

Journal of Image  
and Graphics

# 中国图象图形学报



ISSN1006-8961  
CN11-3758/TB

2012 **6**  
Vol.17 No.

中国科学院遥感应用研究所  
中国图象图形学学会主办  
北京应用物理与计算数学研究所

# 中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2012年6月 第17卷 第6期(总第194期)

## 目次

### 综述

- 人脸图像的年龄估计技术研究 ..... 王先梅, 梁玲燕, 王志良, 胡四泉(603)
- 图像统计模型参数估计中的期望最大值算法 ..... 李旭超(619)

### 图像处理和编码

- 鲁棒的梯度驱动图像修复算法 ..... 叶学义, 王靖, 赵知劲, 陈华华(630)
- 快速去块效应的线性规划方法 ..... 金剑秋, 刘春晓, 王勋, 章志勇(636)
- DWT-SVD 域全盲自嵌入鲁棒量化水印算法 ..... 叶天语(644)
- Mean-shift 滤波和直方图增强的图像弱边缘提取 ..... 冀峰, 高新波, 谢松云(651)

### 图像分析和识别

- 梯度点对约束的结构化车道检测 ..... 王永忠, 王晓云, 文成林(657)
- 多水平外区抑制的轮廓检测 ..... 闫超, 张建州(664)
- 基于局部投影与块 LBP 特征的图像检索 ..... 邹彬, 潘志斌, 胡森(671)

### 图像理解和计算机视觉

- 视觉陌生度驱动的增量自主式视觉学习算法 ..... 瞿心昱, 姚明海, 顾勤龙(678)
- 霍夫空间中多足球机器人协作目标定位算法 ..... 许家铭, 解仑, 王志良, 倪善超(687)

### 计算机图形学

- 虚拟 CCD 线阵星载光学传感器内视场拼接 ..... 张过, 刘斌, 江万寿(696)

“第八届数字电视与无线多媒体通信国际论坛”会议专栏

追踪盗版者的抗共谋数字指纹 ..... 李晓强, 张煌, 赵洋洋, 汪晶晶(702)

LBP 与鉴别模式结合的热红外人脸识别 ..... 谢志华, 伍世虔, 方志军(707)

电影中复杂事件的检索与识别 ..... 杜吉祥, 郭一兰, 翟传敏(712)

方向特征和网格特征融合的离线签名鉴别 ..... 杨丹凤, 吕岳(717)

视觉关注度的立体图像质量评价 ..... 张艳, 安平, 张秋闻, 张兆杨(722)

低开销的异构数据交换 ..... 赵凯, 赵正德(726)

灵活的多视点视频编码预测结构 ..... 张艳, 蔡灿辉(730)

跨层反馈在无线视频自适应编码中的应用 ..... 王垚中, 郑世宝, 张重阳, 刘勃(736)

中国图象图形学报

刊名题字: 宋 健

月刊(1996 年创刊)

第 17 卷 第 6 期

2012 年 6 月 16 日出版

主管单位 中国科学院

主 办 中国科学院遥感应用研究所  
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  
北京 9718 信箱 邮编 100101  
电子信箱:jig@irsa. ac. cn  
电话:010-64807995 010-82614429  
网 址:www. cjjg. cn

印刷装订 北京北林印刷厂

广告经营许可证 京朝工商广字第 0346 号

总 发 行 北京报刊发行局

订 购 全国各地邮局

国外发行 中国国际图书贸易总公司  
(中国国际书店)  
(北京 399 信箱 邮编 100044)

**Superintended by** Chinese Academy of Sciences  
**Sponsored by** Institute of Remote Sensing Application,  
CAS China Society of Image and Graphics  
Institute of Applied Physics and Computational  
Mathematics

**Chief editor** LI Xiaowen  
**Editor, Publisher** Editorial and Publishing Board  
of Journal of Image and Graphics  
(P. O. Box 9718, Beijing 100101, China)  
E-mail:jig@irsa. ac. cn

**Distributed by** Beijing Bureau for Distribution of Newspapers  
and Journals

**Domestic** All Local Post Offices in China  
**Foreign** China International Book Trading Corporation  
(P. O. Box 399, Beijing 100044, China)

**Printed by** Beijing Beilin Printing House

ISSN 1006-8961 CN11-3758/TB CODE ZTTFXZ 国内邮发代号: 82-831 国外发行代号: M1406 国内定价: 45.00 元

# Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 17 No. 6 June 2012

## Contents

### Review

- Age estimation by facial image: a survey ..... Wang Xianmei, Liang Lingyan, Wang Zhiliang, Hu Siqian(603)  
Expectation maximization method for parameter estimation of image statistical model ..... Li Xuchao(619)

### Image Processing and Coding

- Robust gradient driving image inpainting method ..... Ye Xueyi, Wang Jing, Zhao Zhijing, Chen Huahua(630)  
Fast image de-blocking by linear programming ..... Jin Jianqiu, Liu Chunxiao, Wang Xun, Zhang Zhiyong(636)  
Perfectly blind self-embedding robust quantization-based watermarking scheme in DWT-SVD domain ..... Ye Tianyu(644)  
Weak edge detection using Mean-shift filtering and histogram enhancement ..... Ji Feng, Gao Xinbo, Xie Songyun(651)

### Image Analysis and Recognition

- Gradient-pair constraint for structure lane detection ..... Wang Yongzhong, Wang Xiaoyun, Wen Chenglin(657)  
Contour detection based on multilevel inhibition ..... Yan Chao, Zhang Jianzhou(664)  
Image retrieval method based on local projection and block LBP feature ..... Zou Bin, Pan Zhibin, Hu Sen(671)

### Image Understanding and Computer Vision

- Visual novelty driven incremental and autonomous visual learning algorithm ..... Qu Xinyu, Yao Minghai, Gu Qinlong(678)  
Target localization algorithm for cooperative multi-soccer robots based on Hough space .....  
..... Xu Jiaming, Xie Lun, Wang Zhiliang, Ni Shanchao(687)

### Computer Graphics

- Inner FOV stitching algorithm of spaceborne optical sensor based on the virtual CCD line .....  
..... Zhang Guo, Liu Bin, Jiang Wangshou(696)

### Special Issue of the IFIC' 2011

- Anti-collusion fingerprinting scheme capable of tracing pirate ..... Li Xiaoqiang, Zhang Huang, Zhao Yangyang, Wang Jingjing(702)  
Infrared face recognition using LBP and discrimination patterns ..... Xie Zhihua, Wu Shiqian, Fang Zhijun(707)  
Recognize and retrieval complex events in real movies ..... Du Jixiang, Guo Yilan, Zhai Chuanmin(712)  
Off-line signature verification based on combination of direction feature and grid feature ..... Yang Danfeng, Lv Yue(717)  
Stereo image quality assessment based on visual attention ..... Zhang Yan, An Ping, Zhang Qiuwen, Zhang Zhaoyang(722)  
Low overhead of heterogeneous data exchange ..... Zhao Kai, Zhao Zhengde(726)  
Flexible prediction structure for multi-view video coding ..... Zhang Yan, Cai Canhui(730)  
Cross-layer feedback based adaptive coding for wireless video transmission .....  
..... Wang Yaozhong, Zheng Shibao, Zhang Chongyang, Liu Bo(736)

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2012)06-0712-05

论文引用格式: 杜吉祥, 郭一兰, 翟传敏. 电影中复杂事件的检索与识别[J]. 中国图象图形学报, 2012, 17(6): 712-716.

## 电影中复杂事件的检索与识别

杜吉祥, 郭一兰, 翟传敏

华侨大学计算机科学与技术学院, 厦门, 361021

**摘要:** 提出一种基于局部时空兴趣点的电影中复杂事件检索与识别的方法。先将一个独立的事件视频序列表示成一个自组织映射像密度图, 再将此密度图与支持向量机相结合用于识别事件。该方法使用局部时空特征描述子来捕捉视频中的局部事件, 可以适应事件的模式的不同的大小和速度。为了验证该方法的有效性, 使用公开的Hollywood 视频数据库, 其中的镜头序列收集自 32 部不同的 Hollywood 电影, 包含了 8 个事件类别。综合实验, 得到平均正确率、平均查准率和平均查全率分别为 0.601、0.530 和 0.566。实验结果表明本文方法明显提高了平均正确率和平均查准率。

**关键词:** 局部时空兴趣点; 局部时空特征; 自组织特征映射; 事件识别

## Recognize and retrieval complex events in real movies

Du Jixiang, Guo Yilan, Zhai Chuanmin

College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China

**Abstract:** We propose a new method based on local space-time interest points and self-organization feature maps (SOFM) to recognize and retrieval complex events in real movie. In this method, an individual video sequence is represented as a SOFM density map, We integrate this density map with a support vector machine (SVM) to recognize events. Local space-time features are introduced to capture the local events in video and can be adapted to size and velocity of the pattern of the event. To evaluate the effectiveness of this method, we use the public Hollywood dataset. In this dataset shot sequences are collected from 32 different Hollywood movies and it includes eight event classes. According to the experiment, the average accuracy rate, the average precision rate, and average recall rate were 0.601, 0.530 and 0.566 respectively. The presented results justify the proposed method explicitly improving the average accuracy and average precision compared with other relative approaches.

**Key words:** local space-time interest points; local space-time features; self-organization feature map; event recognition

## 0 引言

视频数据的高速增长, 促使我们需要从大量的实际视频数据中挖掘出有用的信息, 或者检索自己

感兴趣的视频等, 这些需求促进了自动视频理解的研究。视频事件经常被定义为吸引用户注意的感兴趣事件。目前的视频事件分析技术主要涉及两个问题<sup>[1]</sup>: 1) 如何提取视频序列中有意义、有助于识别的信息(特征提取及表示); 2) 如何识别感兴趣的事

收稿日期: 2011-09-26; 修回日期: 2012-01-15

**项目基金:** 国家自然科学基金项目(60805021, 61175121); 教育部新世纪优秀人才支持计划项目(NCET-10-0117); 福建省自然科学基金项目(2011J01349); 福建省高等学校杰出青年科研人才培养计划项目(JA10006); 福建省教育厅科技计划项目(JA11004); 华侨大学侨办科研基金项目(11QZR05); 华侨大学基本科研业务费专项基金项目(JB-SJ1003); 华侨大学高层次人才科研启动基金项目(09BS515)

**第一作者简介:** 杜吉祥(1977—), 男, 副教授, 2005年于中国科学技术大学获模式识别与智能系统专业博士学位, 主要研究方向为模式识别、神经网络。E-mail: jxd@hqu.edu.cn

件(事件识别)。

视频事件分析的主要内容包括输入一个视频序列,进行特征的提取,并用全局或者局部的特征来表示视频序列,接着对事件建立模型,最后分析、输出的结果可以是与输入视频序列相似的其他视频序列、也可以是对某特定的事件是否发生的判断;还可以进一步将结果用于生成视频事件摘要,以便于后续处理。从视频中提取的特征包括两类<sup>[2]</sup>:全局特征和局部特征。

全局特征表示,是将视觉观察值作为一个整体。该方法很强大,因为包含了很多信息,然而,它们依赖于精确的定位,而且对视角、噪声、遮挡非常敏感。当这些因素都能够得到很好的控制时,这种表示方法经常能表现得很好。

最近,局部时空特征已经变成一个流行的视频表示方法用于事件识别。局部特征表示描述的观察值是一系列独立块(patch)的集合。这种特征首先检测时空兴趣点,再计算这些点周围的局部patches,最后将这些patches组合成为最后的表示。相对来说,局部特征对噪声并不敏感,并不需要背景差以及精确定位,而且在一定程度上,局部特征能够在视角变化、人的表现变化以及部分遮挡的情况下保持不变。然而,它们依赖于足够的相关兴趣点的提取,有时候需要一些预处理,比如对摄像头运动的补偿。

在过去的几年里,已经提出了很多局部时空特征检测以及表示的方法。局部时空特征检测方法包括 Harris3D detector<sup>[3]</sup>, Cuboid detector<sup>[4]</sup>, Hessian detector<sup>[5]</sup>以及密集采样。局部特征的表示方法包括 HOG/HOF<sup>[6]</sup>, HOG3D<sup>[7]</sup>, extended SURF(ESURF)<sup>[5]</sup>。

先前很多的视频事件检测的研究,严重依赖于特定领域知识和先验模型,导致很难推广应用,如文献[8]中提出的方法只能用于检测吸烟事件,以及和吸烟类似的事件,如喝水和吃东西。为了能够提供一个通用的框架用于在真实的电影场景中解决事件识别问题,本文使用了文献[3]中提出的 Harris3D 检测器用于检测局部时空兴趣点,并使用兴趣点的 jet 响应作为兴趣点的特征。由于视频片段的帧数不一样,而且帧中含有的时空兴趣点也不一样,为了能够统一视频片段特征的表示,采用自组织特征映射的思想,用于训练的局部时空特征首先通过自组织学习过程构造一个2维的神经元平面,最后,每个视频序列表示成一个自组织映射像密度图。本文事

件识别的框架,不需要任何的目标检测和跟踪,不需要建立模型,也不用区分是单个人的或者多个人的交互,或者人车之间的交互或者其他的。和其他相关的方法相比,实验结果表明本文用于事件识别的框架明显提高了平均正确率和平均查准率。

## 1 特征描述与分类

### 1.1 局部时空特征描述

为了表示一个视频序列,使用文献[3]中提出的局部时空特征。为在视频序列 $f(x, y, t)$ 检测局部特征,构造视频序列的尺度空间表示为

$$L(g, \sigma^2, \tau^2) = f * g(\sigma^2, \tau^2)$$

式中,  $\sigma^2, \tau^2$  为尺度空间参数,  $g$  定义为高斯卷积核,即

$$g = \exp(-(x^2 + y^2)/2\sigma_i^2 - t^2/2\tau_i^2) / \sqrt{(2\pi)^3 \sigma_i^4 \tau_i^2}$$

使用空间-时间图像在  $x, y, t, 3$  个方向上的一阶偏导  $L_x, L_y, L_t$ , 定义二阶矩阵为

$$\mu = g(\cdot; \sigma_i^2, \tau_i^2) * \begin{pmatrix} L_x^2 & L_x L_y & L_x L_t \\ L_x L_y & L_y^2 & L_y L_t \\ L_x L_t & L_y L_t & L_t^2 \end{pmatrix} \quad (1)$$

式中缺省的参数表示  $x, y, t, 3$  个一阶偏导分别定义为

$$\begin{aligned} L_x(\cdot; \sigma_i^2, \tau_i^2) &= \partial_x(g * f) \\ L_y(\cdot; \sigma_i^2, \tau_i^2) &= \partial_y(g * f) \\ L_t(\cdot; \sigma_i^2, \tau_i^2) &= \partial_t(g * f) \end{aligned} \quad (2)$$

通过计算在  $(x, y, t)$  上的时空角函数式(3)局部最大,获得特征的位置。

$$H = \det(\mu) - \text{tr}^3(\mu) \quad (3)$$

定义归一化的  $L$  的二阶偏导为

$$\begin{aligned} L_{xx, \text{norm}} &= \sigma^{2a} \tau^{2b} L_{xx} \\ L_{yy, \text{norm}} &= \sigma^{2a} \tau^{2b} L_{yy} \\ L_{tt, \text{norm}} &= \sigma^{2c} \tau^{2d} L_{tt} \end{aligned} \quad (4)$$

假定在特定的空间和时间长度上  $L_{xx, \text{norm}}, L_{yy, \text{norm}}, L_{tt, \text{norm}}$  取得极值,计算式(4)差分,并令其差分为0,可获得  $a = 1, b = 1/4, c = 1/2, d = 4/3$ 。为了能够适应于不同的尺度,使用归一化的二阶偏导定义拉普拉斯算子为

$$\begin{aligned} \nabla_{\text{norm}}^2 L &= L_{xx, \text{norm}} + L_{yy, \text{norm}} + L_{tt, \text{norm}} = \\ &= \sigma^2 \tau^{1/2} (L_{xx} + L_{yy}) + \sigma \tau^{3/2} L_{tt} \end{aligned} \quad (5)$$

同时满足空间-时间角函数式(3)以及归一化的空

间-时间拉普拉斯算子式(5)最大化,即可求得期望的局部时空兴趣点及其相应的尺度。如图 1 所示,这 6 帧图像来自于一个打电话事件的视频序列,白色圆圈是检测到的时空兴趣点。局部特征的空间-

时间邻域包含了关于动作以及局部事件的空间表现信息,为了捕捉这些信息,计算时空 jets 响应

$$j = (L_x, L_y, L_t, L_{xx}, \dots, L_{ttt}) \quad (6)$$

对于检测到的每个兴趣点使用 jet 响应来表示。



图 1 时空兴趣点

Fig. 1 Space-time interest points

1.2 自组织特征映射及自组织分析

这部分主要展示基于自组织特征映射(SOFM)方法用于特征分类。由于视频片段的帧数不一样,而且每一帧中含有的时空兴趣点也不一样,为了能够统一视频片段特征表示,采用自组织特征映射<sup>[9]</sup>的思想,将输入的高维向量向 2 维平面映射。对于每个事件类别,用于训练的局部时空特征首先通过自组织学习过程构造一个 2 维的神经元平面,然后,每个视频序列表示成一个自组织映射像密度图。

自组织特征映射神经网络(图 2)定义为从一个输入数据空间  $R^n$  到一个普通的 2 维数组节点  $R^2$  的映射。经过充足的自组织学习阶段,网络趋向于根据输入数据集的结构而变得有空间组织性。在自组织网络节点平面上相邻或相隔较近的节点对应的类别,它们的相似性要比相隔较远的类别之间大。根据这个特性,自组织特征映射可以提供对原始数据集的聚类分析。

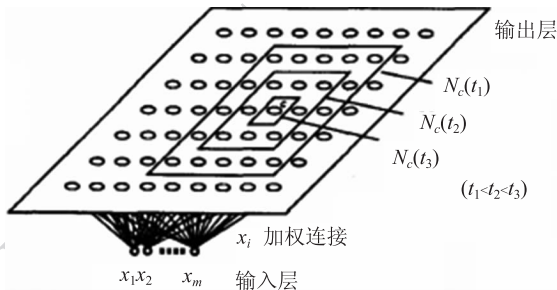


图 2 自组织特征映射神经网络

Fig. 2 Self-organization feature map neural networks

给定一个视频序列样本  $S = (j_1, j_2, \dots, j_m)$ , 该样本包含了  $m$  个特征,即  $m$  个时空兴趣点。如果输入特征  $j_i$  对应于映射平面中的节点  $i$ ,则称节点  $i$  或者节点  $i$  的位置为特征  $j_i$  的像,而该特征则成为这个节点的原像。一个节点可以是多个特征的像。

如果聚类分析的样本集分为较少的几个类,自组织映射网络虽然有一定优势,但并不明显。为此,采用了一种改进的方法——自组织映射分析,该方法以自组织特征映射为基础,在节点平面上统计各个节点的原像数目(称作像密度),得到如图 3 所示的像密度图  $H = (h_1, h_2, \dots, h_n)$ ,其中  $h_i$  是像为第  $i$  个节点的特征的个数。图 3(a)(b)来自于不同的事件类别样本——打电话和下车事件类别。以上所使用的神经网络由打电话事件类别中的正样本特征训练得到。

0	0	1	0	0	0	0	3	0	1
0	7	0	1	1	2	0	0	0	0
0	0	0	8	0	0	0	0	1	0
0	14	9	0	0	5	0	0	0	0
7	0	0	0	0	0	1	0	0	0
0	0	1	1	0	2	1	0	0	0
2	0	0	0	0	0	0	0	2	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0

(a) 打电话事件

(b) 下车事件

图 3 自组织特征映射像密度图

Fig. 3 Self-organization feature density maps

1.3 支持向量机分类

支持向量机(SVM)<sup>[10]</sup>是一个能够使两类的分

类间隙最大的分类器,目前已经在视觉模式识别方面得到了普遍应用。选用 SVM 作为分类器的原因有两个:1)SVM 的基础是统计学习理论,其理论研究是在有限样本情况下的机器学习问题,有别于传统的统计模式识别方法,只有在样本趋向无穷大的时候,传统的统计模式识别方法的性能才有理论上的保证;2)由于每个事件类别的分类识别问题是线性不可分的,而 SVM 可以对线性不可分的问题通过一个非线性的变换将原问题转化为某个高维空间中的线性问题,再在矩阵变换后的高维空间中求解广义最优分类面。

考虑如何将一个训练样本集合  $(s_1, y_1), (s_2, y_2), \dots, (s_n, y_n)$  分为两类,其中  $s_i$  是一个特征向量,而  $y_i \in \{-1, +1\}$  是类标签,而我们并没有关于数据分布的先验知识。假定两个类可以被一个最优分类面  $w \times s + b = 0$  所分割,而该分类面是一个能够使分类间隙最大的平面。其中  $w$  和  $b$  的最优值可以通过一个有约束的求最小值问题解决,其中使用了拉格朗日算子  $\alpha_i (i = 1, \dots, m)$ 。

$$f(s) = \text{sgn} \left( \sum_{i=1}^m \alpha_i y_i K(s_i, s) + b \right) \quad (7)$$

式中,  $\alpha_i$  和  $b$  可以通过使用一个 SVC 学习算法得到,而核函数,本文使用  $K(s_i, s) = s \times s_i$ 。

## 2 实验结果与分析

实验采用 Hollywood 数据库 (<http://www.irisa.fr/vista/Equipe/People/Laptev/download.html>), 数据

来自于 32 部电影的镜头序列,其中训练样本来自于 12 部电影,而测试样本来自于另外的 20 部电影。该数据集总共包含 8 个事件类别:打电话、下车、握手、拥抱、接吻、坐下、起身、站起,每个样本都已经标记。

在实验中,对这 8 类事件类别的正训练样本分别训练得到一个自组织神经网络,其中映射层的节点个数为  $10 \times 10$ 。对于每个类别的样本,都和另外的 7 个样本集组合,分成 7 组实验,从而得到总平均正确率、总平均查准率和总平均查全率分别为 0.601、0.530 和 0.566(实验结果如表 1 所示)。其中站起来事件类别的正确率如图 4 所示,7 条曲线分别为打电话事件类别与另外的 7 个事件类别分别组合测试得到的正确率。由于测试样本的不同,导致每条曲线的横坐标并不相同,综合 7 个组的实验,得到平均正确率为 0.609。

与文献[7,11]中的实验结果相比,本文的平均查准率有很大的提高。而在文献[12]中作者使用了 4 种局部特征检测的方法与 6 种不同的特征表示方法分别组合用于识别事件,使用 Hollywood 数据库的最好的平均正确率也只好 0.458。

## 3 结论

本文不使用任何的目标检测和跟踪,不需要建立模型,提供一个相对通用的事件检测方法,用于在真实的视频场景中检测事件。使用的 harris3D 检测

表 1 实验结果

Table 1 The experimental results

事件	本文			文献[7]平均查准率	文献[11]平均查准率		
	平均正确率	平均查全率	平均查准率		SIFT	HoG + HoF	SIFT + HoG + HoF
打电话	0.557	0.543	0.526	0.321	0.105	0.088	0.107
下车	0.672	0.327	0.387	0.415	0.191	0.090	0.116
握手	0.591	0.536	0.486	0.323	0.123	0.116	0.141
拥抱	0.597	0.611	0.545	0.406	0.129	0.135	0.138
接吻	0.603	0.702	0.664	0.533	0.348	0.496	0.556
坐下	0.564	0.554	0.680	0.386	0.161	0.316	0.278
起身	0.616	0.349	0.413	0.182	0.142	0.072	0.078
站起	0.610	0.753	0.690	0.505	0.262	0.350	0.325
总平均	0.601	0.530	0.566	0.384	0.183	0.208	0.217

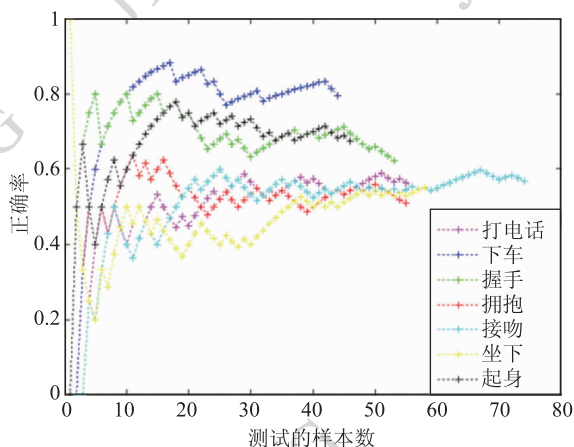


图4 站起来事件类别的平均正确率

Fig.4 The average correct rate of standup event

器能够解决多尺度的问题,可适应事件模式的不同大小和速度。检测到的兴趣点的特征使用 jet 响应表示,能够捕捉到关于动作以及局部事件的空间表现信息。和其他方法相比,实验结果表明本文提出的事件识别的框架明显提高了平均正确率和平均查准率。

### 参考文献 (References)

- [ 1 ] Lavee G, Rivlin E, Rudzsky M. Understanding video events; a survey of methods for automatic interpretation of semantic occurrences in video [ J ]. IEEE Transactions on Systems, Man and Cybernetics: Part C, 2009, 39(5):489-504.
- [ 2 ] Poppe R. A survey on vision-based human action recognition [ J ]. Image and Vision Computing, 2010, 28(6):976-990.
- [ 3 ] Laptev I, Lindeberg T. Space-time interest points [ C ] // Proceedings of the 9th IEEE International Conference on Computer Vision. Washington DC, USA: IEEE Computer Society, 2003:432-439.
- [ 4 ] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features [ C ] // Proceedings of the International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Washington DC, USA: IEEE Computer Society, 2005:65-72.
- [ 5 ] Willems G, Tuytelaars T, VanGool L. An efficient dense and scale-invariant spatial-temporal interest point detector [ C ] // Proceedings of the 10th European Conference on Computer Vision. Berlin: Springer, 2008:650-663.
- [ 6 ] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies [ C ] // Proceedings of the 18th IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2008:1-8.
- [ 7 ] Kläser A, Marszalek M, Schmid C. A spatio-temporal descriptor based on 3d-gradients [ C ] // Proceedings of the 19th British Machine Vision Conference. Leeds, Leeds, UK: BMVA Press, 2008:995-1004.
- [ 8 ] Wu P, Wei J, Cheng J, et al. Human smoking event detection using visual interaction clues [ C ] // Proceedings of the 20th International Conference on Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2010:4344-4347.
- [ 9 ] Zhang X, Li Y. Self-organizing map as a new method for clustering and data analysis [ C ] // Proceeding of International Joint Conference on Neural Networks. Washington DC, USA: IEEE Computer Society, 1993:2448-2451.
- [ 10 ] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach [ C ] // Proceedings of the 17th International Conference on Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2004:32-36.
- [ 11 ] Marszalek M, Laptev I, Schmid C. Actions in context [ C ] // Proceedings of the 12th IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2009:2929-2936.
- [ 12 ] Wang H, Muneeb M, Klaser A, et al. Evaluation of local spatio-temporal features for action recognition [ C ] // Proceedings of the 20th British Machine Vision Conference. London, UK: BMVA Press, 2009:127-138.