

单目视频中的多视角行为识别方法

李妍婷, 罗予频, 唐光荣

(清华大学 自动化系, 北京 100084)

(liyanting99@mails.tsinghua.edu.cn)

摘要:针对多视角人体行为识别中,因选取单一特征所造成的识别困难,提出了一种基于分类特征提取的多视角行为识别方法。该方法通过有效利用视频的二维图像序列所包含的三维信息、实现对人体朝向的聚类,解决因观察角度不同造成的行为特征难以选取的问题。实验证明,该方法对单目视频中不同角度的行为动作有较高的识别率。

关键词:图像序列;运动骨架;行为识别;多视角

中图分类号: TP391.41 **文献标识码:** A

Activity recognition method of multiple view angles from monocular videos

LI Yan-ting, LUO Yu-pin, TANG Guang-rong

(Department of Automation, Tsinghua University, Beijing 100084, China)

Abstract: An activity recognition method of multiple view angles from monocular videos was introduced. To aim at recognition in un-calibrated videos, this method could solve the problem choosing a uniform activity characteristic sector of different view angles by utilizing 3-D information included in 2-D image sequence effectively and classifying human skeleton. The experiments show that it has good recognition rate for activity of different view angles from monocular videos.

Key words: image sequence; motion skeleton; activity recognition; multiple view angles

0 引言

近年来,视频人体行为识别逐渐成为计算机视觉领域的一个研究热点,它在视觉监控系统、智能家居应用、体育运动分析等多方面都具有广阔的应用前景。目前,在银行、商店、停车场等处广泛应用的摄像机,通常只用在异常情况发生后,通过回放录像查找可疑人员,无法实时报警。若能对视频中人体行为自动进行识别,则可以在事件发生时给出提示,避免损失产生。

常见的视频人体行为识别方法有两种^[1]:第一种是模板匹配方法,将图像序列转换为一组静态形状模式,然后在识别过程中和预先存储的行为标本相比较得到识别结果,如Polana与Nelson利用二维网格的特征进行人的运动识别^[2]。模板匹配技术的优点是计算复杂度低、实现简单,但对噪声比较敏感,识别结果比较依赖模板的选择,应用于非特定人的复杂场合困难较大。第二种是状态空间方法,将每个静态姿势作为一个状态,这些状态之间通过某种概率联系起来。任何运动序列可以看作这些静态姿势不同状态之间的一次遍历过程,在这些遍历期间计算联合概率,其最大值被选择作为分类行为的标准。状态空间法虽然涉及迭代运算,计算相对较为复杂,但受噪声影响较小,能更好刻画动作本质特征,应用范围更为广泛。

从视频图像序列识别人体行为时,视频中人体朝向不固定,相同的动作从不同的角度观察所得的二维图像差别很大。如何由二维图像序列中重新恢复有用的三维信息,对不同角度的人体行为实现识别,成为该领域的一个重要研究课题。

一种常见的方法是利用多个不同角度的摄像机,利用标定的方法,对人体模型进行三维信息恢复^[3]。但在实际的应用场合,摄像机的个数不易满足,其内外参数也不如实验时那么容易获得,对摄像机进行标定的方法难以应用。另一种方法是利用透视投影及相关先验知识^[4],或在直接线性变换技术以及人体生物学模型的基础上,对人体进行了三维空间重构^[5]。这些方法在一定条件下都可以由二维图像序列获得三维坐标,在获得模型的三维坐标后可以解决多视角行为识别的问题,但计算相对复杂,并且比较依赖先验知识。本文通过直接利用人体二维模型包含的三维信息,提出了一种基于状态空间,可应用于未标定单目视频,对多角度的人体行为进行识别的方法。

无论是作为人体行为识别的基础研究,还是应用于现实生活(如商店监控系统),人的“走”与“跑”这两种行为都有非常重要的研究价值,本文对人在日常生活中常见的步行及跑步两种行为进行了识别。

1 基于状态空间的多视角行为识别系统

本文使用状态空间法对人体行为进行识别,该类识别系统通常包括三个步骤:模型建立、特征提取和动作识别,其系统框架如图1所示。

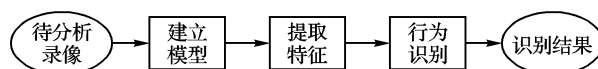


图1 基于状态空间的行为识别系统框架

模型建立的方法大致可分为两类,一种是利用人体轮廓建立模型^[6],另一种是利用人体骨架建立模型^[3]。两种方法

收稿日期:2006-01-09

作者简介:李妍婷(1982-),女,贵州遵义人,硕士研究生,主要研究方向:图像处理、人体行为识别; 罗予频(1959-),男,湖南湘潭人,教授,博士,主要研究方向:计算几何、图论、人工智能; 唐光荣(1946-),男,湖北武汉人,教授,主要研究方向:计算机图形图像处理。

各有利弊,轮廓信息获得相对容易,但是无法表征动作细节,对存在自遮挡现象处理不好;骨架模型获取较为困难,但对人体运动关节信息描述良好,能较好地刻画运动本质,更适用于多视角行为识别问题。

特征选取是人体行为识别中的重点与难点,所选特征的好坏将直接影响识别结果。不同场合、不同方法下的特征千差万别。Feng Niu^[6]在识别人体行、跑、站、坐等动作时利用优化后的表征运动的128维向量及表征形态的90维向量作为送入识别器的特征向量。胡长勃等^[7]在太极拳识别中将每帧中头与手的相对位置关系排列起来,组成该动作的特征向量。在单目视频多视角的人体行为识别中,由于同一动作在不同朝向时的二维表现形式相差较大,选取特征成为这一问题中的难点。

行为识别的方法有很多,如隐马尔可夫模型(Hidden Markov Model, HMM)^[6],主元分析(Principal Component Analysis, PCA)^[7],贝叶斯网络(Bayesian Networks, BN)^[8]等。这些方法各有利弊,在行为识别场合中都有较为广泛的应用。在建立人体骨架模型后,该类模型维数比轮廓模型相对要低,待识别动作类型并不繁杂,贝叶斯网络对此类问题有较好的分类识别能力,故本文在行为识别部分选用了贝叶斯网络方法进行分类。

贝叶斯网络计算的实现过程主要包括以下几个步骤:

- 1) 建立一个表达有向无环的网络结构图,计算联合概率;
- 2) 设置局部概率分布;
- 3) 贝叶斯网络的学习。

1.1 建立人体骨架模型

本文选择人体骨架模型,建立样本集 $S(S_1, S_2, \dots, S_n)$,每个样本 S_i 均从单目视频中的连续31帧图像中得到,即第 i 个样本 S_i 是一个骨架序列,由31个人体骨架 $S_{i,1}, S_{i,2}, \dots, S_{i,31}$ 组成,每个骨架均有16个关节点,如图2所示。

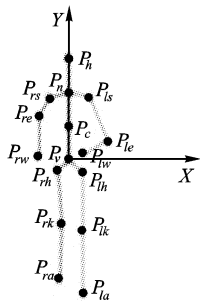


图2 $S_{i,j}$ 的骨架示意图

以身体腹部为原点,建立如图2所示的坐标系,其中, P_h, P_n, P_c, P_e 为人体骨架头部、颈部、胸部、腹部关节点的坐标; $P_{ls}, P_{le}, P_{lw}, P_{lh}, P_{lk}, P_{la}$ 为人体左侧肩、肘、腕、臀、膝、踝关节点的坐标; $P_{rs}, P_{re}, P_{rw}, P_{rh}, P_{rk}, P_{ra}$ 为人体右侧肩、肘、腕、臀、膝、踝关节点的坐标。

1.2 单一特征提取与行为识别

在通常的行为识别中,最直观和常见的想法,是对每个待识别动作按照某种规则提取出统一的特征,将此特征送入识别器训练,得到最终的动作识别器。这里选取的特征能反映出不同动作间的本质差别(如人体骨架模型大小腿之间的夹角 θ 能反映出“走”与“跑”两个动作的区别)。该方法流程如图3所示。

具体实现方法如下:

- 1) 对训练样本集 S_{tr} 中每个样本 S_i 的31个骨架 $S_{i,j}(j=1, \dots, 31)$ 计算各自大小腿之间的夹角 $\theta_{i,j}$,则第 i 个样本 S_i 的特征向量为 $\mathbf{K}_i = (\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,31})^T (i=1, 2, \dots, n)$,训练样本集 $S_{tr} = \{S_1, S_2, \dots, S_n\}$ 得到与之对应的特征向量集 $\mathbf{K}_{tr} = \{\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_n\}$ 。具体计算方法为: $\theta_{i,j} = \text{ang}(P_{lk}, P_{la}) + \text{ang}(P_{lk}, P_{lh})$,其中, $\text{ang}(P_0, P_1) = \left| \arctan\left(\frac{y_1 - y_0}{x_1 - x_0}\right) \right|$, P_0 为 (x_0, y_0) , P_1 为 (x_1, y_1) 。
- 2) 将训练样本集 S_{tr} 所对应的特征向量集 \mathbf{K}_{tr} 按动作分类,分别送入贝叶斯网络进行学习训练,得到分类器。
- 3) 对测试样本集 S_{te} 按1)中的方法计算其对应的特征向量集 \mathbf{K}_{te} 送入2)中所得的分类器,得到识别结果。

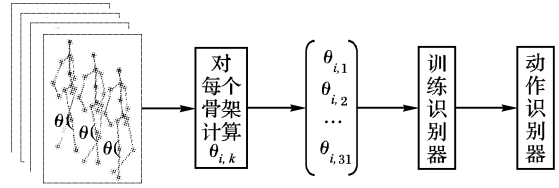


图3 单一特征提取方法的流程

2 多视角分类特征提取算法

当待识别动作的视角不确定时,前面的单一特征提取方法,对不同视角的同一动作,很难选取统一的特征。同时注意到,当人体朝向不同时,同一个方向的骨架模型具有某些相同的特征。如图4所示,(a),(b),(c)中的躯干(头、两肩、腹部的相对位置关系)互不相同,这种特征与动作无关,而与视角有关。因此,可以利用躯干特征对人体模型的视角进行预处理,对模型进行大致分类后,再对每个视角中的不同动作选取相应特征,避免了用一种特征表征所有方向的动作,使问题得到解决,即多视角分类特征提取方法。

2.1 视角的定义

由于人体具有的对称性及步行和跑步动作具有的周期性,本文在实验时将视角分为三类,如图4所示,(a)中身体平面垂直于观察方向时为正视角,(b)中身体平面与观察方向成一定夹角时为斜视角,(c)中身体平面平行于观察方向时为侧视角。

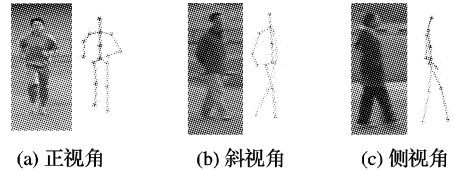


图4 视角分类

2.2 算法流程

该方法流程如图5所示,首先对所有样本计算其视角特征,训练得到视角识别器后,即可将样本集划分为三个视角的子样本集,再对三个子集分别选择该视角下的动作特征,训练得到该视角下的动作识别器。

2.2.1 提取视角特征

对训练样本集 S_{tr} 中的每个样本 S_i 的31个骨架 $S_{i,j}(j=1, \dots, 31)$ 计算其各自视角特征:对前15个骨架 $S_{i,1}, \dots, S_{i,15}$ 计算其肩部与脊柱之间的夹角 $\alpha_{i,j}(j=1, \dots, 15)$,对后16个骨架 $S_{i,16}, \dots, S_{i,31}$ 计算其两肩之间规格化后的距离 $d_{i,j}(j=16, \dots, 31)$,则第 i 个样本 S_i 的表征视角信息的特征

向量为 $V_i = (\alpha_{i,1}, \dots, \alpha_{i,15}, d_{i,16}, \dots, d_{i,31})^T (i = 1, 2, \dots, n)$, 训练样本集 $S_{tr} = \{S_1, S_2, \dots, S_n\}$ 得到与之对应的特征向量集 $V_{tr} = \{V_1, V_2, \dots, V_n\}$ 。

具体计算方法为:

$$\alpha_{i,j} = \text{ang}(P_v, P_h) + \text{ang}(P_{lk}, P_{rs})$$

$$d_{i,j} = \frac{\text{dis}(P_{lk}, P_{rs})}{\text{dis}(P_h, P_v)}$$

其中, $\text{dis}(P_0, P_1) = \sqrt{(x_0 - x_1)^2 + (y_0 - y_1)^2}$, P_0 为 (x_0, y_0) , P_1 为 (x_1, y_1) 。

2.2.2 训练视角分类器

将训练样本集 S_{tr} 所对应的特征向量集 V_{tr} 按视角分类, 分别送入贝叶斯网络进行学习训练, 得到视角分类器。

2.2.3 分视角提取动作特征

将训练样本集 S_{tr} 按照视角分类, 得到正视角、斜视角、侧视角的三个子样本集 S_{front} , $S_{oblique}$, S_{side} 。对这三个子样本集分别选择不同的动作特征, 对正视角样本集 S_{front} 计算其小腿规格化后的长度 $L_{i,j}$, 特征向量 $K_i = (L_{i,1}, L_{i,2}, L_{i,3})^T$, 对应的特征样本集 $K_{front} = \{K_1, K_2, \dots, K_n\}$; 对侧视角样本集 S_{side} 计算大小腿之间的夹角 $\theta_{i,j}$, 特征向量 $K(\theta_{i,1}, \theta_{i,2}, \dots, \theta_{i,31})^T$, 对应的特征样本集 $K_{side} = \{K_1, K_2, \dots, K_n\}$; 对斜视角 $S_{oblique}$ 样本的前 15 个骨架 $S_{i,1}, \dots, S_{i,15}$ 计算其大小腿之间的夹角 $\theta_{i,j} (j = 1, \dots, 15)$, 对后 16 个骨架 $S_{i,16}, \dots, S_{i,31}$ 计算其小腿规格化后的长度 $L_{i,j} (j = 16, \dots, 31)$, 特征向量 $V_i = (\alpha_{i,1}, \dots, \alpha_{i,15}, d_{i,16}, \dots, d_{i,31})^T$, 对应的特征样本集 $K_{oblique} = \{K_1, K_2, \dots, K_n\}$ 。

具体计算方法为:

$$\text{正视角 } L_{i,j} = \frac{\text{dis}(P_{lk}, P_{la})}{\text{dis}(P_h, P_v)}$$

$$\text{斜视角 } k = \text{ang}(P_{lk}, P_{la}) + \text{ang}(P_{lk}, P_{lh})$$

$$\text{侧视角 } k = \text{ang}(P_{lk}, P_{la}) + \text{ang}(P_{lk}, P_{lh})$$

2.2.4 分别训练三个视角的动作分类器

将三个子样本集 S_{front} , $S_{oblique}$, S_{side} 对应的三个特征向量集 K_{front} , $K_{oblique}$, K_{side} 分别送入贝叶斯网络进行训练, 得到该类视角的动作识别器 (这里, 动作识别器的个数 = 视角识别器个数 \times 动作类别数 = $3 \times 2 = 6$)。

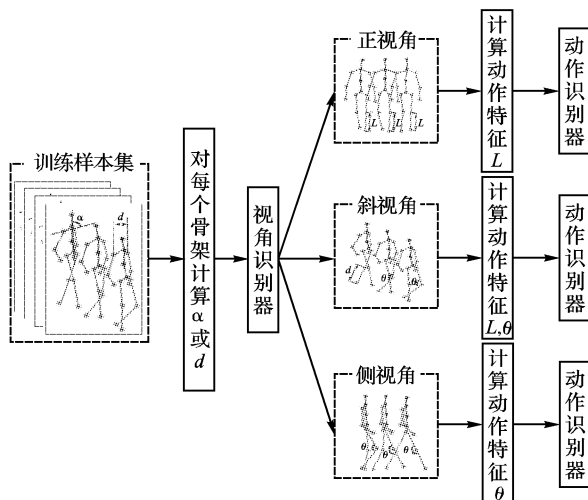


图5 分类特征提取方法流程

2.3 行为识别

对测试样本集 S_{te} 计算其对应的视角特征向量集 V_{te} 送入

视角识别器, 进行分类, 将测试样本集分为正视角 S_1 , 斜视角 S_2 , 侧视角 S_3 三类。再对 S_1, S_2, S_3 这三类找到其对应的动作特征计算方法, 将计算所得的动作特征向量按其所属类别送入动作分类器, 得到最终的识别结果。

3 实验结果

本文选择了学校操场上同学不同朝向的步行及跑步的录像作为实验数据, 分别用单一特征提取的方法和分类特征提取的方法进行实验, 两种方法所用训练及测试样本完全相同。

从正视角、斜视角和侧视角样本中各选出 10 个步行、10 个跑步样本图像序列作为训练样本 (共 30 个步行、30 个跑步样本)。

在测试时, 共使用了 166 个样本, 其中步行与跑步的样本各 83 个。测试所得实验结果如表 1 所示。

表1 识别结果

动作	视角	测试 样本数	单一特征提取		分类特征提取	
			正确数	识别率(%)	正确数	识别率(%)
步行	正	23	17		18	
	斜	30	20	74.70	26	86.75
	侧	30	25		28	
跑步	正	23	10		15	
	斜	30	18	65.06	24	83.13
	侧	30	24		29	

4 结语

针对多视角人体行为识别中, 难以选取单一特征造成的识别困难的问题, 本文提出了一种单目视频中基于分类特征提取的识别方法。实验结果表明, 使用本文方法做预处理, 进行分类特征提取的人体行为识别率能达到 85% 左右, 比单一特征提取的方法识别率提高了 10% ~ 15%, 同时获得人体模型的大致朝向, 为后续研究提供便利。

目前本文的分类方法适用于较为简单的动作, 在后续研究中, 还需要对动作的视角划分作更细致的定义, 对骨架模型的快速有效提取也有更高的要求。

参考文献:

- [1] 王亮, 胡卫明, 谭铁牛. 人运动的视觉分析综述[J]. 计算机学报, 2002, 25(3): 225-237.
- [2] POLANA R, NELSON R. Low level recognition of human motion [A]. Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects[C], 1994. 77-82.
- [3] 罗忠祥. 视频流中的人体运动提取与运动合成[D]. 浙江大学, 2002.
- [4] 陈家实, 庄越挺, 朱强, 等. 透视投影下三维运动重建[J]. 计算机辅助设计与图形学学报, 2002, 14(11): 1041-1046.
- [5] 杨锋, 丁立, 杨春信, 等. 利用图像序列和人体模型重构人体的三维运动[J]. 生物医学工程学杂志, 2005, 22(2): 307.
- [6] NIU F, ABDEL-MOTTALEB M. View-invariant human activity recognition based on shape and motion features[A]. Proceedings of IEEE Sixth International Symposium on Multimedia Software Engineering[C], 2004. 546-556.
- [7] 胡长勃, 冯涛, 马颂德, 等. 基于主元分析法的行为识别[J]. 中国图象图形学报, 2000, 5A(10): 818-824.
- [8] REHG JM, MURPHY KP, FIEGUTH PW. Vision-based speaker detection using Bayesian networks[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C], 1999.