

基于光流速度分量加权的人体行为识别

张飞燕, 李俊峰

(浙江理工大学机械与自动控制学院, 杭州 310018)

摘要: 为了提高光流特征对人体行为的描述性,提出一种新的人体行为识别方法。首先,将提取的光流特征分解为 u 和 v 两个速度分量分别来描述行为,通过直接构造视觉词汇表分别得到不同行为两个速度分量的标准视觉词汇码本,并利用训练视频得到每个行为的不同分量的标准词汇分布;进而根据不同行为两个分量的标准视觉词汇码本,分别计算测试视频相应的速度分量的词汇分布,并利用与各行行为两个速度分量的标准词汇分布的加权相似性度量进行行为识别;最后在 KTH 数据库和 Weizmann 数据库中进行实验。实验结果表明,与其它行为识别方法相比笔者提出的方法可以明显提高行为平均识别率。

关键词: 行为识别; 光流特征; 速度分量 u 和 v ; 视觉词汇表; 加权

中图分类号: TP391.4 **文献标志码:** A

0 引言

近年来,人体行为识别是计算机视觉领域的一个研究热点,被广泛应用于视频监控和人机智能交互等领域^[1]。由于大部分人体动作具有相似性,还有遮挡、光照变化等原因,使得人体行为识别的研究面临重大的挑战,许多人体行为识别的研究都在致力解决这些问题。

根据当前的研究方法,人体行为识别研究可以分为二类:基于全局特征和局部特征的行为识别方法。基于全局特征的方法通常采用边缘、光流、剪影轮廓等信息对检测出的整体感兴趣的人体区域进行描述,对噪声、视角变化、部分遮挡比较敏感。如杜鉴豪等^[2]利用混合高斯模型来自适应更新背景模型,对视频序列提取的运动前景进行区域标记,然后采用 Lucas-Kanade^[3]光流算法得到运动区域内的光流信息,并采用基于幅值的加权方向直方图描述行为;傅博等^[4]利用双背景模型来自适应更新背景模型,对视频序列运动前景中的最小邻接矩形区

域采用 Lucas-Kanade 算法计算光流信息,进而利用运动目标的单位加权光流能量进行行为识别;Zhen 等^[5]对运动历史图像(MHI)和三个正交平面(TOPS)提取的时空立方体的运动和结构信息进行编码,并采用二维拉普拉斯金字塔编码描述符;Guo 等^[6]利用经验协方差矩阵对提取的光流信息降维得到协方差描述符,并利用对数化的协方差描述符进行行为识别;Chaudhry 等^[7]提取了一种面向光流特征直方图的特征来描述运动行为,该特征提取时不需要任何人体分割和背景减除;Li 等^[8]提出了一种基于深度图像 3D 关节点抽样特征包的抽样系统行为识别方法,通过从深度图像序列中提取表征人体姿态的 3D 关节点来描述人体行为;在光流场特征的基础上,王乔等^[9]融合了运动角度变化特征形成新特征——能量特征来表征行为。

基于局部特征的方法是对人体中感兴趣的块或者点提取后进行描述,不需要对人体进行精确定位和跟踪,而且局部特征对部分遮挡、视角变化等不敏感。因此在行为识别中局部特征使用频率比较高。

收稿日期: 2014-07-08

基金项目: 国家自然科学基金项目(61374022);国家“863”高技术项目研究与发展计划项目(2009AA04Z139);浙江省自然科学基金项目(Y1100028)

作者简介: 张飞燕(1990-),女,浙江上虞人,硕士研究生,主要从事智能视频监控系统中人体异常行为识别方面的研究。

通信作者: 李俊峰, E-mail: ljf2003@zstu.edu.cn

如 Tom 等^[10]从压缩视频序列中提取量化参数和运动矢量作为特征; Mota 等^[11]利用 3D-HOG 特征和光流特征来描述视频行为; Tang 等^[12]提取了视频序列中 3D-SIFT 特征; Laptev 等^[13]将 HOG 特征和 HOF 特征结合起来描述视频序列中的时空立方体; Holte 等^[14]从四维时空兴趣点附近区域估计三维光流信息,并在提取的局部三维运动描述子和三维光流直方图(HOF3D)特征的基础上利用词袋库和支持向量机进行行为识别; Hao 等^[15]通过在视频帧的感兴趣区域提取密度轨迹特征,进而利用时间金字塔来实现不同行为速度的适应机制; Solmaz 等^[16]提出使用动态系统的稳定性分析来进行拥挤场景的行为识别; Mahbub 等^[17]在光流分析和随机抽样共识的基础上,根据兴趣点位置的水平和垂直方向的平均差和标准差对人体运动进行定位,并利用运动区域的光流均值表征行为。

如何从图像序列中获取能够有效表达人体运动信息的特征是人体行为识别的关键。光流特征是比较好的时空特征,也是运动识别中经常使用的一种运动特征。上述方法中,一部分文献在提取视频运动前景后对前景运动区域进行标记,进而对其进行光流计算。对于人体的各种行为,运动不明显的肢体部位的光流信息是可以忽略的,而上述方法需要计算整个人体区域的光流,这样不仅增加了计算量,而且还会降低识别精度;另外一部分文献在行为识别时没有考虑不同速度分量对行为的描述性,例如将跑步和步行做比较,这两种行为在速度分量 u 上(X 轴)的差别非常显著,而在速度分量 v 上(Y 轴)的差别比较小。如果忽略光流特征的不同速度分量对行为的描述性,那么这两种行为在识别过程中可能会出现一定程度的误判。基于此,为了提高光流特征对行为的描述性,本文提出基于光流速度分量加权的人体行为识别方法。首先对图像序列进行兴趣点检测,根

据兴趣点提取出人体局部运动区域的视频块,并计算这些视频块的光流信息,利用光流信息的两个速度分量 u 和 v (下文简称两个速度分量)分别来描述行为;其次,为了充分利用训练视频数据,直接构造了光流特征的两个速度分量的视觉词汇表^[18],并分别得到了不同行为两个速度分量的标准视觉词汇码本及其在这些标准码本上的标准词汇分布;进而利用测试视频在各行行为标准视觉词汇码本上的词汇分布与标准词汇分布之间的加权相似性度量进行行为识别。

1 光流特征的计算

1.1 兴趣点的检测

为了可以有效检测出图像序列 $I(x, y, t)$ 的兴趣点,本文采用 Burgos-Artizzu 等^[19]提出的一种可分离线性滤波器时空兴趣点检测算法。该时空兴趣点检测算法能够检测出较多的兴趣点,且对周期性行为和非周期性行为都有较好的鲁棒性。算法中空间域仍采用高斯滤波器,时间域上则采用 Gabor 滤波器,响应函数 R 为:

$$R(x, y, t) = (I * g * h_{ev})^2 + (I * g * h_{od})^2 \quad (1)$$

式(1)中, $*$ 为卷积运算符, I 为视频图像, $g = g(x, y, \sigma)$ 为二维高斯平滑核, h_{ev} 和 h_{od} 是在空间域上正交的一维 Gabor 滤波器。

h_{ev} 和 h_{od} 的定义为:

$$h_{ev}(t; \tau; \omega) = -\cos(2\pi t\omega) e^{-\frac{t^2}{\tau^2}} \quad (2)$$

$$h_{od}(t; \tau; \omega) = -\sin(2\pi t\omega) e^{-\frac{t^2}{\tau^2}} \quad (3)$$

式(2)、式(3)中, $\omega = \frac{4}{\pi}$, σ 和 τ 分别为空间域和时间域上的检测尺度,文中取 $\sigma = 2$ 和 $\tau = 3$, 高斯平滑滤波尺度为 2。

图 1 为 KTH^[20] 数据库中 person17 挥拳行为视频中兴点 1 的视频块展开图,从图 1 中可以看出该视频块可以很好地描述挥拳行为。



图 1 KTH 数据库中 person17 挥拳行为兴趣点 1 的视频块展开图

1.2 光流特征的提取

光流场是能描述图像序列如何随时间变化的向量场,包含了像素点的瞬时运动速度矢量信息^[21],是比较好的时空特征。但是光流场计算量较大,为了通过只计算提取的视频块的光流而减少计算量,文本选用 Lucas-Kanade 方法来计算光流。假设在大小为 $x \times x$ 的窗口的光流是保持恒定的,由此可通过求该窗口光流的约束方程得到其光流(u 和 v),

即:

$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xi} & I_{yi} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_{t1} \\ I_{t2} \\ \vdots \\ I_{ti} \end{bmatrix} \quad (4)$$

式(4)中, i 为特征窗口内的像素个数($i = x \times x$), I_x 和 I_y 为图像的空间梯度, I_t 为时间梯度。

求解式(4)可得:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I_{x_i}^2 & \sum I_{x_i} I_{y_i} \\ \sum I_{x_i} I_{y_i} & \sum I_{y_i}^2 \end{bmatrix} \begin{bmatrix} -\sum I_{x_i} I_{t_i} \\ -\sum I_{y_i} I_{t_i} \end{bmatrix} \quad (5)$$

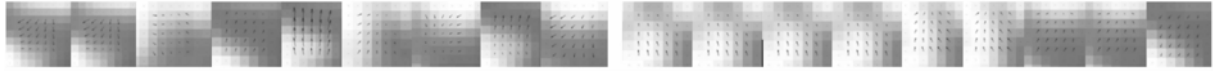


图 2 KTH 数据库 person17 挥拳行为兴趣点 1 的视频块的光流

2 视觉词汇表构造和行为识别

2.1 光流特征的速度分量的视觉词汇表构造

由于视频帧数不同,且不同的人在做相同行为时动作幅度和速度上存在差异,所以同一种行为的不同视频得到的兴趣点及相应提取的视频块数量不一致,导致这些视频提取的光流特征数量也不同。此外,不同行为视频的光流特征的两个速度分量在数值上存在相关性,直接利用光流的两个速度分量进行行为识别会降低对行为的描述性。基于此,为了克服由于不同视频光流特征数量不同造成无法识别的缺陷,本文提出采用视觉词汇表 (bag of words)^[22] 来描述人体行为;为了保证光流特征对行为描述的全面性和有效性,分别提取光流的两个速度分量作为视觉词汇来表征视频中人体运动的特征。

笔者提出如图 3 所示的视觉词汇表构造方法,分别对视频光流特征的两个速度分量的视觉词汇描述子进行聚类,得到两个分量的码本;然后计算每个分量的视觉词汇描述子在该分量码本的直方图分

图 2 为 KTH 数据库中 person17 的挥拳行为兴趣点 1 的视频块的光流,从图 2 中可知随着手部行为的变化,光流也随之发生变化。

布。具体构造过程如下:

a) 对视频进行预处理后,检测时空兴趣点,以兴趣点为中心提取视频块,然后计算各视频块的光流信息 (u 和 v),并将所有视频块的光流信息按两个速度分量的视觉词汇描述子分别聚合作为视觉词汇描述子集合 M_u 和 M_v 来描述行为;

b) 利用聚类算法对得到的两个速度分量的视觉词汇描述子集合分别进行聚类,聚类后每个速度分量都会得到 k 个聚类中心 (码字),每个速度分量的码字就构成了相应的码本 k_u 和 k_v ;

c) 对得到的码本 k_u 和 k_v ,利用最小距离法计算该视频的每个速度分量的视觉词汇描述子集合在相应的两个速度分量码本的码字上的直方图分布,利用它们来表征该视频的人体运动行为。

设 $k_u = \{c_{u1}, c_{u2}, \dots, c_{uk}\}$ 为视频速度分量 u 的码本, M_u 为该视频速度分量 u 的视觉词汇描述子集合,则 M_{ui} (表示 M_u 中第 i 个特征描述子, $i=1, 2, \dots, n$) 到 k_u 中第 j 个码字的距离可以用下式来表示:

$$d_{uj} = \|M_{ui} - c_{uj}\|^2, j=1, 2, \dots, k \quad (6)$$

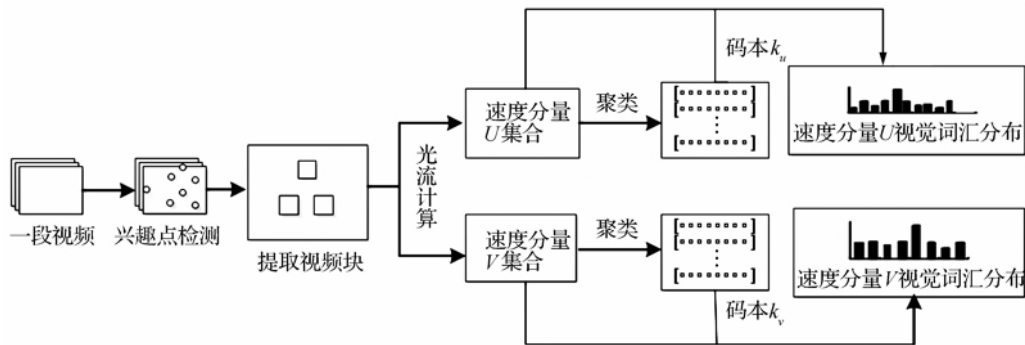


图 3 词汇表构造流程

根据上述视觉词汇表构造方法,选取 KTH 数据库中 person1 表演的挥拳 (box)、跑 (run)、慢跑 (jog)、拍手 (clap)、步行 (walk)、挥手 (wave) 等 6 个行为的视频,分别构造出这 6 种行为的光流特征两个速度分量 u 和 v 的直方图分布,如图 4、图 5 所示。

观察图 4—图 5 可知,在速度分量 u 上, jog、run

及 clap 行为的直方图分布与其它行为的直方图分布区分度很大,原因是这些行为在动作时光流的速度分量 u 变化差异比较大,从而使得其在该速度分量上的直方图区分度比较明显;在速度分量 v 上, box 和 wave 行为与其它行为相比变化差异比较大,从而导致其在速度分量 v 的直方图分布与其它行为明显不同。然而对于一些动作比较相似的行为,它

们在某一个速度分量上的直方图分布区别性就比较小,产生这种现象的原因是这些相似行为在动作时其光流特征在该速度分量的数值变化不明显,从而

引起直方图分布的形状比较相似,如 jog 和 run 行为在速度分量 v 的直方图分布及 walk 和 box 行为在速度分量 u 的直方图分布。

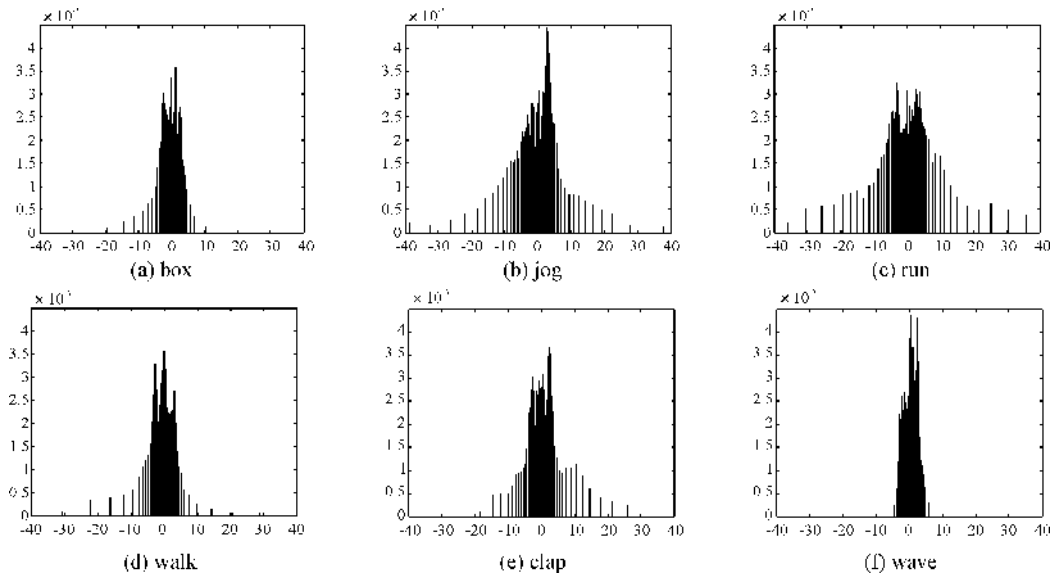


图4 六种行为速度分量 u 的直方图分布

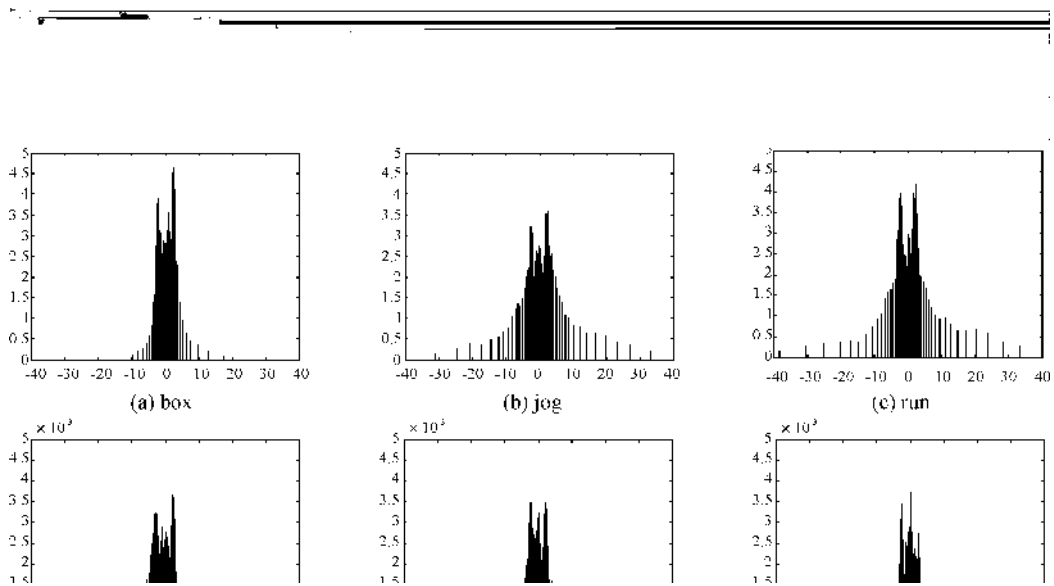


图5 六种行为速度分量 v 的直方图分布

2.2 不同行为两个速度分量的标准直方图分布构造

上文中的6种行为都是同一个人的行为,然而不同人做同种行为时在动作上是有差异的,这会影响到该行为光流特征的两个速度分量的直方图分布,产生差异的主要原因是不同人的身高、着装及在动作时肢体的运动空间范围和运动速度不同。如果直接利用同一个人的6种行为相应的光流特征两个速度分量的直方图分布作为标准进行人体行为识

别,会降低识别率。为了解决这个问题,本文提出构造六个行为的光流特征两个速度分量的标准直方图分布,图6以KTH数据库中box行为为例,给出了该行为光流特征的两个速度分量标准直方图分布的构造原理,具体构造过程如下:

a) 对预处理后的所有box行为视频分别进行时空兴趣点检测,以兴趣点为中心分别提取它们的视频块;计算所有box行为视频的视频块的光流信

息,把光流特征的两个速度分量分别作为视觉词汇描述子;

b) 分别把 box 行为所有视频的视频块的两个分量视觉词汇描述子进行合并得到代表 box 行为的两个速度分量的视觉词汇描述子集合,利用聚类算法分别对两个速度分量的视觉词汇描述子集合进

行聚类,得到两个速度分量的标准码本;

c) 对得到的两个速度分量的标准码本,利用最小距离法计算两个速度分量的视觉词汇描述子集合在对应的两个标准码本的码字上出现的频率,分别形成 box 行为的两个速度分量的标准直方图分布,利用它们来表征 box 行为。

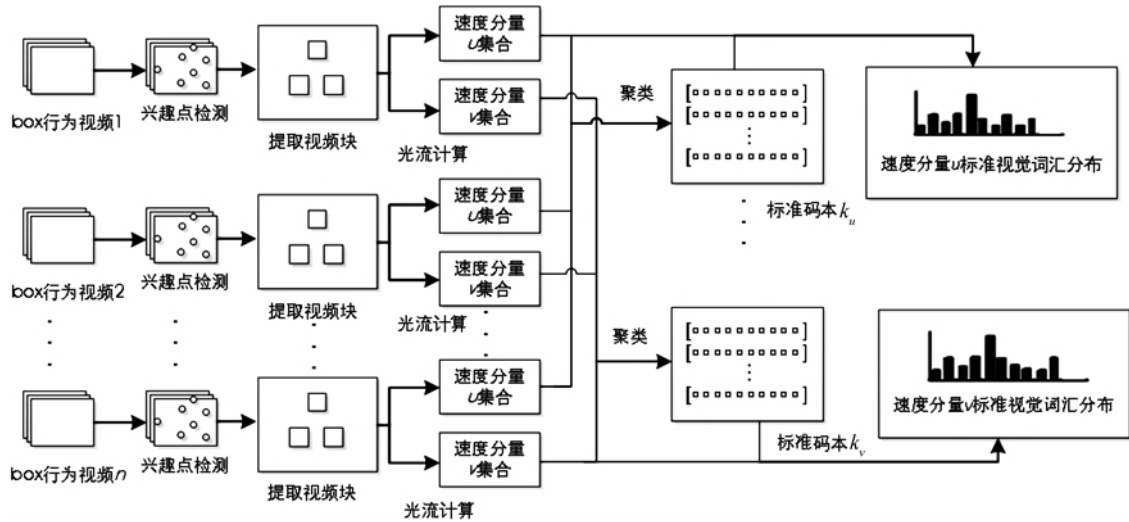


图 6 box 行为两个速度分量的标准直方图分布构造原理

根据 box 行为两个速度分量的标准直方图分布的构造方法,利用 KTH 数据库中 box 行为的所有训练视频来构造该行为两个速度分量的标准直方图分布,图 7—图 8 为所构造的 box、run、jog 等 6 种行为的两个速度分量的标准直方图分布。

从图 7—图 8 中可以看出,大多数行为在两个

速度分量上的标准直方图分布与其它行为在相同速度分量上的标准直方图分布区分度都很明显,可以有效与其它行为区分开来,只有动作比较相似的 jog 和 run 行为在速度分量 u 上的标准直方图分布区分度比较小,但它们在在速度分量 v 上的标准直方图分布区分度很明显,可以有效区分开来。

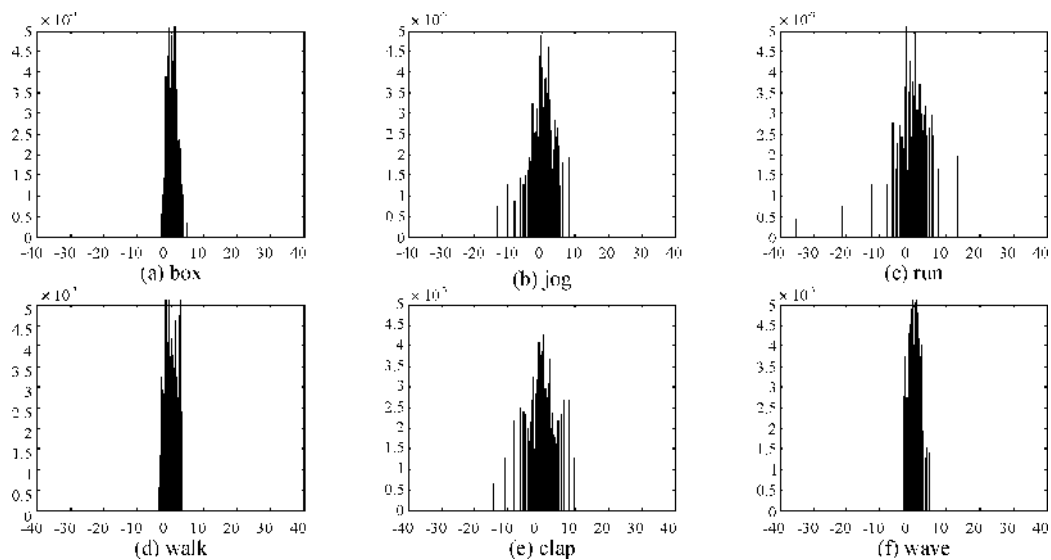
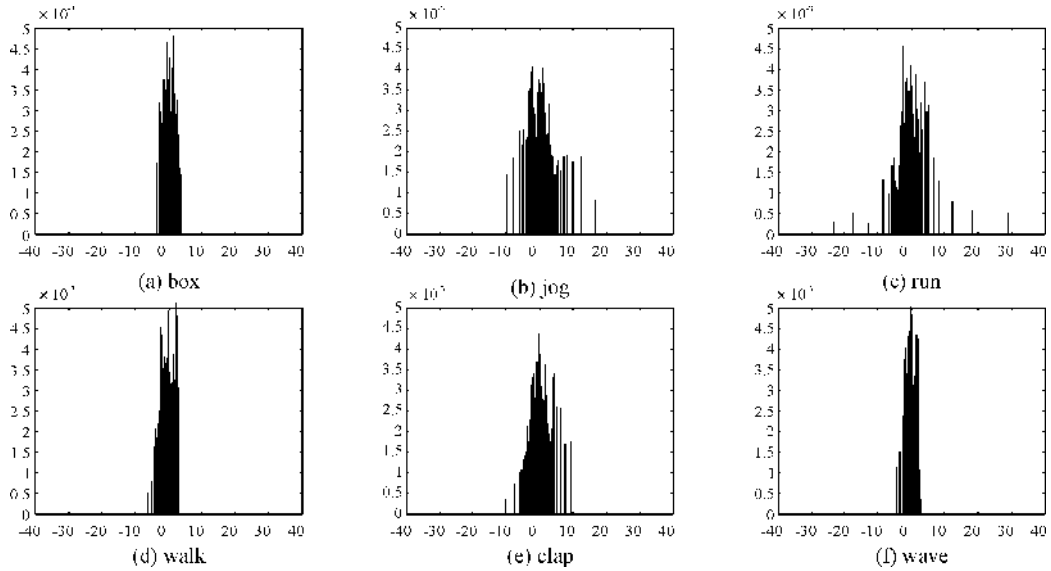


图 7 六种行为速度分量 u 的标准直方图分布

图8 六种行为速度分量 v 的标准直方图分布

2.3 基于光流特征两个速度分量加权的人体行为识别

根据上述分析,不同行为的光流特征的两个速度分量的直方图分布存在差异。即使相似的行为,在两个速度分量的直方图分布中,至少在一个分量上的分布是不同的。如果可以结合视频的两个速度分量的直方图分布进行人体行为识别,可以大大提高人体行为识别率。基于此,本文提出基于光流特征两个速度分量加权的人体行为识别方法,识别原理如图9所示,具体步骤如下:

a) 对待识别视频进行兴趣点检测,提取视频块计算光流特征,把所有视频块的两个速度分量集合作为视觉词汇描述子集合,然后分别计算两个速度

分量的视觉词汇描述子集合在各种行为的两个速度分量的标准码本上的直方图分布得到两个速度分量的测试直方图分布;

b) 分别计算两个速度分量的测试直方图分布与各行为标准两个速度分量的直方图分布的欧氏距离,为了使不同行为区分更为明显,需要对两个速度分量的欧氏距离进行加权,加权系数的范围为 $[0,1]$ 之间,且各分量的加权系数之和为1。各速度分量的欧氏距离的加权系数可以根据各行为两个速度分量的标准词汇分布对行为表示的贡献来优化设计;

c) 根据计算出的加权欧氏距离,利用最近邻原则对行为进行识别。

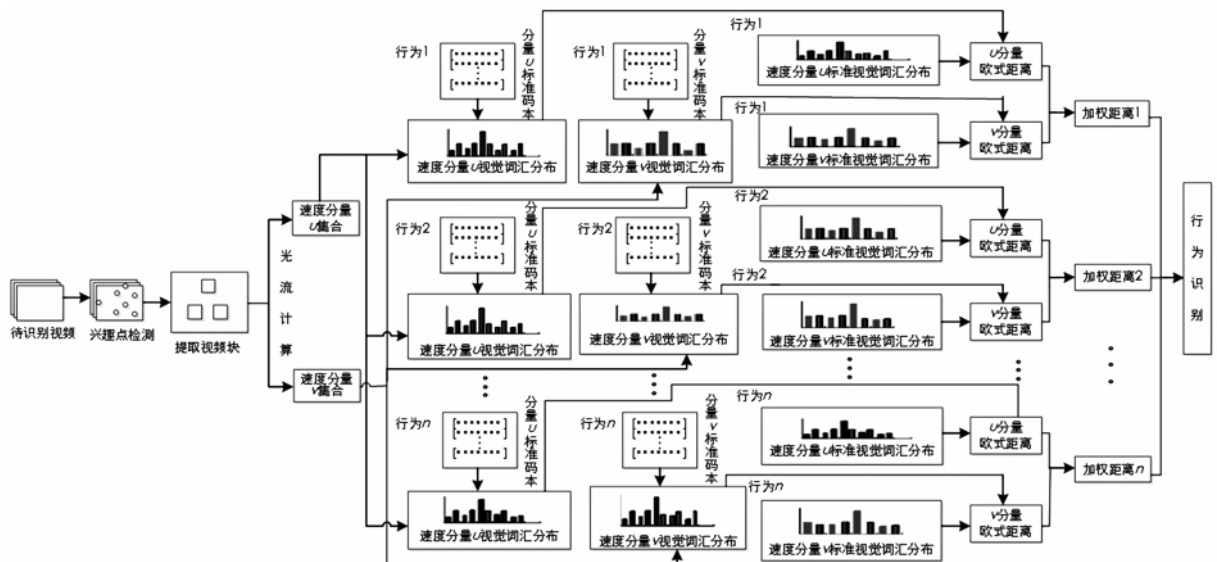


图9 行为识别原理

3 实验

本文在两个公开的标准数据(KTH Action Dataset 和 Weizmann Action Dataset^[23])上进行实验。这两个数据库包含杂乱的背景和单一背景、移动摄像机和静止摄像机等几种不同的环境。本文利用混淆矩阵的形式来表示本文所提方法的行为平均识别率,通过和其它文献的平均识别率比较来体现本文方法对平均行为识别率的提高程度。

3.1 KTH 行为数据库

KTH 数据库包含由 25 个人表演的 6 种行为:挥拳(box)、慢跑(jog)、跑(run)、步行(walk)、拍手(clap)和挥手(wave),共 599 个视频。本文实验采用文献[5、11、13、24]的实验分组方法,将所有视频分为测试集(表演者 2,3,5,6,7,8,9,10,22 的视频)和训练集(其余表演者的视频)。采用 Niebles 等^[24]的“词袋模型”来分别构造两个速度分量的视觉词汇表,其中每个分量的词汇数目设为 500。

图 10 是以混淆矩阵形式表示的 KTH 数据库中的行为识别率,第 i 行第 j 列的数值表示第 i 个行为和第 j 个行为的识别率。在 KTH 数据库中,本文的平均识别准确率达到 100%,可以识别其中的所有行为。

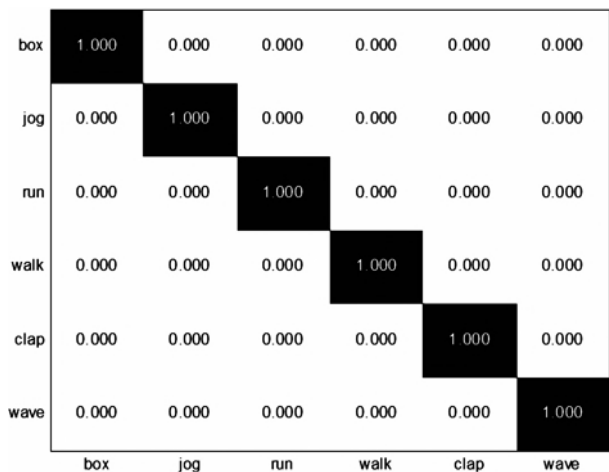


图 10 KTH 数据库的平均识别率

表 1 是其他文献在 KTH 数据库中的平均识别率。从表 1 中可以看出,与其他文献的方法相比,本文方法可以明显提高行为识别率。

表 1 KTH 数据库中不同文献的平均识别率

行为识别方法	发表年份	平均识别率/%
本文方法		100.0
文献[13]	2008	91.8
文献[5]	2013	92.2
文献[25]	2013	93.0
文献[11]	2014	93.2

3.2 Weizmann 行为数据库

Weizmann 数据库包含由 9 个人表演的 10 种行为:弯腰(bend)、原地跳步(jack)、原地纵跳(pjump)、跳步前行(jump)、跑(run)、横向步行(side)、单腿跳行(skip)、步行(walk)、单手挥舞(wave1)和双手挥舞(wave2),共 90 个视频。由于 Weizmann 数据库的视频比较少,在实验中采用文献[6、7、10、26]中的实验分组方法,选取其中一个表演者的 10 个行为视频作为测试样本,其余 8 个表演者的行为视频作为训练样本,这样重复 9 次实验。采用 Niebles 等^[24]的“词袋模型”来分别构造两个速度分量 u 和 v 的视觉词汇表,其中每个分量的词汇数目设为 500。

图 11 是以混淆矩阵的形式表示的 Weizmann 数据库中的行为识别率。由混淆矩阵可以看出 side 和 skip 行为的识别率相对较低,side 行为主要被误判为 jump、skip、run 行为,误判概率分别为 0.037、0.037、0.025。主要原因是 side 行为在动作时有类似跳步和慢跑的动作,与这三个行为都有相似的动作,只不过个别方向的动作速度和幅度稍有不同;skip 行为被误判为 run 行为的概率为 0.037,误判为 jump 行为的概率为 0.025。因为 skip 行为在动作时是跳步前行,与 run 和 jump 行为部分动作相似,只是肢体动作速度和幅度不同。

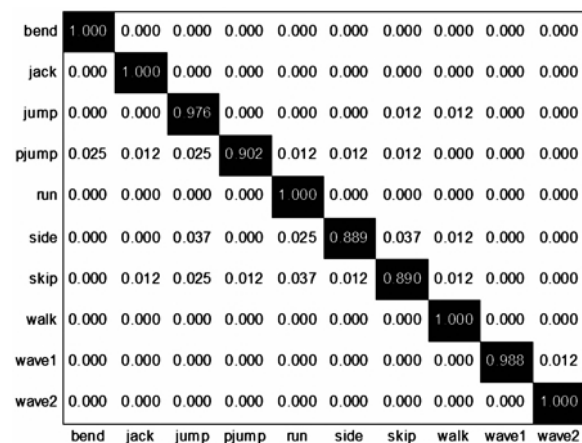


图 11 Weizmann 数据库的平均识别率

表 2 是其他文献在 Weizmann 数据库中的平均识别率。从表 2 中可以看出,本文的平均识别准确率为 96.45%,与其他文献的方法相比,本文方法可以明显提高行为识别率。

表 2 Weizmann 数据库中不同文献的平均识别率

行为识别方法	发表年份	平均识别率/%
本文方法		96.45
文献[7]	2009	95.70
文献[6]	2010	94.40
文献[10]	2013	85.56
文献[26]	2014	92.47

4 结 论

为了提高光流特征对行为的描述性及避免计算人体非运动肢体部位的光流信息,本文先检测视频的兴趣点,根据兴趣点提取人体运动部位的视频块,然后将视频块的光流特征按其两个速度分量分解为 u 和 v 两个分量分别来描述行为,并根据不同行为的动作特点,对 u 和 v 两个分量的直方图加权来进行人体行为识别。由于本文利用人体肢体运动区域的视频块的光流特征两个速度分量,分别直接构造相应分量视觉词汇表,保证了两个速度分量对行为的高描述性,提高了识别的效率,并通过加权考虑了不同行为光流特征的两个速度分量对行为描述的差异性,提高了识别率。实验结果表明,该方法可以明显提高行为平均识别率,其中 KTH 数据库平均识别率高达 100%;Weizmann 数据库平均识别率高达 96.45%。同时,本文在构造视觉词汇表时,没有对各速度分量词汇表的词汇数目进行优化,选取均为 500。下一步的研究重点是改进“词袋模型”的词汇分类方法,优化设计各分量的词汇数目,以进一步提高行为平均识别率和行为识别效率。

参考文献:

- [1] 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.
- [2] 杜鉴豪,许力. 基于区域光流特征的异常行为检测[J]. 浙江大学学报:工学版, 2011, 45(7): 1161-1166.
- [3] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]// Proceedings of the 7th International Joint Conference on Artificial Intelligence. Vancouver, BC, Can: [s. n.], 1981, 81: 674-679.
- [4] 傅博,李文辉,陈博,等. 基于加权光流能量的异常行为检测[J]. 吉林大学学报:工学版, 2013(6): 1644-1649.
- [5] Zhen X T, Shao L. A local descriptor based on Laplacian pyramid coding for action recognition[J]. Pattern Recognition Letters, 2013, 34(15): 1899-1905.
- [6] Guo K, Ishwar P, Konrad J. Action recognition using sparse representation on covariance manifolds of optical flow[C]// 2010 7th IEEE International Conference on Advanced Video and Signal Based, AVSS 2010. Boston, MA, United States: IEEE Computer Society, 2010: 188-195.
- [7] Chaudhry R, Ravichandran A, Hager G, et al. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions[C]// 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops: CVPR Workshops 2009. Miami, FL, United states: IEEE Computer Society, 2009: 1932-1939.
- [8] Li W Q, Zhang Z Y, Liu Z C. Action recognition based on a bag of 3d points[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops: CVPR Workshops 2010. San Francisco, CA, United states: IEEE Computer Society, 2010: 9-14.
- [9] 王乔,雷航,郝宗波. 基于整体能量模型的异常行为检测[J]. 计算机应用研究, 2013, 29(12): 4782-4785.
- [10] Tom M, Venkatesh B R. Rapid human action recognition in H. 264/AVC compressed domain for video surveillance[C]// 2013 IEEE International Conference on Visual Communications and Image Processing, IEEE VCIP 2013. Kuching, Sarawak, Malaysia: IEEE Computer Society, 2013: 1-6.
- [11] Mota V F, Perez E A, Maciel L M, et al. A tensor motion descriptor based on histograms of gradients and optical flow[J]. Pattern Recognition Letters, 2014, 39: 85-91.
- [12] Tang X Q, Xiao G Q. Action Recognition Based on Maximum Entropy Fuzzy Clustering Algorithm[M]. Foundations of Intelligent Systems. [S. l.]: Springer Berlin Heidelberg, 2014: 155-164.
- [13] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]// 26th IEEE Conference on Computer Vision and Pattern Recognition: CVPR. Anchorage, AK, United States: IEEE Computer Society, 2008: 1-8.
- [14] Holte M B, Chakraborty B, Gonzalez J, et al. A local 3-D motion descriptor for multi-view human action recognition from 4-D spatio-temporal interest points[J]. IEEE Journal on Selected Topics in Signal Processing, 2012, 6(5): 553-565.
- [15] Hao Z B, Zhang Q N, Ezquierdo E, et al. Human action recognition by fast dense trajectories[C]// 21st ACM International Conference on Multimedia, MM 2013. Barcelona, Spain: Association for Computing Machinery, 2013: 377-380.
- [16] Solmaz B, Moore B E, Shah M. Identifying behaviors in crowd scenes using stability analysis for dynamical systems[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 2064-2070.
- [17] Mahbub U, Imtiaz H, Ahad R. An optical flow based approach for action recognition[C]// 14th International

- Conference on Computer and Information Technology: ICCIT 2011. Dhaka, Bangladesh: IEEE Computer Society, 2011: 646-651.
- [18] Wang J J Y, Bensmall H, Gao X. Joint learning and weighting of visual vocabulary for bag-of-feature based tissue classification[J]. Pattern Recognition, 2013, 46(12): 3249-3255.
- [19] Burgos-Artizzu X P, Dollár P, Lin D, et al. Social behavior recognition in continuous video [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2012. Providence, RI, United states: IEEE Computer Society, 2012: 1322-1329.
- [20] Laptev I. Recognition of human actions [EB/OL]. (2005-01-18)[2014-07-08]. <http://www.nada.kth.se/cvap/actions/>.
- [21] Vishwakarma S, Agrawal A. A survey on activity recognition and behavior understanding in video surveillance[J]. The Visual Computer, 2013, 29(10): 983-1009.
- [22] Li W X, Yu Q, Sawhney H, et al. Recognizing activities via bag of words for attribute dynamics [C]// 26th IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2013. Portland, OR, United States: IEEE Computer Society, 2013: 2587-2594.
- [23] Lena G. Actions as space-time shapes [EB/OL]. (2007-10-24)[2014-07-08]. <http://www.wisdom.weizmann.ac.il/~vision/SpaceTimeActions.html#Database>.
- [24] Niebles J C, Wang H, LI F F. Unsupervised learning of human action categories using spatial-temporal words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.
- [25] Shi F, Petriu E, Laganieri R. Sampling strategies for real-time action recognition [C]//26th IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2013. Portland, OR, United states: IEEE Computer Society, 2013: 2595-2602.
- [26] Cao X, Zhang H, Deng C, et al. Action recognition using 3D DAISY descriptor[J]. Machine Vision and applications, 2014, 25(1): 159-171.

Human Behavior Recognition Based on Weighted Optical Flow Velocity Component

ZHANG Fei-yan, LI Jun-feng

(School of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310012, China)

Abstract: In order to improve the descriptiveness of optical flow characteristic on human behavior, this paper proposes a kind of new human behavior recognition method. Firstly, the extracted optical flow characteristic is decomposed as 2 velocity components (u and v) to describe the behavior respectively, and then the standard visual vocabulary code book of 2 velocity components can be obtained via direct construction visual vocabulary table, and the testing video is utilized to obtain the distribution of standard vocabulary of different components of each behavior; then, according to standard visual vocabulary code book of 2 components of different behavior, it is able to respectively calculate the vocabulary distribution of corresponding velocity component of testing video, and utilize the weighted similarity measurement of standard vocabulary distribution of 2 velocity components of each behavior to carry out behavior recognition; finally, the test is made in KTH database and Weizmann database; through comparison between experimental result and other behavior recognition methods, it can be found that the average recognition rate of behavior is obviously improved.

Key words: behavior recognition; optical flow characteristic; velocity component u and v ; visual vocabulary table; weighted

(责任编辑: 陈和榜)