



JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
03
VOL.20

ISSN1006-8961
CN11-3758/TB

非真实感图像处理 P437





第20卷第3期（总第227期）

2015年3月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@radi.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@radi.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

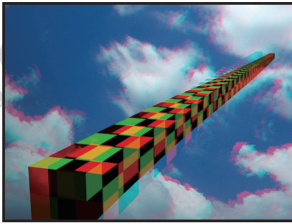
ISSN 1006-8961

CODEN ZTTXFZ

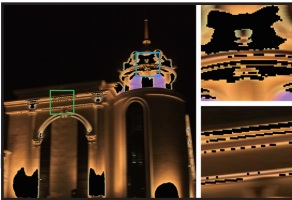
国外发行代号 M1406

国内邮发代号 82-831

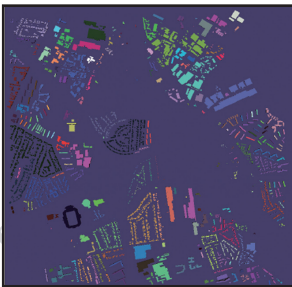
国内定价 50.00元



实时立体视觉系统中的深度映射(第0366页)



基于辐照度的运动模糊图像去模糊(第0395页)



面向智慧城市的3维城市在线可视化(第0445页)

综述

面向血管疾病诊断及预测分析的血流动力学模拟综述

徐梦佳, 杨金柱, 赵宏, 赵大哲 0297

图像处理和编码

多特征融合的车辆阴影消除

邱一川, 张亚英, 刘春梅 0311

图像分析和识别

基于局部时空特征方向加权的人体行为识别

李俊峰, 张飞燕 0320

对表情鲁棒的面部轮廓线3维人脸识别

常朋朋, 达飞鹏, 梅俊 0332

融合双层信息的显著性检测

姜霞霞, 李宗民, 匡振中, 刘玉杰 0340

面向动摄像机的高速运动目标检测

穆春迪, 谢剑斌, 闫玮, 刘通, 李沛秦 0349

匹配滤波在SAR目标单个散射部件检测中的应用

丁柏圆, 文贡坚, 钟金荣, 马聪慧 0357

图像理解和计算机视觉

实时立体视觉系统中的深度映射

李大锦, 白成杰 0366

遥感图像处理

GPU支持的SAR影像几何校正大规模并行处理

杨景辉, 程春泉, 张继贤, 黄国满 0374

基于正则化约束的遥感图像多尺度去模糊

谭海鹏, 曾炫杰, 牛四杰, 陈强, 孙权森 0386

第十届中国计算机图形学大会专栏

基于辐照度的运动模糊图像去模糊

方贤勇, 阚未然, 周健, 李黎, 郭延文 0395

对比度、颜色一致性和灰度像素保持的消色算法

王勋, 罗婷婷, 刘春晓, 彭浩宇 0408

结合因子图的多目的地地图布局优化

罗振珊, 张俊松, 范接鹏 0418

高阶连续的形状可调三角多项式曲线曲面

严兰兰, 韩旭里 0427

改进的自然图像铅笔画效果生成

黄志勇, 姚远, 张晶晶, 马凯, 任东 0437

面向智慧城市的3维城市在线可视化

潘斌, 郭小明, 陈明明, 陈为, 彭群生 0445

CONTENTS

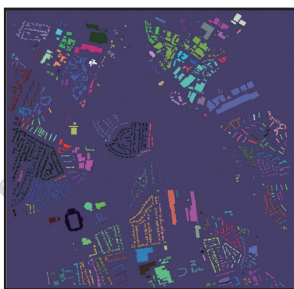
JOURNAL OF IMAGE AND GRAPHICS



Depth mapping in real-time stereoscopic systems (P0366)



Deblurring of motion blurred images based on the irradiance(P0395)



Online 3D city visualization for smarter cities(P0445)

Review

Survey of hemodynamics simulation for diagnosis and prediction in vascular disease

Xu Mengjia, Yang Jinzhu, Zhao Hong, Zhao Dazhe 0297

Image Processing and Coding

Vehicle shadow removal with multi-feature fusion

Qiu Yichuan, Zhang Yaying, Liu Chunmei 0311

Image Analysis and Recognition

Human behavior recognition based on directional weighting local space-time features

Li Junfeng, Zhang Feiyan 0320

Expression-robustness 3D face recognition based on facial profiles

Chang Pengpeng, Da Feipeng, Mei Jun 0332

Image saliency detection based on two-layer information fusion

Jiang Xiaxia, Li Zongmin, Kuang Zhenzhong, Liu Yujie 0340

Detecting high-speed moving targets in moving camera environments

Mu Chundi, Xie Jianbin, Yan Wei, Liu Tong, Li Peiqin 0349

The application of matched filter in detecting single scattering parts of SAR targets

Ding Boyuan, Wen Gongjian, Zhong Jinrong, Ma Conghui 0357

Image Understanding and Computer Vision

Depth mapping in real-time stereoscopic systems

Li Dajin, Bai Chengjie 0366

Remote Sensing Image Processing

GPU supported massively parallel processing for geometric correction of SAR imagery

Yang Jinghui, Cheng Chunquan, Zhang Jixian, Huang Guoman 0374

Remote sensing image multi-scale deblurring based on regularization constraint

Tan Haipeng, Zeng Xuanjie, Niu Sijie, Chen Qiang, Sun Quansen 0386

Column of Chinagraph'2014

Deblurring of motion blurred images based on the irradiance

Fang Xianying, Kan Weiran, Zhou Jian, Li Li, Guo Yanwen 0395

Decolorization with contrast, color consistency and grayscale pixel preservation

Wang Xun, Luo Tingting, Liu Chunxiao, Peng Haoyu 0408

Synthesizing multi-destinations map with factor graph

Luo Zhenshan, Zhang Junsong, Fan Jiepeng 0418

Higher-order continuous shape adjustable trigonometric polynomial curve and surface

Yan Lanlan, Han Xuli 0427

Improved algorithm for natural image pencil drawing production

Huang Zhiyong, Yao Yuan, Zhang Jingjing, Ma Kai, Ren Dong 0437

Online 3D city visualization for smarter cities

Pan Bin, Guo Xiaoming, Chen Mingming, Chen Wei, Peng Qunsheng 0445

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2015)03-0320-12

论文引用格式: Li J F, Zhang F Y. Human behavior recognition based on directional weighting local space-time features[J]. Journal of Image and Graphics, 2015, 20(3): 0320-0331. [李俊峰, 张飞燕. 基于局部时空特征方向加权的人体行为识别[J]. 中国图象图形学报, 2015, 20(3): 0320-0331.] [DOI:10.11834/jig.20150303]

基于局部时空特征方向加权的人体行为识别

李俊峰, 张飞燕

浙江理工大学自动化研究所, 杭州 310012

摘要: 目的 对人体行为的描述是行为识别中的关键问题, 为了能够充分利用训练数据从而保证特征对行为的高描述性, 提出了基于局部时空特征方向加权的人体行为识别方法。方法 首先, 将局部时空特征的亮度梯度特征分解为3个方向(X 、 Y 、 Z)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3方向特征描述子集合的标准视觉词汇码本, 并利用训练视频得到每个行为的标准3方向词汇分布; 进而, 根据不同行为3方向特征描述子集合的标准视觉词汇码本, 分别计算测试视频相应的3方向的词汇分布, 并利用与各行为标准3方向词汇分布的加权相似性度量进行行为识别; 结果 在Weizmann数据库和KTH数据库中进行实验, Weizmann数据库中的平均识别率高达96.04%, KTH数据库中的平均识别率也高达96.93%。结论 与其他行为识别方法相比可以明显提高行为平均识别率。

关键词: 行为识别; 局部时空特征; 视觉词汇表; 方向加权

Human behavior recognition based on directional weighting local space-time features

Li Junfeng, Zhang Feiyan

Institute of Automation, Zhejiang Sci-Tech University, Hangzhou 310012, China

Abstract: **Objective** Human action recognition aims to detect and analyze human behavior intelligently on the basis of information captured by cameras. The applications of this technology include surveillance, video content retrieval robotics, and human-computer interfaces. Human behavior description is a key problem of behavior recognition. To utilize training data fully and to ensure a highly descriptive feature descriptor of behavior, a new human activity recognition method is proposed in this study. **Method** First, the brightness gradient was decomposed into three directions (X , Y , Z) to describe the behavior from different perspectives. Second, the standard visual vocabulary codebooks of the three directions for different behaviors could be obtained by directly constructing a visual vocabulary. Moreover, the standard visual vocabulary codebooks of the three directions for each behavior serve as bases to calculate the corresponding vocabulary distributions of the test video separately. The behavior of the test video might be recognized by using the weighted similarity measure between the standard vocabulary distribution of each behavior and the vocabulary distribution of the test video. **Result** The performance was investigated in the KTH and Weizmann action datasets. We obtained an average recognition rate of 96.04% accu-

收稿日期: 2014-08-20; 修回日期: 2014-11-24

基金项目: 国家自然科学基金项目(61374022, 61203177); 浙江省公益性技术应用研究计划项目(2014C33109); 浙江省新型网络标准及应用技术重点实验室开放课题(2013E10012)

第一作者简介: 李俊峰(1978—), 男, 副教授, 2010年于东华大学获机械工程专业博士学位, 主要研究方向为人体行为识别、图像质量评价等。E-mail: ljf2003@zstu.edu.cn

Supported by: National Natural Science Foundation of China (61374022, 61203177)

racy in the Weizmann action dataset and 96.93% accuracy in the KTH action dataset. **Conclusion** Our method could generate a comprehensive and effective representation of action videos. Furthermore, this approach can reduce clustering time by producing the codebooks of each direction. Experimental results show that the proposed method significantly improves action recognition performance and is superior to all available identification methods.

Key words: behavior recognition; local space-time features; visual vocabulary; directional weighting

0 引言

在过去的十几年里,人体行为检测、动态跟踪、场景建模和行为理解(人体行为识别)是计算机视觉和机器学习的热点,受到越来越多的研究者的关注^[1]。人的行为识别属于人体运动分析中的高级处理部分,它对包含运动目标的视频序列进行目标的检测、提取、分类和跟踪,最终对其中运动人体目标的行为进行理解和识别^[2]。

目前,人体行为识别主要分为3大类:时空分析法、序列分析法和层次化分析法。时空分析法是把视频图像按时间轴排列组成一个3维时空体,提取其中的局部时空立方体后再进行特征描述,最后实现行为识别。Laptev^[3]检验了时空兴趣点方法在简单行为的识别中的可靠性。Li^[4]提出一种3D加速鲁棒特征(3D SURF)描述器来详细阐述和描述时空兴趣点周围区域。Wang^[5]通过探索视频中不同时空视觉单词之间的相关性来对人体行为进行分类。Burghouts^[6]考虑了行为的时空布局,采用一个时空特征管道来描述行为,然后利用随机森林法把特征量化成直方图,利用SVM分类器进行行为分类。序列分析法是从训练数据里得到模版行为序列,然后比对测试行为序列与模版行为序列来进行行为识别,Chen^[7]首先提取每个视频时空兴趣点的3D-SIFT描述符,然后利用词袋来表示这些视频的行为,最后利用贝叶斯网络来识别人体的行为。层次化分析法是将一个复杂行为分解为多个简单易识别的子行为,然后通过组合这些子行为来识别复杂行为。Alazrai^[8]提出基于联结的两层动作识别系统(CHARS),第1步在基于运动存在的基础上利用KNN分类法将人体行为分为两个层次,第2层包含多个二层递归神经网络来区分各个子层与层之间的关系。Khan^[9]提出了一个分层人类活动识别系统(HAR系统)来识别老年人日常生活中的异常行为。

对人体行为的描述是行为识别中的首要问题。

Li^[10]提出从深度图像序列中提取人体姿态的3D关节节点来描述人体行为。Wang^[11]采用密集的运动轨迹和运动边界来描述人体行为,提取轨迹来描述形状特征(点坐标)、外观(梯度方向直方图)和运动(光流直方图)。文献[12-15]都是通过兴趣点检测来提取时空特征,其中Zhang^[12]利用方差滤波的方法来提高时空特征的行为识别精度;Dollar^[13]提取特征后对特征进行亮度、梯度、光流描述后,利用PCA^[16]对描述子进行降维,最后采用1-NN法进行行为分类识别;Zhai^[14]对特征描述子降维后利用局部时空兴趣点特征词包来表示视频序列中一个独立事件中的人体行为,然后结合支持向量机来识别事件;Lyu^[15]在提取时空特征后对其进行亮度梯度描述,在描述子降维后采用词袋模型生成视觉词汇表,最后采用最近邻分类器进行行为分类识别。

上述方法对时空特征描述子集合进行PCA降维后再构造词袋码本,即对训练数据进行抽样后再进行聚类计算生成“词典”,这种方法使得训练样本不能被充分利用;而且为了保证一定的平均识别率,即使降维,样本数据量仍然过高,聚类速度比较慢。此外,各方向的特征数据可能存在一定的相似性,所有方向一起聚类就会降低不同方向特征对行为的描述性。为了既能充分利用训练数据和保证特征描述子对行为的描述性,又使聚类速度不受影响,本文提出基于局部时空特征方向加权的人体行为识别方法。把时空特征的亮度梯度描述子集合分 X 、 Y 、 Z 3个方向(下文简称3方向)分别来描述行为,通过直接构造视觉词汇表^[17]分别得到不同行为3方向特征描述子集合的标准视觉词汇码本,并利用训练视频和测试视频在标准视觉词汇码本上的词汇分布之间的加权相似性度量进行行为识别。该方法在Weizmann数据库^[18]中10种行为的平均识别率高达96.04%,在KTH数据库^[19]中6种行为的平均识别率也高达96.93%。

1 人体行为描述

1.1 时空兴趣点的检测

局部特征点^[20] (下文称兴趣点) 是计算机视觉和数字图像处理研究领域常用的一种特征。国内外很多学者直接在整幅图像上寻找兴趣点作为特征, 方法为对 2 维空间内的检测图像 $I(x, y)$ 定义一个响应函数 R , 然后计算响应函数的极值, 一般在极值处出现兴趣点, 其中最常见兴趣点检测方法是 Harris 角点检测^[21]。时空域上兴趣点的检测与空间域上很相似, 可以把视频看成是多幅图像组成的图像序列 $I(x, y, t)$ 。

为了可以有效检测出图像序列 $I(x, y, t)$ 的时空兴趣点, 采用 Dollar^[13] 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, 响应函数为

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

式中, I 为视频图像, $g = g(x, y, \sigma)$ 为 2 维高斯平滑核, h_{ev} 和 h_{od} 是在空间域上正交的 1 维 Gabor 滤

波器。

h_{ev} 和 h_{od} 的定义为

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

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

式中, $\omega = \frac{4}{\tau}$, σ 和 τ 分别为空间域和时间域上的检测尺度, 本文中取 $\sigma = 2$ 和 $\tau = 3$ 。

响应函数 R 的每一个极大值点的邻近区域包含了 $I(x, y, t)$ 中的人体局部运动信息。

1.2 特征的提取和描述

对图像序列进行兴趣点检测后, 就可以得到一系列时空兴趣点, 但只靠这些兴趣点无法有效描述人体行为。本文以时空兴趣点为中心定义一个时空立方体, 抽取这个时空立方体的像素来构造时空特征, 立方体的边长取其空间尺度的 6 倍。这种时空立方体包含了有利于响应函数 R 取极大值的大多数点, 图 1 为 Weizmann 数据库中弯腰 (bend)、原地跳步 (jack)、单手挥舞 (wave1)、双手挥舞 (wave2)、横向步行 (side)、单腿跳行 (skip)、跑 (run)、步行 (walk)、原地纵跳 (pjump)、跳步前行 (jump) 等 10 种行为的时空立方体的视频帧。

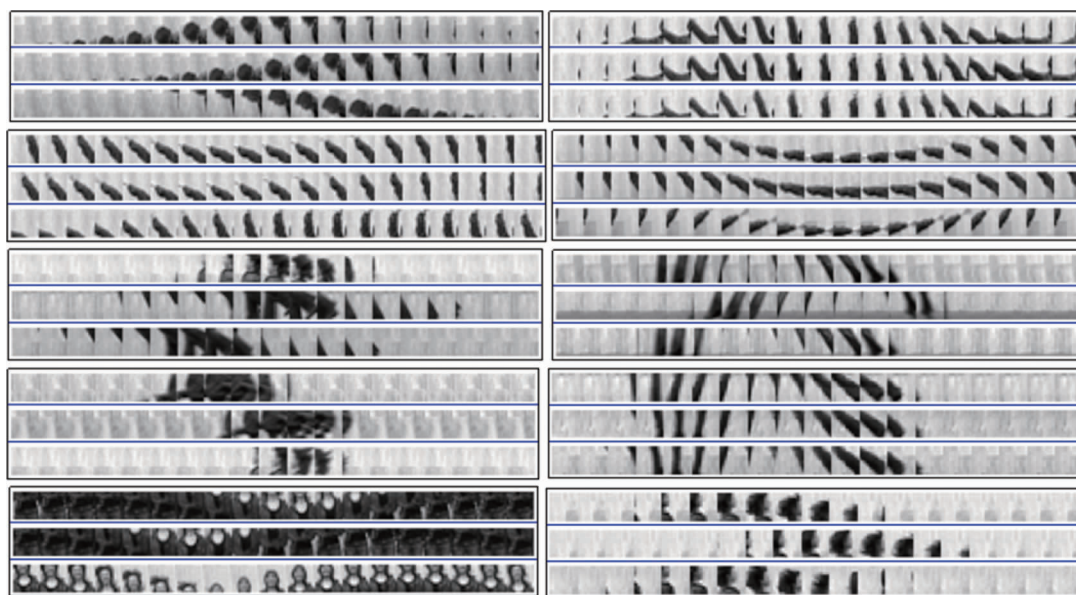


图 1 10 种行为的时空立方体的视频帧

Fig. 1 The frames from the spatial-temporal cube for each behavior

描述时空立方体的方法有将立方体展开值作为一个向量来进行描述、像素归一化描述和直方图描述等。由于人体在运动时, 兴趣点附近区域的图像

亮度变化非常剧烈, 而且人体的运动行为不同时, 兴趣点附近区域的图像亮度变化也不同。因此, 可以利用兴趣点附近区域的图像亮度变化来描述不同人

体行为的兴趣点。不同人体行为的兴趣点附近区域的图像亮度变化可以通过时空立方体的亮度沿 X 轴(水平方向)、 Y 轴(竖直方向)、 Z 轴(时间轴)方向的梯度来反映,本文提取这些梯度作为特征进行人体行为识别。

设时空立方体为 $I(x, y, t)$, 则其 X 轴、 Y 轴、 Z 轴方向的梯度 G_x 、 G_y 、 G_z 可分别定义为

$$G_x(x, y, t) = L(x+1, y, t) - L(x-1, y, t) \quad (4)$$

$$G_y(x, y, t) = L(x, y+1, t) - L(x, y-1, t) \quad (5)$$

$$G_z(x, y, t) = L(x, y, t+1) - L(x, y, t-1) \quad (6)$$

式中, $L(x+1, y, t)$ 为该点经过卷积滤波后的高斯函数的梯度。

2 基于局部时空特征方向加权的人体行为识别

2.1 3 方向视觉词汇表构造

时空立方体的亮度沿 X 轴、 Y 轴、 Z 轴方向的梯度可以作为特征进行人体行为识别。但是,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等方面也存在差异,所以每一帧的兴趣点也不同。因此,同一种行为的不同视频所提取的时空立方体数量是有差异的,相应时空立方体的亮度沿 X 轴、 Y 轴、 Z 轴方向的梯度数量也是不同的。所以,国内外学者经常采用视觉词汇表(bag of

words)^[22]来描述和识别人体行为。但是,目前基于时空特征构造视觉词汇表的人体行为识别方法均把视频 X 、 Y 、 Z 方向的梯度特征描述子集合综合起来进行聚类,如文献[13-15]等,而视频 3 方向的梯度特征描述子集合具有一定的相似性,体现在 3 方向的部分梯度特征描述子数值比较接近,综合聚类就会降低不同方向梯度特征对行为的描述性。例如原地纵跳和步行两种行为,原地纵跳主要是 Y 轴方向运动, Y 轴方向梯度特征变化显著;而步行主要为 X 轴方向运动,相应的 X 轴方向梯度特征变化显著,如果把所有方向特征综合聚类,这两种行为在识别过程中可能会出现一定程度的误判。产生该现象的主要原因是没有考虑不同行为的方向性。

为了综合考虑行为动作的幅度和方向,保证特征描述子集合对行为描述的完整性和差异性,提出如图2所示的视觉词汇表构造方法。分别对视频 3 个方向的梯度特征描述子集合进行聚类,得到 3 个方向的码本;然后计算每个方向特征描述子集合在该方向码本的码字上出现的概率,得到 3 方向直方图分布。具体构造过程如下:

1) 对视频进行预处理并检测该视频的时空兴趣点,以兴趣点为中心提取它的时空立方体;计算其 X 、 Y 、 Z 3 个方向的亮度梯度 G_x 、 G_y 、 G_z , 分别作为 3 个方向上的特征描述子集合。

2) 利用聚类算法对上述 3 个方向上的特征描述子集合 G_x 、 G_y 、 G_z 分别进行聚类,每个方向均得到 k 个聚类中心(码字),各方向的 k 个码字就形成了相应方向上的码本 V_x 、 V_y 、 V_z 。

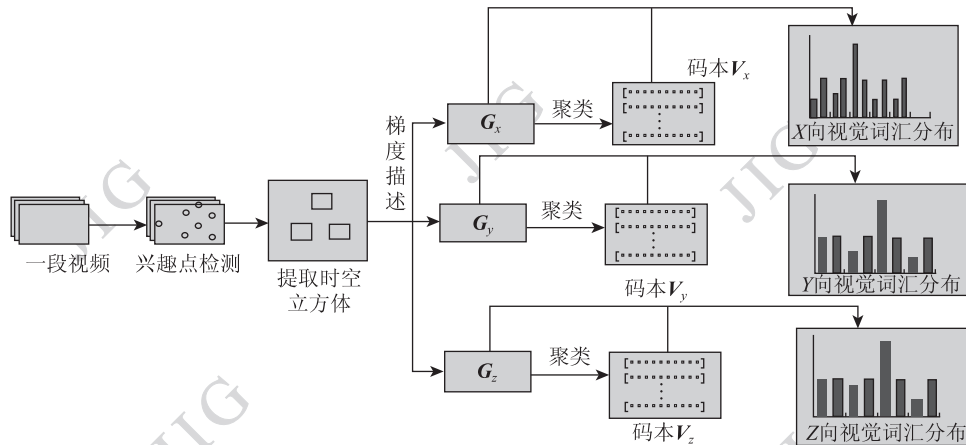


图2 词汇表构造流程

Fig. 2 The flow chart of constructing vocabulary

3) 利用最小距离法计算该视频的 X 、 Y 、 Z 方向特征描述子集合 G_x 、 G_y 、 G_z 在相应方向的码本 V_x 、 V_y 、 V_z 码字上出现的频率, 分别形成 3 方向直方图分布, 利用它们来表征这段视频中的人体运动行为。

设 $V_x = \{c_{x1}, c_{x2}, \dots, c_{xk}\}$ 为视频 X 方向的码本, G_x 为该方向的特征描述子集合, 则 G_{xi} (G_{xi} 表示 G_x 中第 i 个特征描述子, $i = 1, 2, \dots, n$) 到 V_x 中第 j 个码字的距离表示为

$$d_{xij} = \|G_{xi} - c_{xj}\|^2, j = 1, 2, \dots, k \quad (7)$$

G_{xi} 到 V_x 中所有码字的距离可表示为 $d_{xi1}, d_{xi2}, \dots, d_{xik}$, 取这 k 个距离中的最小值 $d_{\min} = \min(d_{xi1}, d_{xi2}, \dots, d_{xik})$, 如果 $d_{\min} = d_{xij}$, 则特征描述子 G_{xi} 属于 V_x 中第 j 类码字。

2.2 不同行为 3 方向视觉词汇分布分析

根据上文 3 方向视觉词汇表构造方法, 选取

Weizmann 数据库中 10 个行为的视频, 分别构造出这 10 个行为的 3 方向直方图分布, 图 3—图 5 为 bend、run、wave1、wave2 等部分行为的 3 方向直方图分布示例图。

根据对 10 个行为的 3 方向直方图分布的研究发现, bend 行为在 3 个方向的直方图分布与其他行为的直方图分布的区分度都比较要明显, 因为该行为为只在一个方向上运动, 且动作速度相对缓慢; 由于 wave1、wave2 行为在挥手时速度和位移与其他行为的区别比较大, 所以 wave1、wave2 行为在 X 和 Z 方向上的直方图分布与其他行为相比区分度比较明显; run 行为在动作时相对于其他行为运动速度比较快, 故与其他行为相比 Z 轴上直方图分布的区分度比其他两个方向的明显, 这与 Z 方向上的梯度描述子差别比较大有关; jack、pjump 行为在 Y 和 Z 方向

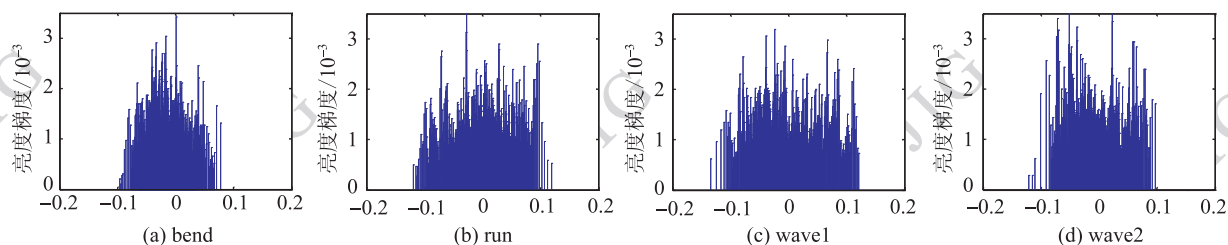


图 3 部分行为的 X 方向直方图分布

Fig. 3 The histogram distributions of X direction for part behavior

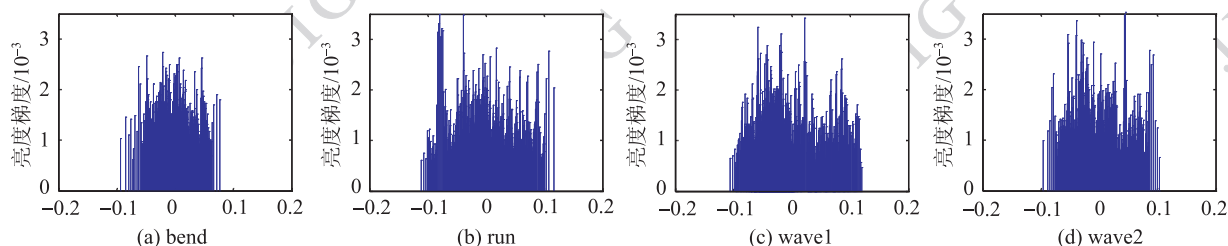


图 4 部分行为的 Y 方向直方图分布

Fig. 4 The histogram distributions of Y direction for part behavior

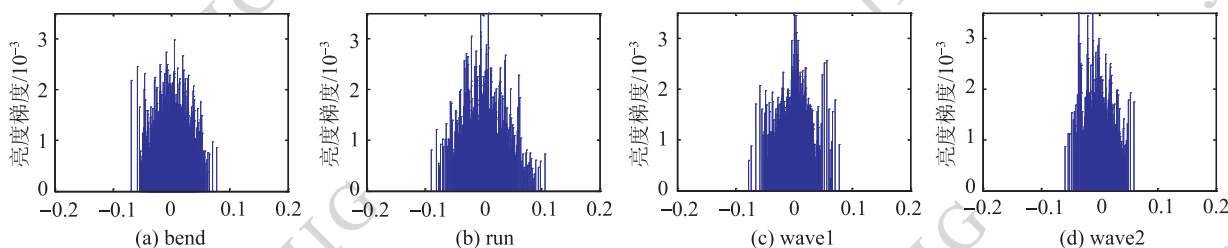


图 5 部分行为的 Z 方向直方图分布

Fig. 5 The histogram distributions of Z direction for part behavior

上的直方图分布与其他行为的直方图分布相比区分度都比较明显,因为这两个行为在动作时在 X 方向上的位移变化都比较小;walk、side 行为在 Y 方向上的直方图分布与其他行为的直方图分布相比差异比较大,因为这两个行为在动作时在身高方向上的位移变化都比较小;此外,一些在某个动作上比较相似的行为,它们在相应方向的直方图的形状比较相似,如同样有双手舞动的 jack、wave2 行为在 Y 方向上的直方图分布比较相似,因为它们在动作时 Y 方向上的位移变化差别小,所以 Y 方向的梯度描述子比较相似,但它们在 X 和 Z 方向上的直方图分布区分度比较大;由于 pjump、skip 行为在动作时的速度相近,从而 Z 方向的梯度描述子在数值上区分比较小,所以这两个行为在 Z 方向上的直方图分布形状相似,但它们在 X 和 Y 方向上的直方图分布区分度比较大,其他动作上比较相似的行为,也存在类似的情况。

总而言之,在 3 方向直方图分布中,不同的行为至少在一个方向上分布是不同的,产生上述现象的主要原因是不同行为的时空立方体亮度变化不同,相应的 3 个方向的梯度变化不同时相同。因此,认为如果结合视频的时空立方体 3 个方向直方图分布进行人体行为识别,会大大提高人体行为识别率。

2.3 不同行为 3 方向标准直方图分布构造

上文中的 10 种行为都是同一个人的行为,然而不同人的同一种行为在动作上是有差异的,这会影响到该行为时空立方体的 3 方向直方图分布,产生差异的主要原因是人的身高、着装及其在动作时肢体的动作空间范围和动作速度不同。如果直接利用同一个人 10 种行为相应的 3 方向直方图分布作为标准进行人体行为识别,识别率比较低。为了解决这个问题,提出构造 10 个行为的 3 方向标准直方图分布,图 6 以 Weizmann 数据库中 run 行为为例,给出了该行为 3 方向标准直方图分布的构造原理,具体构造过程为:

1) 计算 n 个 run 行为视频的 3 方向特征描述子集合 G_{rx} 、 G_{ry} 、 G_{rz} ,并分别把它们进行合并得到代表 run 行为的 3 方向特征描述子总集合 $\Omega_{G_{rx}}$ 、 $\Omega_{G_{ry}}$ 、 $\Omega_{G_{rz}}$,即

$$\Omega_{G_{rx}} = G_{rx1} \cup G_{rx2} \cup \dots \cup G_{rxn} \quad (8)$$

$$\Omega_{G_{ry}} = G_{ry1} \cup G_{ry2} \cup \dots \cup G_{ryn} \quad (9)$$

$$\Omega_{G_{rz}} = G_{rz1} \cup G_{rz2} \cup \dots \cup G_{rzn} \quad (10)$$

2) 对 3 方向特征描述子总集合分别聚类得到 3 方向标准码本 M_{rx} 、 M_{ry} 、 M_{rz} ,并利用最小距离法分别计 $\Omega_{G_{rx}}$ 、 $\Omega_{G_{ry}}$ 、 $\Omega_{G_{rz}}$ 在所相应方向的标准码本 M_{rx} 、 M_{ry} 、 M_{rz} 码字上出现的频率,分别形成 run 行为的 3 方向标准直方图分布,利用它们来表征 run 的标准行为。

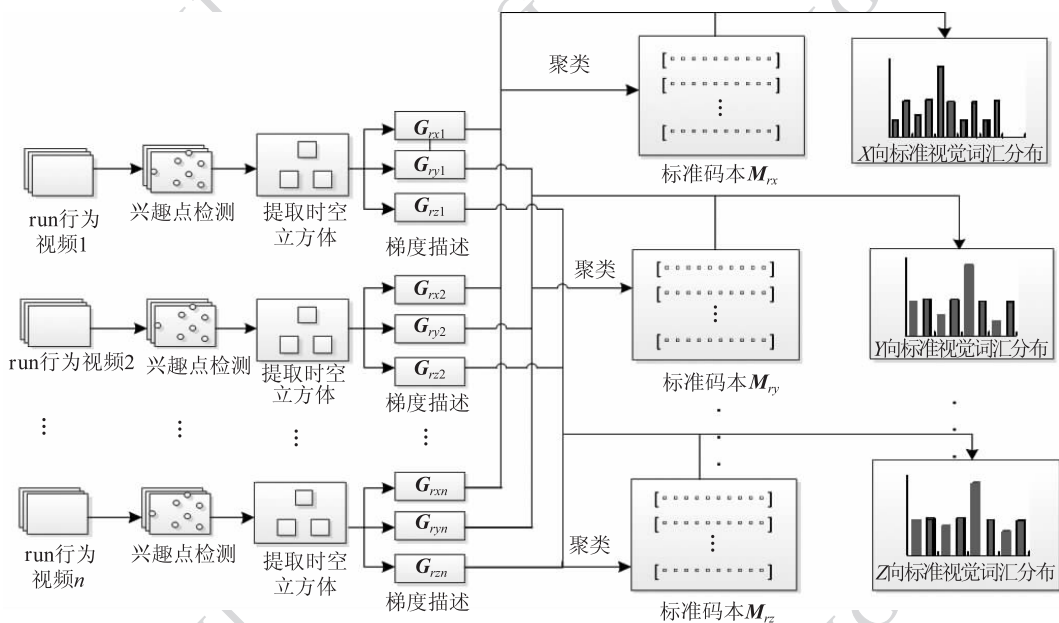


图 6 run 行为 3 方向标准直方图分布构造原理

Fig. 6 The construction principle of three directional standard histogram distribution for run action

根据 run 行为 3 方向标准直方图分布的构造方法,利用 Weizmann 数据库中相同行为的所有视频(daria 的行为除外,daria 的行为作为测试视频)来

构造相应行为的 3 方向标准直方图分布。图 7—图 9 为所构造的 bend、run、wave1、wave2 等部分行为的 3 方向标准直方图分布。

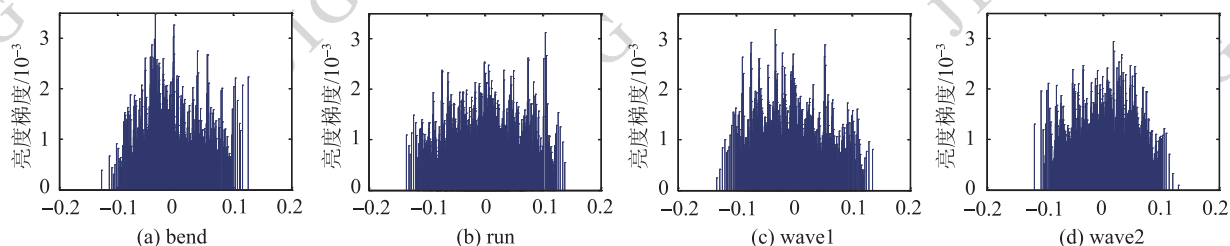


图 7 部分行为的 X 方向标准直方图分布

Fig. 7 The standard histogram distributions of X direction for part behavior

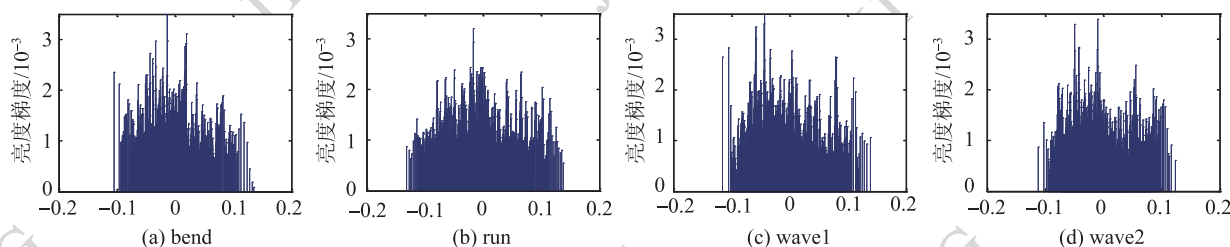


图 8 部分行为的 Y 方向标准直方图分布

Fig. 8 The standard histogram distributions of Y direction for part behavior

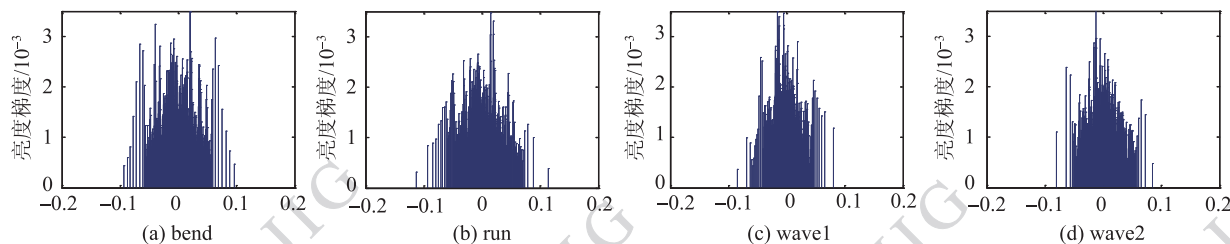


图 9 部分行为的 Z 方向标准直方图分布

Fig. 9 The standard histogram distributions of Z direction for part behavior

2.4 基于局部时空特征方向加权的人体行为识别

根据上述分析,不同行为的 3 方向直方图分布存在差异。即使相似的行为,在 3 方向直方图分布中,至少在一个方向分布是不同的。如果可以结合视频的时空立方体 3 方向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,本文提出基于局部时空特征方向加权的人体行为识别方法,识别原理如图 10 所示,具体步骤如下:

1) 对待识别视频进行兴趣点检测,提取时空立方体并对其进行 3 方向的亮度梯度描述得到相应方向特征描述子集合;分别计算待识别视频 3 方向特征描述子集合在各种行为的 3 方向标准码本上的直

方图分布,得到其 3 方向直方图分布。

2) 分别计算待识别视频的 3 方向直方图分布与各行为标准 3 方向直方图分布的距离 d_x 、 d_y 、 d_z ,为了使不同行为区分更为明显,需要对 3 方向距离 d_x 、 d_y 、 d_z 进行加权得到加权距离 D_{ist} 。加权距离

$$D_{ist} = \omega_1 \times d_x + \omega_2 \times d_y + \omega_3 \times d_z \quad (11)$$

式中, ω_1 、 ω_2 、 ω_3 为加权系数, $\omega_1 + \omega_2 + \omega_3 = 1$,且 $\omega_1, \omega_2, \omega_3 \in [0, 1]$ 。

各方向距离的加权系数 ω_1 、 ω_2 、 ω_3 可以根据各行为标准 3 方向词汇分布对行为表示的贡献来优化设计。

3) 根据计算出的加权距离,利用最近邻分类器对行为进行识别。

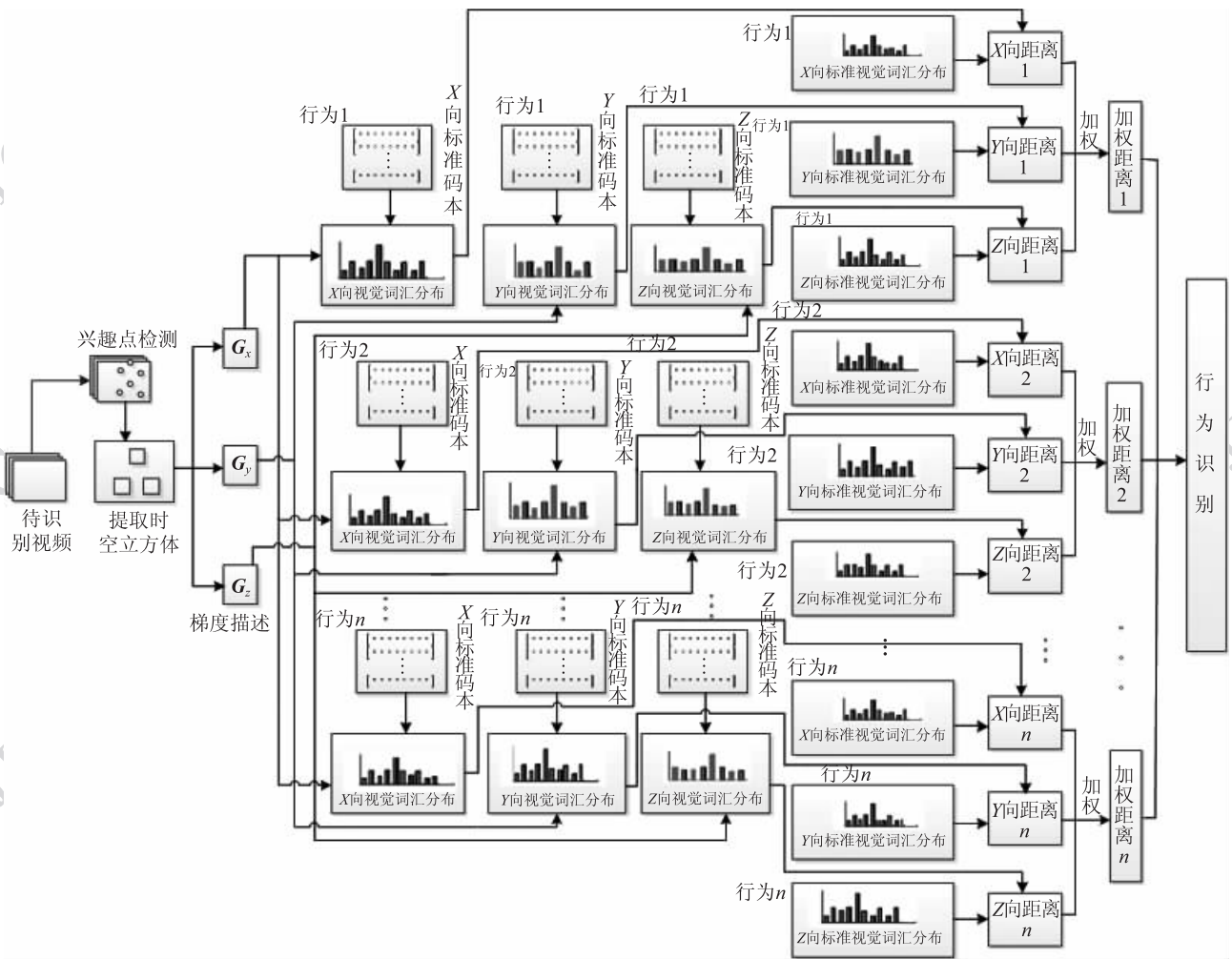


图10 识别原理图

Fig. 10 The principle of recognition

3 实验结果及分析

为了验证本文人体行为识别方法的性能,在 Weizmann Action Dataset 和 KTH Action Dataset 两个常用的标准数据库上进行测试。此外,为了保证测试的公平性,根据数据库特点并结合参考文献常用的实验分组方法,采用两种不同实验方法对两个数据库分别进行测试,测试结果以混淆矩阵的形式给出各行为的平均识别率。

3.1 Weizmann 行为数据库

Weizmann 数据库包含由 9 个不同的人的 90 个低分辨率视频(像素为 180×140)。每个人的 10 种不同的行为: bend、jack、jump、pjump、run、side、skip、walk、wave1 及 wave2,如图 11 所示。

由于 Weizmann 数据库的视频比较少,实验方

法采用文献[4,24-26]中的 leave-one-out 实验分组方法,即选取其中一个人的 10 个行为视频作为测试样本,其余 8 个人的行为视频作为训练样本。采用 Niebles^[23]中的“词袋模型”来构造 3 方向视觉词汇表,其中每个方向的词汇数目设为 700。最后利用基于最小距离法的最近邻分类器进行行为分类(本文实验采用的是欧氏距离)。

由于不同人在做同一类行为时具有差异,且一些相似行为比较难以准确区分,为了解决这些问题,提高行为识别率,通过优化设计了多组加权系数组合进行行为识别。由于多组加权系数最后可以得到多组加权距离,因此在识别过程中选取加权距离最小的那组来对行为进行分类。表 1 为 Weizmann 数据库中 10 个行为的 3 向加权系数表,从表 1 中可以看出多数行为只需两个方向的权值组合便可以进行有效识别。



图 11 Weizmann 数据库视频单帧示例

Fig. 11 Frames of video in Weizmann action dataset

表 1 Weizmann 数据库中的 3 方向的加权系数表

Table 1 The weighted coefficients table of three direction in Weizmann action dataset

行为	X	Y	Z
bend	0.6	0.2	0.2
jack	0.3	0.6	0.1
jump	0.9	0.1	0
pjump	0.8	0.1	0.1
run	0	0.9	0.1
side	0.7	0.2	0.1
skip	0.8	0.2	0
walk	0	0.3	0.7
wave1	0	0.7	0.3
wave2	0.3	0	0.7

图 12 是以混淆矩阵的形式表示的 Weizmann 数据库中的行为识别率,其中第 i 行第 j 列的数值表示第 i 个行为被判为第 j 个行为的概率,同一行的所有数值之和为 1。由混淆矩阵可以看出,run 和 side 行为的识别率相对较低,run 行为主要被误判为 jump 和 side,而 side 行为主要被误判为 jump,主要原因是这 3 种行为在 3 方向都有相似的动作,只不

过个别方向的动作速度和幅度稍有不同。此外, jack 行为被误判为 wave2 的概率也高达 0.036,主要原因是原地跳步 jack 在动作过程中也有类似双手挥舞 wave2 的行为。

bend	0.988	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000
jack	0.000	0.940	0.012	0.000	0.000	0.000	0.012	0.000	0.000	0.000	0.036
jump	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pjump	0.000	0.024	0.000	0.952	0.000	0.000	0.000	0.024	0.000	0.000	0.000
run	0.000	0.012	0.036	0.000	0.892	0.024	0.012	0.012	0.012	0.000	0.000
side	0.000	0.012	0.036	0.000	0.012	0.928	0.000	0.012	0.000	0.000	0.000
skip	0.000	0.000	0.012	0.000	0.000	0.012	0.964	0.000	0.012	0.000	0.000
walk	0.000	0.012	0.000	0.000	0.012	0.000	0.000	0.976	0.000	0.000	0.000
wave1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
wave2	0.000	0.012	0.012	0.000	0.000	0.000	0.000	0.000	0.012	0.964	0.000
bend	jack	jump	pjump	run	side	skip	walk	wave1	wave2		

图 12 Weizmann 数据库的混淆矩阵

Fig. 12 The mean confusion matrix in Weizmann action dataset

表 2 为本文及其他文献识别方法在 Weizmann 数据库中的平均识别率。从表 2 中可以看出,与其他文献的识别方法相比,本文方法可以明显提高行

为识别率,平均识别准确率高达 96.04%。

表2 Weizmann 数据库中不同文献的平均识别率

Table 2 The average recognition rate of different articles in Weizmann action dataset

行为识别方法	发表年份	平均识别率/%
本文	2015	96.04
文献[4]	2013	90
文献[24]	2013	85.56
文献[25]	2013	94.44
文献[26]	2013	95.1

3.2 KTH 行为数据库

KTH 数据库(如图 13 所示)包含由 25 个人的 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(clap 行为 99 个),像素为 120×160 。实验采用文献[3,5,8,22]的实验分组方法,将所有视频分为测试集(Person 2,3,5,6,7,8,9,10,22 的视频)和训练集(其余人的视频)。采用 Niebles^[23]中的“词袋模型”来构造 3 方向视觉词汇表,其中每个方向的词汇数目设为 500(本文实验采用的是欧氏距离)。

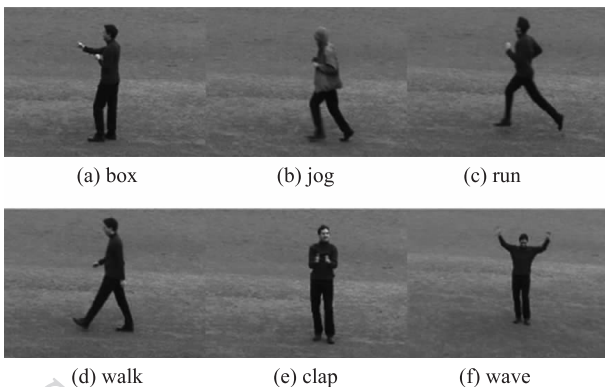


图 13 KTH 数据库视频单帧示例

Fig. 13 Frames of video in KTH action dataset

表 3 是 KTH 数据库中 6 个行为的 3 向加权系数表,图 14 是以混淆矩阵的形式表示的 KTH 数据库中的行为识别率,其中 box 行为的识别率相对较低,box 行为被误判为 walk、clap 和 wave,主要原因是 box 行为与这 3 种行为手部在 X、Y 两个方向上都有相似的动作,只不过动作幅度稍有不同。此外,run 和 wave 行为的误判率也分别高达 0.046,run 行为被误判为 jog,原因是该行为与 jog 行为都是跑步,

只是肢体动作速度和幅度不同;而 wave 行为被误判为 clap,主要原因是双手挥舞 wave 行为和鼓掌 clap 行为都存在相似的手部动作,只不过 clap 行为的动作幅度比 wave 行为小一点。在 KTH 数据库中,本文的平均识别准确率为 96.93%。

表3 KTH 数据库中的 3 方向的加权系数表

Table 3 The weighted coefficients table of three direction in KTH action dataset

行为	X	Y	Z
box	0.1	0.9	0
jog	0.3	0.1	0.6
run	0	0.9	0.1
walk	0	0.3	0.7
clap	0.1	0.2	0.7
wave	0	0.1	0.9

box	0.931	0.000	0.000	0.023	0.023	0.023
jog	0.000	1.000	0.000	0.000	0.000	0.000
run	0.000	0.046	0.954	0.000	0.000	0.000
walk	0.000	0.023	0.000	0.977	0.000	0.000
clap	0.000	0.000	0.000	0.000	1.000	0.000
wave	0.000	0.000	0.000	0.000	0.046	0.954
	box	jog	run	walk	clap	wave

图 14 KTH 数据库的混淆矩阵

Fig. 14 The mean confusion matrix in KTH action dataset

表 4 是本文与其他文献识别方法在 KTH 数据库上平均识别率的比较表。从表 4 中可以看出,与其他文献的方法相比,本文方法仍然可以明显提高行为识别率,平均识别率高达 96.93%。

表4 KTH 数据库中不同文献的平均识别率

Table 4 The average recognition rate of different articles in KTH action dataset

行为识别方法	发表年份	平均识别率/%
本文	2015	96.93
文献[3]	2008	91.8
文献[5]	2012	79.8
文献[8]	2012	95.55
文献[22]	2013	93

4 结 论

为了提高局部时空特征对行为描述的完整性及突出不同方向特征对行为描述的差异性,提出将局部时空特征梯度描述子分解为3方向分别来描述行为,并根据不同行为的动作特点对3方向直方图加权进行人体行为识别。该识别方法不仅可以保持3方向梯度特征描述子集合对行为描述的完整性,又可以兼顾不同行为3方向梯度特征描述子集合对行为描述的差异性;既可以有效提高行为识别正确率,又可以通过降低所需聚类的特征数据量而提高识别的效率。实验结果表明,该方法在 Weizmann 数据库平均识别率高达 96.04%;在 KTH 数据库平均识别率也高达 96.93%。此外,在构造视觉词汇表时,没有考虑不同方向词汇数目差异,3方向词汇表中词汇数目选取均相同,(Weizmann 数据库为 700, KTH 数据库为 500),而且各方向词汇数目也没有进行优化。下一步的研究重点是改进“词袋模型”的词汇分类方法及优化设计各方向词汇数目,以进一步提高行为平均识别率和行为识别效率。

参考文献 (References)

- [1] Popoola O P, Wang K. Video-based abnormal human behavior recognition—a review[J]. IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, 2012, 42(6): 865-878. [DOI:10.1109/TSMCC.2011.2178594]
- [2] Hu S S. Research and implementation of human abnormal behavior identification on intelligent monitoring system[D]. Xi'an: Xi'an University of Electronic Science and Technology, 2012. [胡韶山. 人体异常行为识别在智能监控系统中的研究与实现[D]. 西安: 西安电子科技大学, 2012.] [DOI:10.7666/d.d216693]
- [3] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]// Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, United States: IEEE Computer Society, 2008: 1-8. [DOI:10.1109/CVPR.2008.4587756]
- [4] Li C Z, Su B L, L Y, et al. Human action recognition using spatio-temporal descriptor[C]// Proceedings of the 6th International Congress on Image and Signal Processing. Hangzhou, China: IEEE Computer Society, 2013: 107-111. [DOI:10.1109/CISP.2013.6743966]
- [5] Wang S, Yang Y, Ma Z, et al. Action recognition by exploring data distribution and feature correlation[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Providence, RI, United States: IEEE Computer Society, 2012: 1370-1377. [DOI: 10.1109/CVPR.2012.6247823]
- [6] Burghouts G J, Schutte K. Spatio-temporal layout of human actions for improved bag-of-words action detection[J]. Pattern Recognition Letters, 2013, 34(15): 1861-1869. [DOI: 10.1016/j.patrec.2013.01.024]
- [7] Chen W Q, Xiao G Q, Tang X Q. An action recognition model based on the bayesian networks[J]. Applied Mechanics and Materials, 2014, 513: 1886-1889. [DOI:10.4028/www.scientific.net/AMM.513-517.1886]
- [8] Alazrai R, Lee C S G. A connectionist-based approach for human action identification[C]// Proceedings of IEEE International Conference on Robotics and Automation. RiverCentre, Saint Paul, Minnesota, USA: Institute of Electrical and Electronics Engineers Inc., 2012: 1212-1217. [DOI: 10.1109/ICRA.2012.6224702]
- [9] Khan Z A, Sohn W. A hierarchical abnormal human activity recognition system based on R-transform and kernel discriminant analysis for elderly health care[J]. Computing, 2013, 95(2): 109-127. [DOI:10.1007/s00607-012-0216-x]
- [10] Li W, Zhang Z, Liu Z. Action recognition based on a bag of 3d points[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops. San Francisco, CA, United States: IEEE Computer Society, 2010: 9-14. [DOI:10.1109/CVPRW.2010.5543273]
- [11] Wang H, Klaser A, Schmid C, et al. Dense trajectories and motion boundary descriptors for action recognition[J]. International journal of Computer Vision, 2013, 103(1): 60-79. [DOI: 10.1007/s11263-012-0594-8]
- [12] Zhang H B, Li S A, Wang L Y, et al. Human action recognition with spatial-temporal feature filter[C]// Proceedings of the 6th International Conference on Pervasive Computing and Applications. Port Elizabeth, South Africa: IEEE Computer Society, 2011: 458-461. [DOI:10.1109/ICPCA.2011.6106547]
- [13] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]// Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China: Inst. of Elec. and Elec. Eng. Computer Society, 2005: 65-72. [DOI:10.1109/VSPETS.2005.1570899]
- [14] Zhai C M, Guo Y L, Du J X. Event recognition based on bag of local space-time interest points' features[C]// Proceedings of the 4th International Workshop on Advanced Computational Intelligence. Wuhan, Hubei, China: IEEE Computer Society, 2011: 477-481. [DOI:10.1109/IWACI.2011.6160054]
- [15] Lv W, Xu G L, Chen Y H, et al. Human behavior of soft classi-

- fication based on local space-time features [J]. *Computer and Modernization*, 2014, (3): 94-98. [吕温, 徐贵力, 程月华, 等. 基于局部时空特征的人体行为软分类识别[J]. *计算机与现代化*, 2014, (3): 94-98.] [DOI:10.3969/j.issn.1006-2475.2014.03.023]
- [16] Huang W G, Gu C, Zhu Z K. PCA-SC shape matching for object recognition [J]. *Opt. Precision Eng.*, 2013, 21(8): 2103-2110. [黄伟国, 顾超, 朱忠奎. 用于目标识别的 PCA-SC 形状匹配算法[J]. *光学精密工程*, 2013, 21(8): 2103-2110.] [DOI:10.3788/OPE.20132108.2103]
- [17] 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. [DOI: 10.1016/j.patcog.2013.05.001]
- [18] Blank M, Gorelick L, Shechtman E, et al. Actions as space-time shapes [C]//*Proceedings of the IEEE International Conference on Computer Vision*. Beijing, China; Institute of Electrical and Electronics Engineers Inc., 2005, 2: 1395-1402. [DOI: 10.1109/ICCV.2005.28]
- [19] Schudt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach [C]//*Proceedings of International Conference on Pattern Recognition*. Cambridge, United Kingdom; Institute of Electrical and Electronics Engineers Inc., 2004: 32-36. [DOI:10.1109/ICPR.2004.1334462]
- [20] Fan C X, Zhu H, Lin G P, et al. Person re-identification based on multi-features [J]. *Jouranal of Image and Graphics*, 2013, 18(6): 711-717. [范彩霞, 朱虹, 蔺广逢, 等. 多特征融合的人体目标再识别[J]. *中国图象图形学报*, 2013, 18(6): 711-717.] [DOI:10.11834/jig.20130614]
- [21] Ting J L, Jun Z, Lin L. Speed up junction detector based on azimuth consensus by Harris corner [J]. *Optical Review*, 2014, 21(2): 135-142. [DOI:10.1007/s10043-014-0021-1]
- [22] Shi F, Petriu E, Laganieri R. Sampling strategies for real-time action recognition [C]//*Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Portland, OR, United States; IEEE Computer Society, 2013: 2595-2602. [DOI:10.1109/CVPR.2013.335]
- [23] 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. [DOI: 10.1007/s11263-007-0122-4]
- [24] Tom M, Venkatesh Babu R. Rapid human action recognition in H.264/AVC compressed domain for video surveillance [C]//*Proceedings of IEEE International Conference on Visual Communications and Image Processing*. Kuching, Sarawak, Malaysia: IEEE Computer Society, 2013: 1-6. [DOI: 10.1109/VCIP.2013.6706430]
- [25] Ar I, Akgul Y S. Action recognition using random forest prediction with combined pose-based and motion-based features [C]//*Proceedings of the 8th International Conference on Electrical and Electronics Engineering*. Bursa, Turkey; IEEE Computer Society, 2013: 315-319. [DOI:10.1109/ELECO.2013.6713852]
- [26] Halici U, Gökce O. Mixture of Poses for human behavior understanding [C]//*Proceedings of the 6th International Congress on Image and Signal Processing*. Hangzhou, China; IEEE Computer Society, 2013: 112-116. [DOI:10.1109/CISP.2013.6743968]