



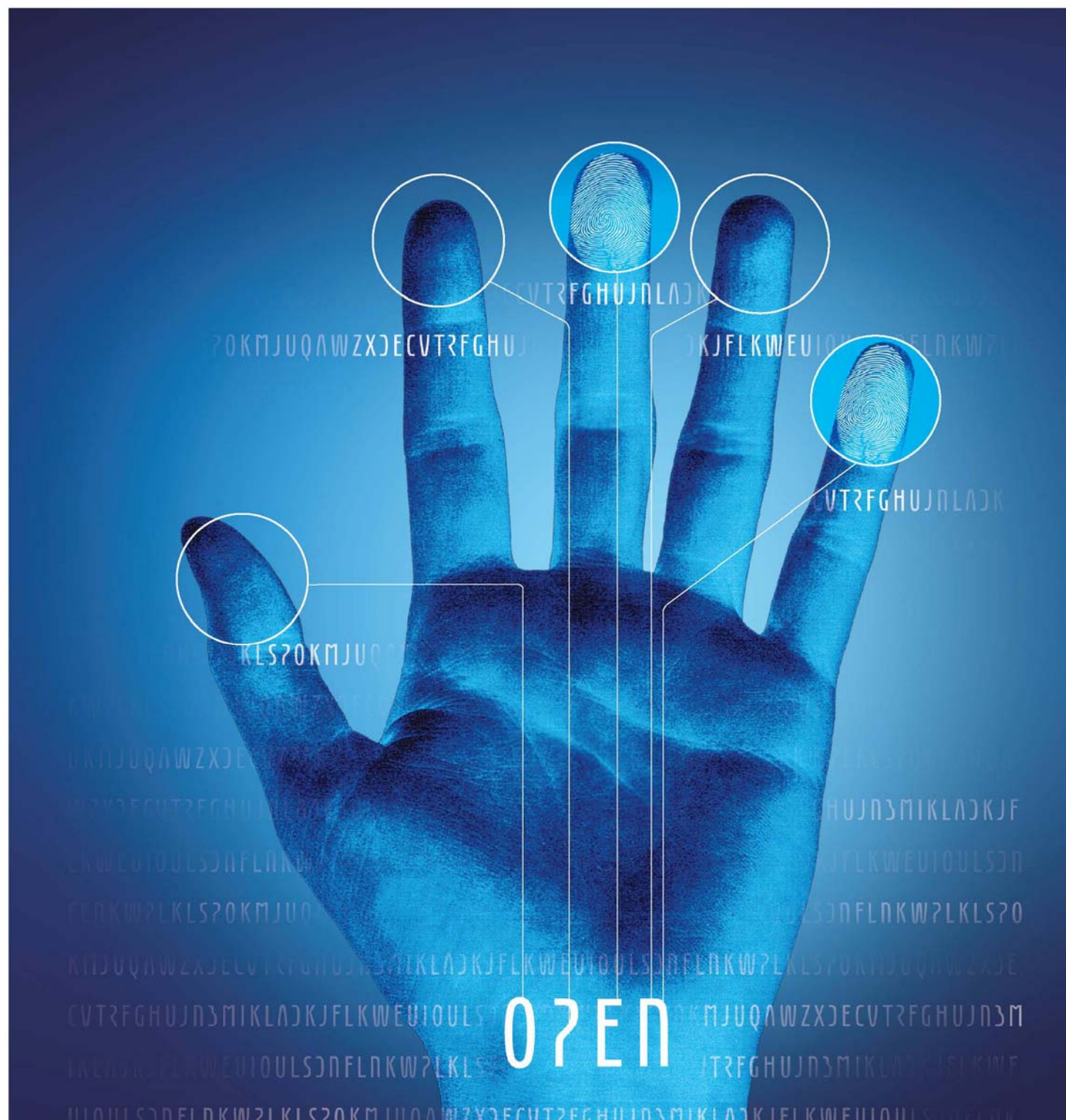
JOURNAL OF IMAGE AND GRAPHICS

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

2013  
07  
VOL.18

ISSN1006-8961  
CN11-3758/TB





第18卷第7期（总第207期）

2013年7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。所有内容，未经许可，不得转载或以其他方式使用。本刊所有图片均为非商业目的使用。

## Copyright

2013 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@irsa.ac.cn

电 话 010-64807995 010-82614429

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

国外发行 中国国际图书贸易总公司

（中国国际书店，邮政信箱：北京399信箱 邮编：100044）

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@irsa.ac.cn

Telephone 010-64807995 010-82614429

Website www.cjig.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, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

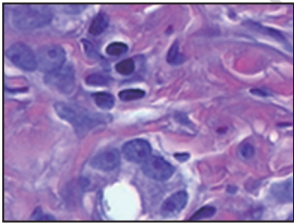
国外发行代号 M1406

国内邮发代号 82-831

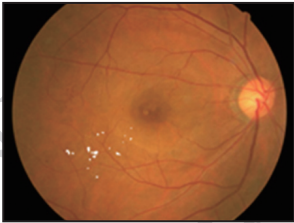
国内定价 50.00元



彩色纹理图像恢复的非局部TV模型 (第753页)



基于Nyström密度值逼近的减法聚类 (第790页)



RBF神经网络和阈值分割实现视网膜硬性渗出自动检测 (第859页)

图像处理 and 编码

|                                     |     |
|-------------------------------------|-----|
| 各向异性的均衡化网状扩散模型<br>熊辉, 赖剑煌           | 731 |
| 特征提取和匹配的图像倾斜校正<br>钟金荣, 杜奇才, 刘炎, 林嘉宇 | 738 |
| 先进边界区分噪声检测的改进算法<br>祁冰露, 黄宴委, 陈少斌    | 746 |
| 彩色纹理图像恢复的非局部TV模型<br>端金鸣, 潘振宽, 台雪成   | 753 |
| 相似性约束的视频超分辨率重建<br>张义轮, 干宗良, 朱秀昌     | 761 |

图像分析和识别

|                                      |     |
|--------------------------------------|-----|
| 低质量指纹图像方向场提取<br>卞维新, 徐德琴, 俞庆英        | 768 |
| 方向微分分块统计的运动模糊方向鉴别<br>李均利, 储诚曦        | 776 |
| 参数自适应的KPCA先验形状约束目标分割<br>沈霁, 李元祥, 周则明 | 783 |
| 基于Nyström密度值逼近的减法聚类<br>孙志海, 孔万增      | 790 |
| 空间目标检测序列图像中恒星目标抑制<br>黄宗福, 孙刚, 陈曾平    | 799 |

图像理解和计算机视觉

|                                     |     |
|-------------------------------------|-----|
| 人体动作的超兴趣点特征表述及识别<br>王扬扬, 李一波, 姬晓飞   | 805 |
| 轮廓编组中的线索合并模型<br>王娇, 罗四维, 钟晶晶, 邹琪    | 813 |
| 全局图像特征分析与实时层次化消失点检测<br>孙愿, 卢鸿波, 张志敏 | 818 |
| 彩色图像特征捆绑的自动模型构建<br>徐洁, 邓红霞, 李海芳     | 829 |

计算机图形学

|                                     |     |
|-------------------------------------|-----|
| 偏离映射的烙画风格绘制<br>钱文华, 徐丹, 岳昆, 官铮, 普园媛 | 836 |
| 点云模型法矢调整优化算法<br>孙金虎, 周来水, 安鲁陵       | 844 |

医学图像处理

|                                            |     |
|--------------------------------------------|-----|
| 基于惩罚最大似然优化模型的各向异性约束磁共振成像方法<br>邓梁, 史仪凯, 张均田 | 852 |
| RBF神经网络和阈值分割实现视网膜硬性渗出自动检测<br>高玮玮, 沈建新, 王玉亮 | 859 |

遥感图像处理

|                                                   |     |
|---------------------------------------------------|-----|
| 衡量结构保持性能的SAR图像去噪质量评价<br>唐益明, 杨学志                  | 866 |
| 地面LiDAR数据中建筑轮廓和角点提取<br>童礼华, 程亮, 李满春, 陈焱明, 王亚飞, 张雯 | 876 |

|                    |     |
|--------------------|-----|
| 《中国图象图形学报》论文摘要写作要求 | 884 |
|--------------------|-----|

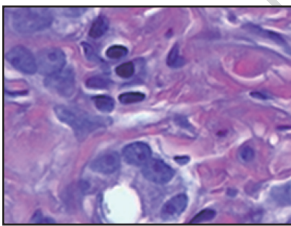


# CONTENTS

## JOURNAL OF IMAGE AND GRAPHICS



Non-Local TV models for restoration of color texture images (P753)



Density value approximation for subtractive clustering based on Nyström method (P790)



Automatic detection of hard exudates based on RBF neural network and threshold segmentation (P859)

### Image Processing and Coding

|                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| Equalized net diffusion model based on the anisotropy diffusion<br>Xiong Hui, Lai Jianhuang .....                         | 731 |
| Robust method of skew correction based on feature points matching<br>Zhong Jinrong, Du Qidai, Liu Ying, Lin Jiayu .....   | 738 |
| Modification of advanced boundary discriminative noise detection algorithm<br>Qi Binglu, Huang Yanwei, Chen Shaobin ..... | 746 |
| Non-Local TV models for restoration of color texture images<br>Duan Jinming, Pan Zhenkuan, Tai Xuecheng .....             | 753 |
| Video super-resolution method based on similarity constraints<br>Zhang Yilun, Gan Zongliang, Zhu Xiuchang .....           | 761 |

### Image Analysis and Recognition

|                                                                                                                                                 |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| The orientation field extraction method for low quality fingerprints<br>Bian Weixin, Xu Deqin, Yu Qingying .....                                | 768 |
| Identification of motion blur direction from motion blurred image by directional derivation and block statistics<br>Li Junli, Chu Chengxi ..... | 776 |
| Shape prior constrained KPCA object segmentation with parameter adaption<br>Shen Ji, Li Yuanxiang, Zhou Zeming .....                            | 783 |
| Density value approximation for subtractive clustering based on Nyström method<br>Sun Zhihai, Kong Wanzeng .....                                | 790 |
| Stellar targets suppression in image sequences for space targets detection<br>Huang Zongfu, Sun Gang, Chen Zengping .....                       | 799 |

### Image Understanding and Computer Vision

|                                                                                                                             |     |
|-----------------------------------------------------------------------------------------------------------------------------|-----|
| Human action recognition based on super-interest points features<br>Wang Yangyang, Li Yibo, Ji Xiaofei .....                | 805 |
| Cue combination model for contour grouping<br>Wang Jiao, Luo Siwei, Zhong Jingjing, Zou Qi .....                            | 813 |
| Hierarchical real-time vanishing point detection based on holistic image feature<br>Sun Yuan, Lu Hongbo, Zhang Zhimin ..... | 818 |
| Construction of the automatic model for feature binding of color image<br>Xu Jie, Deng Hongxia, Li Haifang .....            | 829 |

### Computer Graphics

|                                                                                                                               |     |
|-------------------------------------------------------------------------------------------------------------------------------|-----|
| Rendering pyrography style painting based on deviation mapping<br>Qian Wenhua, Xu Dan, Yue Kun, Guan Zheng, Pu Yuanyuan ..... | 836 |
| Optimal algorithm for normal adjustment of point clouds<br>Sun Jinhu, Zhou Laishui, An Luling .....                           | 844 |

### Medical Image Processing

|                                                                                                                                             |     |
|---------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Anisotropically constrained MR imaging based on penalized maximum likelihood optimality model<br>Deng Liang, Shi Yikai, Zhang Juntian ..... | 852 |
| Automatic detection of hard exudates based on RBF neural network and threshold segmentation<br>Gao Weiwei, Shen Jianxin, Wang Yuliang ..... | 859 |

### Remote Sensing Image Processing

|                                                                                                                                                           |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Image quality assessment of SAR image despeckling for measuring structure-preserving capability<br>Tang Yiming, Yang Xuezhi .....                         | 866 |
| Extraction of building contours and corners from terrestrial LiDAR data<br>Tong Lihua, Cheng Liang, Li Manchun, Chen Yanming, Wang Yafei, Zhang Wen ..... | 876 |

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2013)07-0805-08

论文引用格式: 王扬扬, 李一波, 姬晓飞. 人体动作的超兴趣点特征表述及识别[J]. 中国图象图形学报, 2013, 18(7): 805-812. [DOI: 10.11834/jig.20130710]

# 人体动作的超兴趣点特征表述及识别

王扬扬<sup>1</sup>, 李一波<sup>2</sup>, 姬晓飞<sup>2</sup>

1. 南京航空航天大学自动化学院, 南京 210016; 2. 沈阳航空航天大学自动化学院, 沈阳 110136

**摘要:** 提出一种基于超兴趣点的动作特征描述方法, 用于人体动作的识别。兴趣点特征描述了人体动作时变化显著的局部点信息, 但其最大的缺陷在于离散的兴趣点间缺乏时间和空间上的结构关联。提出根据兴趣点间的时空距离, 使用广度优先搜索邻居算法, 将时空距离相近的兴趣点聚合成超兴趣点, 该结构作为一个整体, 反映人体在一定时空范围内的动作变化特征。与现有的基于局部兴趣点的动作识别算法相比, 本文算法增加了兴趣点间的整体时空结构关系, 提高了特征的区分度。实验采用两层分类方法对超兴趣点特征分类, 实验结果表明该算法具有较好的识别率。

**关键词:** 超兴趣点; 动作识别; 广度优先搜索邻居; 时空距离

## Human action recognition based on super-interest points features

Wang Yangyang<sup>1</sup>, Li Yibo<sup>2</sup>, Ji Xiaofei<sup>2</sup>

1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;

2. College of Automation, Shenyang Aerospace University, Shenyang 110136, China

**Abstract:** A novel feature representation method based on super-interest points is proposed and applied to human action recognition. Interest point features describe the salient change of local point information when humans are in the state of movement, however, the drawback of these features is that the rich spatio-temporal relationships among the discrete interest points are not being used. According to the spatio-temporal distance of interest points, a broad first search neighbors algorithm is used to cluster these adjacent points into super-interest point. The super-interest point features reflect the variety of the human limb in a certainly spatio-temporal scope. Compared to the existing interest point method, the spatial and temporal information among the points, are all included in our super-interest points, making the features more discriminative. Finally, a two hierarchical classifier recognizes the human actions using super-interest point features. The experimental results show that our method achieves a good recognition rate.

**Key words:** super-interest points; action recognition; broad first search neighbors; spatio-temporal distance

## 0 引言

基于视频的人体动作识别是目前机器视觉领域的一个研究热点, 被广泛应用于智能监控, 视频检

索, 人机交互等领域, 其主要任务是在一定的场景环境中, 由计算机自主识别人体动作的含义。

动作识别的研究主要分为动作特征的提取描述和识别两个方面。动作的特征描述分为基于全局的特征和基于局部的特征两大类。基于全局的特征描

收稿日期: 2012-09-17; 修回日期: 2012-11-22

基金项目: 国家自然科学基金项目(61103123)

第一作者简介: 王扬扬(1979—), 女, 工程师, 现为南京航空航天大学自动化学院模式识别与智能系统专业在读博士研究生, 主要研究方向为模式识别、计算机视觉、行为识别。E-mail: wyw2004101@163.com

述运动人体的整体属性,Efros 等人<sup>[1]</sup>提取以人为中心的感兴趣区域,对区域内的前景目标计算光流特征。Yilmaz 等人<sup>[2]</sup>提出用时空体(STV)描述人体的整个轮廓,既利用了动作的轮廓信息又包含了时间信息。该类特征能较准确地反映人肢体各部分间的时空结构关系,但是其特征描述符的动作区分度建立在精确分割并跟踪人体运动目标的基础上,对背景噪声和遮挡比较敏感。

基于局部的特征,只选取人体中的某一部分信息作描述,描述人体某一局部子区域的特征,受背景变化、遮挡等影响较小。局部特征描述符中使用最广泛的是兴趣点特征。该方法的基本流程遵循以下 3 个步骤:首先,设计兴趣点检测器,从视频中检测到人体动作时变化显著的关键点;其次,把这些关键点作为区分动作类别的兴趣点,在以兴趣点为中心的邻域内捕获局部形状、外观和运动等特征进行描述;最后,构建分类器并识别动作。对于兴趣点检测器,Dollar 等人<sup>[3]</sup>直接对视频 3D 图像序列进行空域和时域上的滤波,捕捉人体部位区域的运动巨变;Laptev 等人<sup>[4]</sup>将 2D 空间的角点检测扩展到 3D 空间;Bregonzio 等人<sup>[5]</sup>提出 2D 帧差和 1D Gabor 滤波器结合的检测器。上述各种兴趣点检测器检测到的兴趣点在稀疏度上各不相同,但其根本宗旨是在对执行动作起主要作用的人体部位上,选择尽可能多的反映局部特征信息的兴趣点,同时尽量避免由于背景噪声、摄像机移动等干扰引起的伪检测。在兴趣点特征描述符的表述上,目前最常用的是词袋法(bag of words)<sup>[6]</sup>。它以兴趣点为中心,构造若干规则的小立方体或球体,对每个小立方体,采用梯度直方图/光流直方图(HOG/HOF)<sup>[7]</sup>、扩展 3D 梯度直方图(HOG3D)<sup>[8]</sup>等特征描述。文献[9]将 SIFT(scale invariant feature transform)特征、HOG 特征和像素亮度等组合在一起,构成特征向量,每一个特征向量对应一个特征描述符,通过聚类使每个特征向量对应一个与它距离最近的聚类中心(visual word)。于是,视频序列被转化为词袋形式,可以通过各视频词出现频率等特征的分析,构建相应直方图或者树型结构<sup>[10]</sup>,以供最终分类使用。

虽然基于兴趣点的方法对视觉、遮挡和噪声等干扰不敏感,处理过程简单,但仍存在以下缺陷:1)使用词袋法对兴趣点特征进行描述时,各兴趣点间的时空关系和逻辑关系较难获取,特征描

述符的鲁棒性差,不利于识别率的提高;2)全局信息缺失,对局部运动相似度比较大的动作,如跑、走和向前跳等,无法较好地保证类间差异度大的同时类内差异度小。这些也是国内外目前着力研究探索的方向。文献[11-13]除了捕捉局部动态的兴趣点信息之外,还引入了全局特征(如静态轮廓、动态光流等),在一定程度上依靠身体姿态增加类间特征的区分度,但是两类特征的融合不能动态地改变。文献[14]将每帧中的兴趣点组成一个集合,每个集合对应一个姿态,通过非监督姿态建模实现对动作的描述,但是缺乏帧与帧之间的兴趣点间的联系。

文献[6,15-17]等对局部兴趣点进行聚类,将兴趣点与码本表述相结合。具有相似属性的局部特征被聚合在一起,目前在动作识别领域采用最多的聚类方法为 k 均值(k-means)方法<sup>[6,15-17]</sup>。该方法的优点是算法简单,但需要预先设定类的个数。对于不同类型动作的兴趣点,其数量及位置分布均有很大差异,即使是同一类动作,不同人由于动作方式风格的不同、光照衣着等影响,产生的兴趣点也会存在一定的类内差异,聚类过程中固定类的个数,可能将具有较大类内差异的兴趣点划分到不同类别的聚类中。文献[18]采用凝聚聚类方法,迭代合并兴趣点,直至满足合并终止条件。该方法虽然不需要预先确定类的个数,但一旦制定了合并规则及终止条件,就必须将聚类进行到底,容易将噪声点作为正常数据参与分类。文献[19]提出相互信息最大化聚类法,该方法在 k-means 的基础上,对聚类结果再次优化合并,可以自动确定最优的类的个数,更准确地划分出不同类的兴趣点集合,使描述动作特征的结构更紧凑,但缺点是需要耗费大量的计算时间,并且仍然只能得到局部最优解。

为此提出一种超兴趣点的特征描述方法,将视频中孤立分布在各帧中的兴趣点,依据其时空距离分布聚集度,采用基于广度优先搜索邻居的聚类算法(BFSN)生成若干超兴趣点。该算法相比于上文所述的聚类方法,在保证计算效率的基础上,提高聚类准确度,通过设置两个距离参数,除了能区分开兴趣点是否属于同一类别,还可以获取同一聚类中各点的差异程度,进而生成任意形状的超兴趣点,同时噪声点自成一类减少干扰。超兴趣点将一定时间范围内分布在同一局部肢体上的兴趣点聚合在一起,以超兴趣点整体为单位



构造特征描述符,相比以兴趣点为单位构造特征描述符的方法,既减少了生成特征描述符的计算量,又增加了兴趣点间的时空关联信息,有利于提高人体动作识别的识别率。

## 1 方法概述

采用超兴趣点特征描述并识别人体的动作,识别过程主要由3个步骤构成:

### 1) 兴趣点检测

本文方法需要在离散的兴趣点间建立一定的时空联系,在松散的个体结构之上,聚合生成一个更紧密的时空集合体,因此需要选择能生成较多数量的兴趣点的兴趣点检测器,Bregonzio法<sup>[5]</sup>检测视频中的兴趣点。

由于不同人在视频中出现的位置不同,为提高算法的识别率,需要对兴趣点作必要的规格化处理。通过平移缩放等预处理,将视频中各帧的人运动区域及检测到的兴趣点规格化为以人的质心为中心的新的视频序列。

### 2) 生成超兴趣点描述符

对步骤1)得到的兴趣点,首先采用BFSN算法作聚类分析,将时空距离相近的兴趣点聚合成一个“超兴趣点”(super-interest points)。通过调整BFSN算法的参数,对不同类型的动作可以构造出适合其自身运动变化特点的超兴趣点结构,即将一段微小时间内聚集在人某一肢体部位附近的兴趣点集合成一个超兴趣点。然后,以超兴趣点为一个整体单元构造特征描述符向量,描述人局部肢体层面的特征信息,使得设计的识别系统对动作的空间变化具有更强的鲁棒性。

### 3) 描述符分类

用最近邻分类器对超兴趣点特征描述符进行分类,得到最终的分类结果。

## 2 超兴趣点描述符

### 2.1 超兴趣点含义

超兴趣点,即由时空距离相近的若干兴趣点组成的一个空间集合体。该集合体能同时反映人体运动的时间变化和空间变化,其中空间变化主要反映人体执行动作使用到的局部肢体部位和动作执行方式,时间变化则主要反映在上述空间范围内,人体执

行动作的时间长短。

通常检测到的兴趣点离散分布在多帧图像序列中,各点相互之间的时空关系是割裂的,而本文通过构造超兴趣点,将一定时空范围内的若干兴趣点集合在一起,不仅可以兼顾超兴趣点内部各点的子特征信息,还可获取近似兴趣点间的时空联系,得到一个具备相对整体特征的时空信息,使得本文系统对动作执行的时空变化有更强的鲁棒性。

### 2.2 基于BFSN算法的超兴趣点提取

钱江波等人<sup>[20]</sup>提出了一种基于广度优先搜索邻居的聚类算法。该算法目前多应用于雷达信号分析领域。将该算法应用到动作识别领域,并从2维空间扩展到3维空间,对检测到的兴趣点作聚类分析,生成超兴趣点。

BFSN算法根据兴趣点的空间欧氏距离,定义两个距离阈值 $r$ 和 $\lambda$ ,由这两个参数确定兴趣点间是否存在直接邻居或间接邻居关系,即对视频中兴趣点的时空聚集度进行判定,将满足阈值条件的兴趣点聚集成一个超兴趣点。

#### 1) BFSN算法涉及术语及相关参数定义

$d_{ij}$ 表示兴趣点 $p_i$ 和兴趣点 $p_j$ 之间的空间欧氏距离,反映两个点的时空聚集度,距离公式为

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (t_i - t_j)^2} \quad (1)$$

式中, $(x_i, y_i, t_i)$ ,  $(x_j, y_j, t_j)$ 分别表示兴趣点 $p_i, p_j$ 的空间坐标。

所有兴趣点两两计算 $d_{ij}$ ,构成下三角相似度矩阵

$$M_{\text{simi}}(i, j) = \begin{cases} d_{ij} & i \geq j \\ 0 & i < j \end{cases} \quad (2)$$

参数 $r$ 表示直接邻居阈值,如果兴趣点 $p_i, p_j$ 的 $M_{\text{simi}}(i, j) < r$ , ( $i, j = 1, \dots, m$ ),则这两个兴趣点为直接邻居,简称邻居。兴趣点 $p_i$ 的所有邻居的集合记为 $N_a$ 。

参数 $\lambda$ 表示间接邻居阈值。假设存在 $m$ 个兴趣点 $p_1, p_2, \dots, p_m$ 满足 $p_m$ 是 $p_{m-1}$ 的邻居, $p_1$ 是 $p_2$ 的邻居, $p_k$ 是 $p_{k+1}$ 和 $p_{k-1}$ 的邻居( $k = 2, \dots, m-1$ ),则 $p_3, \dots, p_m$ 均定义为兴趣点 $p_1$ 的间接邻居。

假设已存在一个超兴趣点 $S_{ip} = [p_1, p_2, \dots, p_n]$ ,  $n$ 表示兴趣点集包含的兴趣点个数,现在判定一个新的兴趣点 $p_{n+1}$ 是否属于 $S_{ip}$ ,则必须满足条件

$$\frac{\sum_{u=1}^n X_u}{n} \geq \lambda \quad (3)$$

式中,  $X_u$  为  $p_{n+1}$  是否为  $S_{ip}$  中任一点  $p_u$ , ( $u=1, \dots, n$ ) 的直接邻居的逻辑标量, 即

$$X_u = \begin{cases} 1 & d_{n+1,u} \leq r \\ 0 & d_{n+1,u} > r \end{cases} \quad (4)$$

如果兴趣点  $p_{n+1}$  满足式(3), 则该点属于  $S_{ip}$ , 否则不属于。

## 2) BFSN 算法的基本流程

首先任意选取一个兴趣点, 搜索该兴趣点的所有直接邻居, 然后再分别搜索集合  $N_u$  中的每个兴趣点的所有直接邻居, 对搜索到的所有兴趣点计算是否满足式(3), 如果满足, 则将它们合并, 从而完成一个  $S_{ip}$  的聚类。然后依此过程, 寻找其他的  $S_{ip}$ , 最后不属于任何  $S_{ip}$  的点为噪声点。

每个视频中超兴趣点的个数和空间形态由 BFSN 算法的距离阈值  $r$  和  $\lambda$  的大小确定。在  $r$  值一定的条件下, 当  $\lambda$  的值较大时, 超兴趣点内部各点的差异较小, 一个视频中  $S_{ip}$  的个数较多; 反之, 当  $\lambda$  的值较小时, 超兴趣点内部各点的差异较大, 一个视频中  $S_{ip}$  的个数减少。通过实验选取适当的距离阈值  $r$  和  $\lambda$ , 在一定的时间区域内, 将集中在不同人肢体部位, 如小腿、大脚、脚等部位的兴趣点聚集成不同的超兴趣点, 同时排除掉噪声点。

以 Weizmann 数据库<sup>[21]</sup>为例, 图 1 给出了在不同距离阈值条件下同一段跑步视频的超兴趣点分布

对比图。当距离阈值参数  $r$  和  $\lambda$  保持一致, 分别为 0.60、0.55 时, 选取该数据库的 10 种不同动作视频构造超兴趣点, 其分布效果如图 2 所示。图 2 中第 1 排从左到右 5 个子图分别代表弯腰、原地抬手跳、原地直立双腿跳、原地挥单臂和原地挥双臂的超兴趣点分布, 第 2 排从左到右 5 个子图分别代表双腿向前跳、跑、侧身跑、单腿向前跳、走的超兴趣点分布。图中不同颜色及不同符号的点代表其分属不同的超兴趣点。

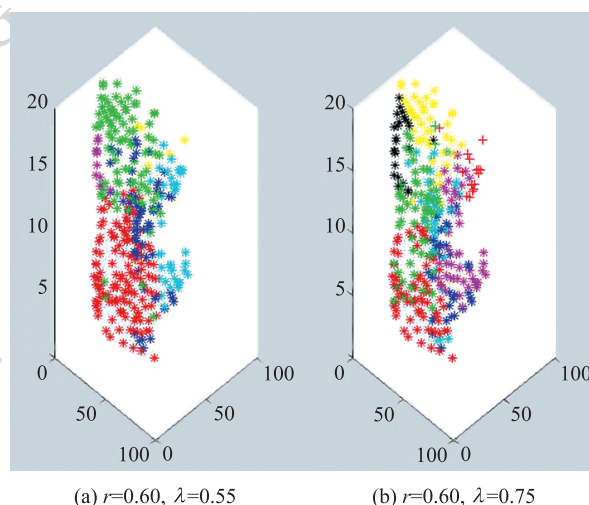


图 1 距离阈值对超兴趣点的影响

Fig. 1 The influence of distance thresholds on super-interest points

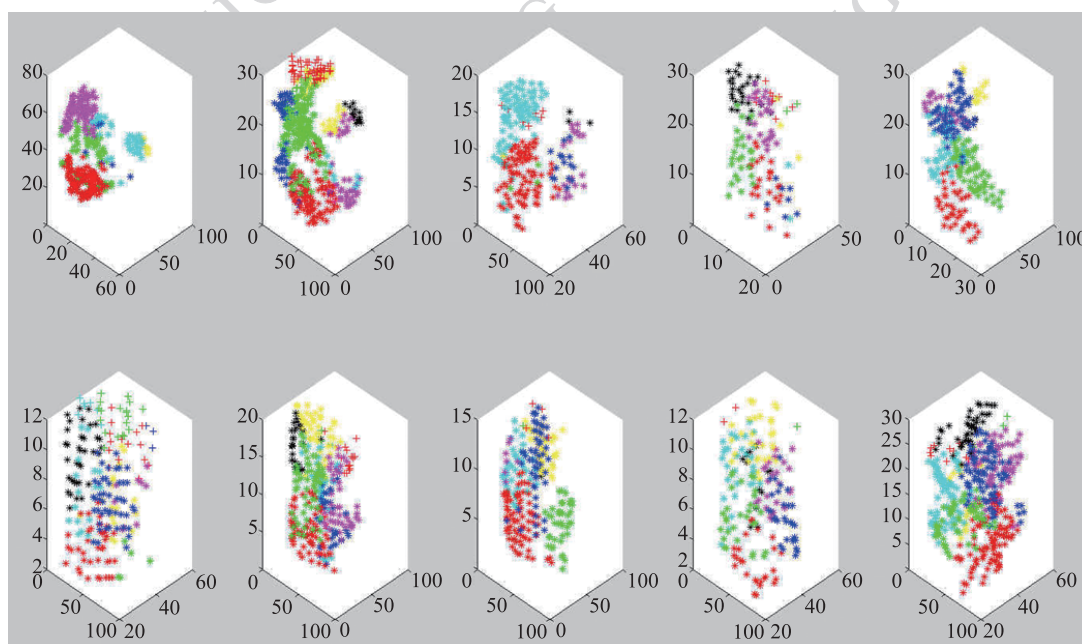


图 2 Weizmann 数据库 10 类动作的超兴趣点分布

Fig. 2 The distribution of super-interest points of ten kinds of actions on Weizmann dataset



与其他聚类算法相比,该算法效率较高,并且能根据距离阈值调整超兴趣点内各兴趣点间的相异度,易于构造有较强区分度的人体动作特征描述符。

### 2.3 超兴趣点描述符

根据 BFSN 算法,一个视频的兴趣点聚类生成若干超兴趣点,对每个超兴趣点  $S_{ip}$  构造描述符  $F$ 。描述符由超兴趣点的整体外观特征  $F_w$  和超兴趣点的 3D-SIFT 特征  $F_s$  两部分构成。其中  $F_w = \langle S_h, S_w, S_d, C_x, C_y, C_t, N_p, H_p \rangle$ , 8 个分量分别表示由超兴趣点集构成的空间体的长、宽、厚,空间体中心点的  $x, y, t$  坐标,超兴趣点集中包含的兴趣点的个数占视频中所有兴趣点的个数的百分比,超兴趣点集在整个人体平面所处的相对位置编码。根据  $(C_x, C_y)$ , 将超兴趣点集投影到 2 维人体平面,该平面大小规格化为  $80 \times 80$ ,划分为 16 行 16 列个正方形子区域,对所有子区域按列向排序,各子区域的序号即位置编码。

$F_s$  计算一个超兴趣点的 3D-SIFT 特征<sup>[22]</sup>。两部分特征  $F_w$  和  $F_s$  加权线性组合生成最终的特征描述符  $F$ 。

## 3 人体动作分类

本文要识别的人体动作,特指弯腰、挥手、跑、走、跳等简单的动作,这些动作根据腿部是否基本位于原地不动,可以分为基于原地不动的和移动的两大类动作。基于原地不动的动作,人整体形状变化较大,兴趣点的分布相对分散;相对而言,移动类的动作,人的整体轮廓变化较小,兴趣点比较集中,因此在利用 BFSN 算法构造超兴趣点时,距离阈值  $r$  和  $\lambda$  的选择对不同类型的动作的影响是不同的。对于移动类的动作(如图 2 第 2 排所示),本身局部运动的相似度就比较大,当  $r$  值不变时,如果选择的  $\lambda$  值较小,会增加超兴趣点内各兴趣点的差异度,不利于动作的区分。而对于原地不动的动作,当  $r$  值不变时,选择较小的  $\lambda$  值,更利于保证超兴趣点与人整体形状变化的一致性。因此采用两级分类的方法对人体动作进行识别。第 1 级基于人体目标区域的质心特征将动作划分成原地不动的和移动的两类;第 2 级对不同类型的动作选取不同的距离阈值  $r$  和  $\lambda$ , 分别对动作分类。

第 1 级分类,使用 Bregonzio 方法检测兴趣点

的同时,可同步检测出视频序列中人体所在的位置,进而提取每帧图像中人的中心。通过判断中心运动轨迹,可以很容易地将两类不同的运动划分开来。

第 2 级分类,虽然对不同类型的动作选取的距离阈值  $r$  和  $\lambda$  不同,构造的超兴趣点集描述符不同,但均采用最近邻分类器对动作分类。

假定一个未知动作的视频含  $n$  个超兴趣点,每个兴趣点集对应一个特征向量  $F_k$ , ( $k = 1, \dots, n$ ), 采用最近邻方法。已知训练集中存在  $t_m$  个训练视频样本,分别将  $F_k$  与  $t_m$  个样本中的每个超兴趣点的特征作最近邻比较,将与其距离最近的超兴趣点所属的动作标号赋给  $F_k$ , 每个  $F_k$  对应一个动作标号,将所有  $F_k$  的标号作多数投票,得票数最多的动作标号即为待识别视频的所属动作。

## 4 实验结果

### 4.1 实验数据库

为与其他算法作比对,采用 Weizmann 数据库作为测试对象。Weizmann 数据库是目前研究人体动作识别的基准数据库,各国研究者均采用该库做算法的识别率和效率测试,其优点是对不同算法可以用统一的评测标准作比较分析。

Weizmann 数据库包含弯腰、原地抬手跳、原地直立双腿跳、原地挥单臂、原地挥双臂、双腿向前跳、跑、侧身跑、单腿向前跳和走 10 种不同的动作。视频的分辨率为  $180 \times 144$ 。每类动作由 9 个不同的人分别执行,视频中人的衣着、动作风格、动作幅度和速率均不相同,每类动作均含有较大的类内差异,能较好验证算法的鲁棒性。测试过程中,采用留一法交叉验证,每次只取一个人的动作作为测试样本,其他人的动作作为训练样本,对数据库中的所有入重复上述过程,最终结果取平均。

### 4.2 算法效率分析

现有的兴趣点算法在检测到兴趣点后,直接以兴趣点为中心计算特征描述符或将兴趣点转换为码本形式进而生成特征描述符。在兴趣点算法基础上,在特征描述阶段,对兴趣点聚类生成超兴趣点,再以超兴趣点为单位进行特征的描述,现就增加聚类算法对识别效率的影响加以分析。

算法采用 Matlab 语言编程,测试环境:处理器 4 核,2.4 GHz,内存 2 GB。

假定平均一段视频中约有 350 个兴趣点,经 BFSN 聚类算法共生成超兴趣点 10 个左右,在聚类同时生成所有超兴趣点的整体外观特征  $F_w$ 。两者所需的总时间约为 0.014 12 s。此外每个超兴趣点的形状是可变的,其空间集合体中包含的像素数量级一般在  $1 \sim 10^2$  之间,因此对应每个超兴趣点的 3D-SIFT 特征  $F_s$  的计算时间为  $10^{-3} \sim 10^1$  s,平均单个  $F_s$  的计算时间约为 1 s,所有  $F_s$  的计算时间约为 10 s。综上,本文算法在特征描述阶段所用的总时间约为  $10 + 0.014\ 12 = 10.014\ 12$  s。

而对兴趣点算法,采用相同的 3D-SIFT 特征描述符,计算 1 个兴趣点特征的平均时间为 0.227 3 s,计算 350 个兴趣点的 3D-SIFT 特征的总时间为 79.555 s。两者相比,虽然本文算法在特征描述过程中引入了聚类算法,但聚类本身所需时间很少,小于计算一个兴趣点 3D-SIFT 特征的时间,可以忽略不计,并且由于聚类,大量兴趣点被合并,极大减少了计算 3D-SIFT 特征的时间,反而减少了计算特征

描述符的总时间。

4.3 实验结果分析

采用本文算法需要设定两个距离阈值  $r$  和  $\lambda$ 。阈值  $r$  主要控制类间差距,阈值  $\lambda$  主要控制类内差异。经实验验证,当  $r$  值固定不变时, $\lambda$  的变化对基于原地不动的和移动的两大类动作的影响不同。 $\lambda$  值较大时,超兴趣点内部包含的各点的差异度较小,对跑、跳、走等动作比较相近的移动类视频识别效果较好;反之, $\lambda$  值较小时,超兴趣点包含的范围较大,能更好地反映人的整体轮廓的变化,更符合原地不动类动作的运动特点。表 1 给出了不同参数下的动作识别结果。表 1 中数据表示正确识别出动作的百分率。由表 1 可见,对原地不动类动作(表 1 中的前 5 个动作),距离阈值  $r$  和  $\lambda$  分别为 0.60 和 0.55 时,超兴趣点的分类效果较好;而对移动类的动作(表 1 中的后 5 个动作),距离阈值  $r$  和  $\lambda$  分别为 0.60 和 0.75 时,超兴趣点的分类效果较好。因此两类动作在各自的最佳参数下,最终得到的识别混淆矩阵如图 3 所示,识别率为 95.6%。

表 1 Weizmann 数据库基于不同距离阈值的动作识别结果  
Table 1 Results using different distance thresholds on Weizmann dataset

|      |         | 识别率/%                      |                            |                            |                            |                            |                            |
|------|---------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|----------------------------|
| 动作序号 | 动作      | $r=0.60$<br>$\lambda=0.75$ | $r=0.60$<br>$\lambda=0.55$ | $r=0.75$<br>$\lambda=0.75$ | $r=0.75$<br>$\lambda=0.55$ | $r=0.45$<br>$\lambda=0.75$ | $r=0.45$<br>$\lambda=0.55$ |
| 1    | 弯腰      | 100                        | 100                        | 100                        | 100                        | 88.89                      | 100                        |
| 2    | 原地抬手跳   | 100                        | 100                        | 100                        | 88.89                      | 88.89                      | 77.78                      |
| 3    | 原地直立双脚跳 | 77.78                      | 88.89                      | 77.78                      | 88.89                      | 50                         | 50                         |
| 4    | 原地挥单臂   | 100                        | 100                        | 100                        | 100                        | 100                        | 88.89                      |
| 5    | 原地挥双臂   | 88.89                      | 100                        | 100                        | 100                        | 100                        | 100                        |
| 6    | 双腿向前跳   | 100                        | 88.89                      | 77.78                      | 88.89                      | 66.67                      | 100                        |
| 7    | 跑       | 100                        | 44.44                      | 77.78                      | 88.89                      | 66.67                      | 77.78                      |
| 8    | 侧身跑     | 88.89                      | 66.67                      | 50                         | 88.89                      | 77.78                      | 88.89                      |
| 9    | 单腿向前跳   | 88.89                      | 33.33                      | 88.89                      | 66.67                      | 44.44                      | 66.67                      |
| 10   | 走       | 88.89                      | 100                        | 100                        | 100                        | 66.67                      | 100                        |

为验证本文 BFSN 聚类算法和超兴趣点结构的有效性,采用相同的兴趣点检测器、特征描述符分别获取未经过聚类的兴趣点特征和用 k-means 聚类方法生成的超兴趣点特征,使用相同的分类算法计算

动作识别率,结果如表 2 所示。此外,比较采用其他特征描述符的识别算法,编辑距离特征<sup>[23]</sup>,基于外形-运动特征<sup>[24]</sup>的动作识别率分别为 94.6%,94.4%,本文算法具有较好的识别结果。

动作序号

|    |     |     |     |     |     |     |     |     |     |    |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 1  | 1.0 | .0  | .0  | .0  | .0  | .0  | .0  | .0  | .0  |    |
| 2  | .0  | 1.0 | .0  | .0  | .0  | .0  | .0  | .0  | .0  |    |
| 3  | .0  | .11 | .89 | .0  | .0  | .0  | .0  | .0  | .0  |    |
| 4  | .0  | .0  | .0  | 1.0 | .0  | .0  | .0  | .0  | .0  |    |
| 5  | .0  | .0  | .0  | .0  | 1.0 | .0  | .0  | .0  | .0  |    |
| 6  | .0  | .0  | .0  | .0  | .0  | 1.0 | .0  | .0  | .0  |    |
| 7  | .0  | .0  | .0  | .0  | .0  | .0  | 1.0 | .0  | .0  |    |
| 8  | .0  | .0  | .0  | .0  | .0  | .0  | .89 | .0  | .11 |    |
| 9  | .0  | .0  | .0  | .0  | .0  | .11 | .0  | .89 | .0  |    |
| 10 | .0  | .0  | .0  | .0  | .0  | .0  | .11 | .0  | .89 |    |
|    | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10 |

动作序号

图 3 Weizmann 数据库识别混淆矩阵

Fig. 3 Confusion matrix on Weizmann dataset

表 2 Weizmann 数据库基于不同算法的动作识别率

Table 2 Results using different algorithm on Weizmann dataset

|       | 兴趣点的<br>3D-SIFT<br>特征 | k-means 聚类<br>生成的超<br>兴趣点特征 | BFSN 聚类<br>生成的超兴趣<br>点特征 |
|-------|-----------------------|-----------------------------|--------------------------|
| 识别率/% | 88.9                  | 93.3                        | 95.6                     |

## 5 结 论

采用超兴趣点作为特征描述符,对跑、跳、走、挥手等人体动作进行识别。超兴趣点这一特征表述结构,将若干距离紧密的兴趣点集成一个整体,降低了特征的数量,减少了后续特征描述分类的计算复杂度。同时采用 BFSN 方法生成的超兴趣点中不仅包含同一帧上的兴趣点,也包含邻近帧的兴趣点,在考虑兴趣点间空间分布的同时,也考虑了时间上的联系,取得了较好的识别结果。

## 参考文献 (References)

- [1] Efros A A, Berg A C, Mori G, et al. Recognizing action at a distance[C]//Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003:726-733.
- [2] Yilmaz A, Shah M. Matching actions in presence of camera motion[J]. Computer Vision and Image Understanding, 2006, 104(2-3):221-231.

- [3] 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: IEEE, 2005:65-72.
- [4] Laptev I. On space-time interest points[J]. International Journal of Computer Vision, 2005, 64(2): 107-123.
- [5] Bregonzio M, Gong S, Xiang T. Recognising action as clouds of space-time interest points[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Florida, USA: IEEE, 2009:1948-1955.
- [6] 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.
- [7] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [8] Klaser A, M Marszalek, C Schmid. A spatio-temporal descriptor based on 3D-gradients[C]//Proceedings of British Machine Vision Conference. Leeds, England: University of Leeds, 2008: 955-1004.
- [9] Gao Z, Chen M, Hauptmann A, et al. Comparing evaluation protocols on the KTH dataset[C]//Proceedings of the 1st International Conference on Human Behavior Understanding. Istanbul, Turkey: Springer-Verlag, 2010:88-100.
- [10] Nater F, Grabner H, Gool L Van. Exploiting simple hierarchies for unsupervised human behavior analysis[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California, USA: IEEE, 2010: 2014-2021.
- [11] Liu S, Liu J, Zhang T, et al. Human action recognition in videos using hybrid motion features[C]// Proceedings of the 16th International Multimedia Modeling Conference. Chongqing, China: Springer-Verlag, 2010: 411-421.
- [12] Sun X, Chen M, Hauptmann A. Action recognition via local descriptors and holistic features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Florida, USA: IEEE, 2009:58-65.
- [13] Liu J, Ali S, Shah M. Recognizing human actions using multiple features[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [14] Wang C X, Liu Y, Li W Q. Research of unsupervised posture modeling and action recognition based on spatial-temporal interesting points[J]. Acta Electronica Sinica, 2011, 39(8):1751-1756. [王传旭, 刘云, 厉万庆. 基于时空特征点的非监督姿态建模和行为识别的算法研究[J]. 电子学报, 2011, 39(8): 1751-1756.]
- [15] Yan X, Luo Y. Recognizing human actions using a new descriptor based on spatial-temporal interest points and weighted-output classifier[J]. Neurocomputing, 2012, 87: 51-61.



- [16] Zhang Z, Hu Y, Chan S, et al. Motion context: a new representation for human action recognition[C]//Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer-Verlag, 2008: 817-829.
- [17] Zhao Z, Elgammal A. Human activity recognition from frame's spatiotemporal representation[C]// Proceedings of the 19th International Conference on Pattern Recognition. Tampa, Florida, USA: IEEE, 2008: 1-4.
- [18] Thureau C, V Hlaváč. Pose primitive based human action recognition in videos or still images[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [19] Liu J, Shah M. Learning human actions via information maximization [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [20] Qian J B, Dong Y S. A clustering algorithm based on broad first searching neighbors[J]. Journal of Southeast University: Natural Science Edition, 2004, 34(1): 109-112. [钱江波, 董逸生. 一种基于广度优先搜索邻居的聚类算法[J]. 东南大学学报: 自然科学版, 2004, 34(1): 109-112.]
- [21] Blank M, Gorelick L, Shechtman E, et al. Actions as space-time shapes[C]//Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005: 1395-1402.
- [22] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition[C]//Proceedings of the 15th International Conference on Multimedia. Augsburg, Germany: ACM, 2007: 357-360.
- [23] Yuan H J, Wang C R. Human action recognition using Markov random walk based semi-supervised learning [J]. Journal of Computer-Aided Design & Computer Graphics, 2011, 23(10): 1749-1757. [袁和金, 王翠茹. 人体行为识别的 Markov 随机游走半监督学习方法[J]. 计算机辅助设计与图形学学报, 2011, 23(10): 1749-1757.]
- [24] Lin Z, Jiang Z, Davis L S. Recognizing actions by shape-motion prototype trees[C]//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009: 444-451.