillance system is a hot research topic in computer vision, and temporal segmentation of human activity sequence is the most fundamental step in human motion analysis. In this paper, an unsupervised online temporal segmentation algorithm is presented, and then the segmentation result is recognized by HMM. Firstly, a robust shape encoding scheme is employed to produce a compact representation of human silhouette, and a new feature called contour point set is proposed. Secondly, the intrinsic dimensionality of feature sequence and the corresponding low-dimensional manifolds are determined using SVD, and the break of projecting error of activity sequence on the determinate manifolds is detected as the segmentation point of the activity sequence. Temporal segmentation results are recognized by HMM finally. Experiments on two public databases show the effectiveness of the segmentation and recognition algorithms in this paper.

Keywords temporal segmentation of activity sequence, activity recognition, intrinsic dimensionality, SVD, HMM

1 引言

近年来, 公共场所都普遍安装了视频监控系统, 智能视频监控已成为计算机视觉领域的一个研究热点。智能监控系统包括运动目标检测, 目标跟踪和行为理解, 目前在行为理解中行为序列分割方面的工作相对较少。行为序列分割是将连续的人体运动

序列分割为单个的有意义的行为模式, 它是后续行为识别与理解的基础。

现有的人体运动序列分割算法可分为有监督方法^[1-3]和无监督方法^[4-7]两类。有监督方法采用手动方法分割视频序列, 分割过程非常烦琐、耗时, 且由于手动分割受到人的主观影响, 分割结果往往不够准确。无监督方法则能自动或半自动地分割行为序列, 包括按照等长、有重叠的时间窗口进行分割^[4-5]和根

据行为模式的不连续性进行分割^[6-7]。前者比较简单,但准确率较低,其分割结果可能是一个不完整的行为模式或是多个行为模式的组合,这些误分割会导致后续的行为误识别。为提高分割准确率,吴晓婕等人提出了由粗到细的分割思想^[8],在采用等长、有重叠的时间窗口进行粗分割的基础上,再结合 Segmental-DTW (dynamic time warping) 和图聚类方法进行细分割,此方法运算量较大。另外,由于该算法基于时间窗口,实际应用中存在窗宽选择的困难。

根据行为模式的不连续性进行行为分割,是将行为序列分割问题转换为时间序列的不连续性检测问题。Yong 等人通过检测特征数据序列在时间上的不连续性分割人体运动^[6],然而,其分割是一种低层分割,分割边界对应行为的关键帧,分割出的片段不是一个完整的行为。本文的视频行为序列分割是一种高层分割,它将长视频序列分割为一个个仅包含一种完整行为模式的行为段,用于后续的行为识别,并最终实现智能监控。

高维数据的本征维数估计是目前高维数据处理领域的研究热点。本文依据行为序列的本征维数变化实现行为序列的分割,采用奇异值分解(SVD)方法估计本征维数。具体分割依据为包含单种行为模式的序列段的本征维数应小于包含多种行为模式的序列段的本征维数,在 3 维动画的运动捕获数据分割^[9]中已证明该思想的有效性。

本文首先采用鲁棒的形状编码方案得到人体二值图像的紧轮廓表示,提取特征数据;然后通过检测本征维数的变化实现基于 SVD 的行为序列分割,最后,采用隐马尔可夫模型(HMM)对分割结果进行识别。实验结果表明本文算法的有效性。

2 特征提取

形状特征是刻画人体行为的基本特征之一,现有分割算法中常使用人体骨架^[10-11]和人体外接矩形框的栅格表示^[8]特征。然而,人体骨架依赖于身体部分的完整检测,不够稳定;而人体外接矩形框的栅格表示则不能准确描述人体的细节特征,且矩形框内非人体部分的冗余信息引入了额外的计算量。本文采用鲁棒的形状编码方案^[12]得到人体轮廓的紧凑表示,并提出轮廓点集特征描述人体轮廓。

首先采用背景减除法获得运动人体的二值图像。假设二值图像中人体像素轮廓为 S ,现采用 $m \times n$ 的 2 维格子逼近轮廓 S ,编码过程如下:

- (1) 获取轮廓 S 的最小外接矩形框 $L(S)$;
- (2) 将 $L(S)$ 纵向等分为 n 个矩形,设每个矩形框内的人体像素集为 $S_j, j = 1, \dots, n$,并将每个矩形框收缩为 S_j 的最小外接矩形框 $L(S_j)$;
- (3) 将 $L(S_j)$ 横向等分为 m 个矩形,设每个矩形框内的人体像素集为 $S_{ij}, i = 1, \dots, m$,并将每个矩形框收缩为 S_{ij} 的最小外接矩形框 $L(S_{ij})$;若 S_{ij} 为空,则相应的 $L(S_{ij})$ 也为空。

图 1 给出了人体轮廓编码示意图,人体轮廓可由非空格子 $L(S_{ij})$ 的集合近似表示。现计算每个格子 $L(S_{ij})$ 的质心点,并采用所有质心点的集合 S_0 近似人体轮廓。随着 $m \times n$ 的增加, S_0 越来越逼近轮廓 S ,如图 2 所示,因此, S_0 能准确描述人体轮廓。



图 1 人体轮廓编码示意图

Fig. 1 The sketch map of encoding scheme

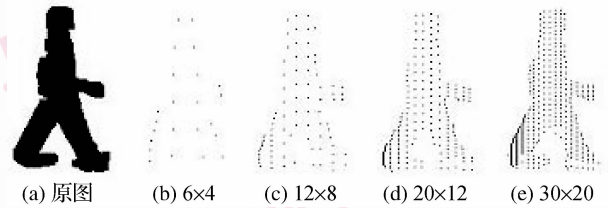


图 2 人体二值图像及不同 $m \times n$ 对应的轮廓点集 S_0

Fig. 2 Human blob and the corresponding S_0 under various $m \times n$

设第 k 帧图像的人体轮廓为 $S(k)$,质心坐标为 $(x_0(k), y_0(k))$,则第 k 帧图像的特征向量 $D_k = (P_{11x}(k), P_{11y}(k), \dots, P_{ijx}(k), P_{ijy}(k), \dots)$, $i = 1, \dots, m, j = 1, \dots, n$ 。

$$\begin{aligned} P_{ijx}(k) &= (\hat{P}_{ijx}(k) - x_0(k)) \\ P_{ijy}(k) &= (\hat{P}_{ijy}(k) - y_0(k)) \end{aligned} \quad (1)$$

式中, $(\hat{P}_{ijx}(k), \hat{P}_{ijy}(k))$ 是点集 $S_0(k)$ 中的点坐标。式(1)使得特征向量 D_k 对目标的平移具有不变性。若视频序列 V 由 K 帧图像构成,则其特征数据为 $D = (D_1, D_2, \dots, D_K)$ 。另外,本文实验在对人体轮廓高度进行规范化处理后再进行轮廓编码,这样既使得特征具有尺度不变性,也使得 $m \times n$ 较小时 S_0 能逼近 S ,从而降低特征向量维度,减小计算量。

3 行为序列分割

视频行为序列分割是指将连续的视频序列 V 分割为 n 个独立的行为段, 每个行为段中有且仅有一种有意义的行为模式。行为段之间的分割边界及行为段的数目 n 是该问题需要求解的变量。

3.1 基于 SVD 的本征维数估计

本征维数反映了高维数据的本征结构, 高维数据近似位于本征维数确定的低维流形上。基于 SVD 的本征维数估计将高维数据映射到一个低维空间中, 根据给定的阈值估计合适的本征维数。

设高维数据 $X \in \mathbf{R}^{N \times M}$ 由 N 个 M 维的样本数据构成, 对数据 X 进行奇异值分解:

$$X = U\Sigma V^T \quad (2)$$

式中, $\Sigma = \text{diag}(\Sigma_p, 0)$ 是广义对角矩阵, 且 $\Sigma_p = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > 0$, p 为矩阵 X 的秩。给定投影误差 $\tau < 1$, 本征维数估计为满足 $1 - E_r < \tau$ 的最小维数 r , 其中, $E_r = \sum_{i=1}^r \sigma_i^2 / \sum_{j=1}^p \sigma_j^2$ 。

假设仅包含一种行为模式的样本数据位于流形 Φ_1 上, 包含两种行为模式的样本数据位于流形 Φ_2 上, 则前者的本征维数 r_1 应小于后者的本征维数 r_2 。因此, 在分割行为序列时, 首先估计某一行为模式的本征维数 r , 固定 r , 若 X 在 r 维流形上的投影误差迅速增加, 则说明数据的本征维数改变, 相应地, 行为模式也发生了变化, 可在该处分割行为序列。

3.2 行为序列分割算法

本文的行为序列分割算法如下: 读取视频序列 V 的前 K 帧 (仅包含一种行为模式), 计算特征数据 D , 给定阈值 τ , 并采用 SVD 估计本征维数 r ; 计算前 n 帧的投影误差 $e_n = \sum_{j=r+1}^M \sigma_j^2$, 以及 $d_n = e_n - e_{n-l}$, l 为采样间隔, 用以消除噪声的影响; 假设属于同一行

为类型的样本数据对应的 d_n 服从高斯分布, 即 $d_n \sim N(\mu, \sigma)$, 则若 d_n 不在置信区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 内, 行为模式在第 n 帧发生变化, 因此, 在第 n 帧分割行为序列。分割算法 $TemporalSegment(D_n, \tau, K, l)$ 的伪代码如表 1 所示。本文参数选择为 $K = 80$, $l = 10$, $\tau = 0.99$ 。

表 1 分割算法 $TemporalSegment$ 的伪代码

Tab. 1 Pseudocode of algorithm "TemporalSegment"

```

初始化变量  $d_n = 0, \mu = 0, \sigma = 0.0001, n = 1$ ;
若  $n \leq K$ , 计算特征向量  $D_n$ , 更新  $D$ ;
估计  $D$  的本征维数  $r$ ;
当  $|d_n - \mu| \leq 3\sigma$ ,
    {  $n = n + 1$ ;
      计算特征向量  $D_n$ , 更新  $D$ ;
       $D = U\Sigma V^T$ , 并计算误差  $e_n$ ;
      若  $n \geq K + l$ , 计算  $d_n = e_n - e_{n-l}$ , 并更新  $\mu, \sigma$ ; }
在第  $n$  帧分割行为序列, 并重新初始化;
  
```

本文算法基于某一类行为数据的本征结构, 不受行为持续时间的影 响, 而采用等长、有重叠时间窗口分割视频序列的方法在行为持续时间变化时难以选择合适的窗宽。

4 实验结果

本文分别在目前使用最多的两个公共数据库^[13-14]上实现行为序列分割和识别算法, 以下简称 Dataset1^[13], Dataset2^[14]。Dataset1 由 9 个人分别完成 10 种行为: bend, jack, pjump, run, side, skip, walk, wave1, wave2。Dataset2 由 25 个人分别完成 6 种行为: box, clap, jog, run, walk, wave。本文将数据库中每个人的行为段随机组合构成长视频行为序列, 然后分割该序列。图 3 给出了 Dataset1 中“dennis”的长视频行为序列中的一些例子帧。

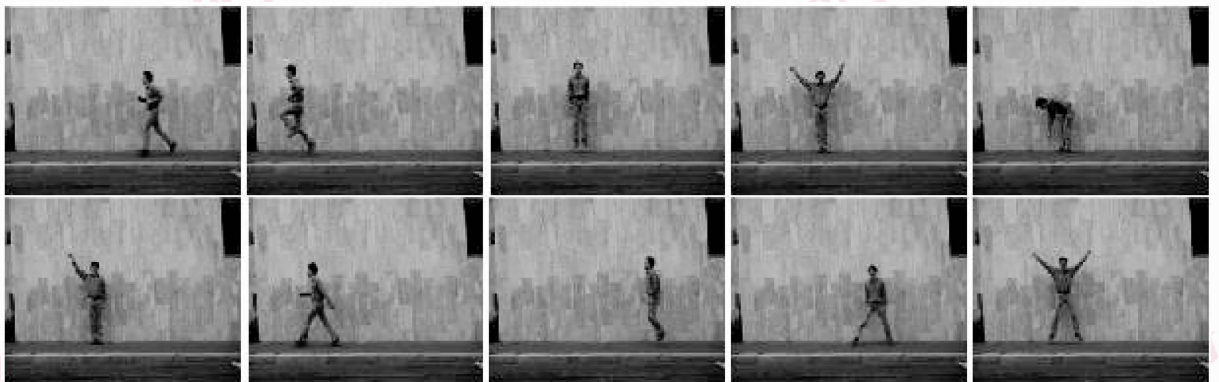


图 3 “dennis”的视频行为序列中的部分图像帧

Fig. 3 Example frames of an activity sequence of “dennis”

4.1 行为序列分割结果

图 4 给出了在 Dataset1 上分别采用本文分割算

法和手动分割算法得到的部分结果比较,横轴对应图像帧号,每条垂线对应行为段的分割边界。

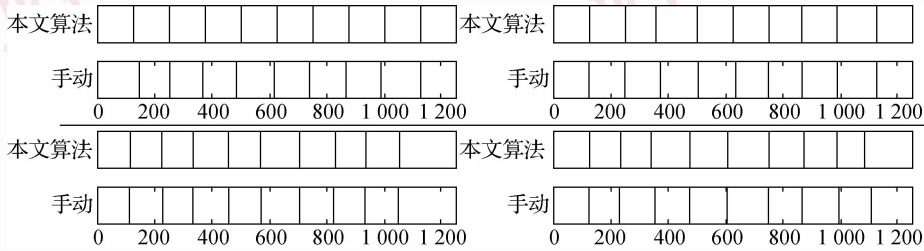


图 4 本文算法与手动分割算法结果比较

Fig. 4 Comparisons of the segmentation results by our algorithm and manually

由图可知,本文算法的大部分分割结果与手动分割结果一致,部分分割边界存在偏移,但这些轻微的位置偏移并不影响后续的行为识别。另外,本文算法得到的分割结果中存在分割边界丢失的情况,这种漏分割将影响后续行为理解的准确率。定义漏分割率为序列被漏分割的行为段数占总行为段数的百分比,在 Dataset1 上进行 20 次随机试验,漏分割率为 8.67%,说明漏分割的发生概率较低。下面统计本文分割算法的分割准确率,即被准确分割的图像帧数目占序列总帧数的百分比,进一步说明分割算法的性能。表 2 给出了 Dataset1 中每个人对应的分割准确率,由表可知,分割算法能准确分割视频行为序列,同时也说明漏分割概率较小。

4.2 行为识别结果

本文采用隐马尔可夫模型^[15](HMM)进行行为的建模和识别。HMM 的隐含节点个数为 3,且每个节点采用 3 个混合高斯模型(GMM)建模。在 Dataset1 中随机选取 4 个人的行为作为训练样本,剩余 5 人的行为随机组合构成长视频序列,采用本文算法进行分割,并使用 HMM 模型进行识别。对数据库 Dataset1 采用手动分割及本文分割算法得到的结果分别进行识别,表 3 给出了 20 次随机试验的

表 2 在 Dataset1 上的分割准确率

Tab. 2 The temporal segmentation accuracy on Dataset1

个体	序列长度	准确率 (%)	个体	序列长度	准确率 (%)
daria	1 167	98.46	lena	1 273	97.03
denis	1 243	98.86	lyova	1 266	98.29
eli	1 239	98.89	moshe	1 240	90.11
ido	1 278	99.38	shahar	1 210	95.54
ira	1 232	93.18			

表 3 在 Dataset1 上的识别率比较

Tab. 3 Recognition accuracy comparison on Dataset1

单位:%

行为类型	手动分割	本文算法	行为类型	手动分割	本文算法
bend	90	84.6	jack	100	99
jump	88	83.6	pjump	74	65.2
run	84	82	side	100	95.4
skip	64	59	wave1	86	84.8
walk	100	94	wave2	98	93.2

平均识别率。表 4 给出了在数据库 Dataset 2 上进行相似实验得到的结果。从表 3 和表 4 可知,对本文算法分割结果的识别率略低于手动分割结果的识别率,说明本文分割算法的有效性。

表 4 在 Dataset2 上的识别率比较

Tab. 4 Recognition accuracy comparison on Dataset2

单位:%

行为类型	box	clap	jog	run	walk	wave
手动分割	72.35	61	69.15	66.7	92.15	87.15
本文算法	71.85	60.05	68.9	66.5	91.85	87.1

5 结论

智能视频监控系统中的行为识别是当前计算机视觉的研究热点之一,现有大多数行为识别算法忽略了行为序列的分割。本文提出了一种基于时间序列不连续性检测的无监督行为分割算法,并采用 HMM 模型识别分割结果,实验结果表明了本文算法的有效性。

参考文献 (References)

- 1 Bobick A F, Davis J W. The recognition of human movement using temporal templates [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(3):257-267.
- 2 Gong Shao-gang, Xiang Tao. Recognition of group activities using dynamic probabilistic networks [A]. In: *Proceedings of International Conference on Computer Vision [C]*, Nice, France, 2003: 742-749.
- 3 Oliver N M, Rosario B, Pentland A P. A bayesian computer vision system for modeling human interactions [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 831-843.
- 4 Zelnik-manor L, Irani M. Event-based analysis of video [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, Hawaii, USA, 2001, **2**:123-130.
- 5 Zhong Hua, Shi Jian-bo, Visontai M. Detecting unusual activity in video [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, Washington, DC, USA, 2004, **2**: 819-826.
- 6 Yong Rui, Anandan P. Segmenting visual actions based on spatio-temporal motion patterns [A]. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C]*, Hilton Head, SC, USA, 2000, **1**:111-118.
- 7 Xiang Tao, Gong Shao-gang. Video behaviour profiling and abnormality detection without manual labeling [A]. In: *Proceedings of International Conference on Computer Vision [C]*, Beijing, China, 2005: 1238-1245.
- 8 Wu Xiao-jie, Hu Zhan-yi, Wu Yi-hong. Unsupervised behavior sequence segmentation based on Segmental-DTW [J]. *Journal of Software*, 2008, **19**(9):2285-2292. [吴晓婕,胡占义,吴毅红. 基于 Segmental-DTW 的无监督行为序列分割[J]. *软件学报*, 2008, **19**(9):2285-2292.]
- 9 Barbic J, Safonova A, Pan J Y, *et al.* Segmenting motion capture data into distinct behaviors [A]. In: *Proceedings of Graphics Interface [C]*, Ontario, Canada, 2004, **62**: 185-194.
- 10 Ali A, Aggarwal J K. Segmentation and recognition of continuous human activity [A]. In: *Proceedings of IEEE Workshop on Detection and Recognition of Events in Video [C]*, Vancouver, BC, Canada, 2001: 28-35.
- 11 Chen Duan-yu, Liao Hongyuanmark, Shih Sheng-wen. Continuous human action segmentation and recognition using a spatio-temporal probabilistic framework [A]. In: *Proceedings of IEEE International Symposium on Multimedia [C]*, San Diego, CA, USA, 2006: 275-282.
- 12 Bauchhage C, Tsotsos, J K, Bunn F E. Automatic detection of abnormal gait[J]. *Image and Vision Computing*, 2009, **27**(1-2): 108-115.
- 13 <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html>, 2007-12-24/2009-06-30.
- 14 <http://www.nada.kth.se/cvap/actions/>. 2005-01-18/2009-06-30.
- 15 <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>, 2005-06-08/.