

# 基于光流模值变化频率的群体骚乱行为检测方法

林杰<sup>1</sup> 林拉<sup>2</sup>

(华南理工大学公共管理学院 广州 510640)<sup>1</sup> (华南师范大学计算机学院 广州 510631)<sup>2</sup>

**摘要** 群体骚乱行为突发性强、破坏性大,是视频监控关注的重点。研究群体骚乱行为的智能检测方法有助于提高视频监控的智能化水平。现有行为检测方法在检测群体骚乱行为时虚警率较高,实用性较差。因此,依据群体骚乱行为发生时光流模值变化大的特性,提出了一种基于光流模值变化频率的群体骚乱行为检测方法。该方法首先计算视频中每一帧图像上各像素点的光流;然后自适应求取反映像素点光流变化大小的二值映射图;接着分区块计算视频片段上的光流变化频率直方图,构建行为描述子;最后采用线性支持向量机进行特征训练与分类。实验表明,所提方法在检测群体骚乱行为时虚警率和漏警率低、识别率高,可广泛用于智能视频监控领域。

**关键词** 群体骚乱行为,行为检测,光流,支持向量机

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.5.054

## Detection Method for Group Riot Activity Based on Change Frequency of Optical Flow's Magnitude

LIN Jie<sup>1</sup> LIN La<sup>2</sup>

(School of Public Management, South China University of Technology, Guangzhou 510640, China)<sup>1</sup>

(School of Computer, South China Normal University, Guangzhou 510631, China)<sup>2</sup>

**Abstract** Group riot activity is the focus of video surveillance, which is emergent and destructive. To detect the group riot activity intelligently is helpful for improving the intelligence level of video surveillance. The common activity detection methods have high false alarm rate while detecting group riot activity from videos. In this paper, a detection method for group riot activity was proposed based on change frequency of optical flow's magnitude. Firstly, the optical flow of each pixel on each frame in a video is calculated. Secondly, a binary map is obtained to reflect the changes of optical flow aptively. Then an activity descriptor by dividing an image into several blobs and the change frequency histogram of optical flow is calculated independently. Finally, the activity features are trained and classified by using a linear support vector machine. Experiments show that the new method has low false alarm rate, low missed alarm rate and high genuine acceptance rate, while detecting group riot activity. So it can be widely used in the field of intelligent video surveillance.

**Keywords** Group riot activity, Activity detection, Optical flow, SVM

随着视频监控设备的普及,仅依靠人工监视难以充分发挥视频监控的价值,开发具有智能预警功能的视频监控系统成为迫切需求。行为识别是指通过对视频数据的处理、分析与理解,识别场景中是否存在某一种或几种行为的过程,对提高视频监控系统的智能化水平有重要意义,是当前计算机视觉领域的研究热点<sup>[1]</sup>。

群体骚乱是指在某一特定场合或局部范围发生的扰乱和冲击社会正常秩序的群体行为,是公众在非正常状态或危机状态下的一种行为方式,具有突发性强、交互感染速度快、破坏性大的特点,是视频监控关注的重点。因此,研究群体骚乱行为的智能检测方法具有重要意义。

目前行为识别领域研究较多的是暴力行为识别(如斗殴)<sup>[2]</sup>和可疑行为(如徘徊、聚集)识别<sup>[3-7]</sup>,常用方法主要包括基于局部兴趣点的方法<sup>[8-11]</sup>和基于全局特征的方法<sup>[12-15]</sup>。前者首先要检测视频片段中的时空关键点,然后提取关键点的时空描述子,最后采用词袋模型等方法进行特征分类,可以较

好地实现局部动作的检测,但当场景中时空关键点太少或者运动目标太多时,此类方法的虚警率较高;后者提取整个视频帧的全局统计特征,通过检测这些特征的时空变化来识别不同的行为,此类方法受目标相互遮挡及目标提取精度的影响小,但受环境变化(如雨雪天气等)的影响较大。

对于群体骚乱行为,由于行为发生时场景中运动目标很多,不宜采用基于局部兴趣点的方法,常采用基于全局统计特征的方法。其中,如何提取能有效描述群体骚乱行为的特征是行为识别的关键。目前由于视频监控系统往往缺失音频信息,因此无音频特征可用;而系统全天候工作的需求以及实际监控视频分辨率低的现状,导致颜色特征的可靠性差;且群体骚乱行为发生时,由于群体成员之间的相互遮挡,难以获取精细的动作特征;进一步地,考虑到实际监控系统的算法的时效性要求,不适宜采用高层次的形状和运动分析,常借鉴动态纹理识别的方法<sup>[16]</sup>,通过提取大量简单的低级特征,来描述复杂的群体行为。目前行为识别领域经典的特征描述子有

到稿日期:2015-04-10 返修日期:2015-05-12 本文受广东省科技厅项目(2013B070207001)资助。

林杰(1981-),男,硕士,助理实验师,主要研究方向为数据挖掘以及计算机应用;林拉(1953-),男,副教授,主要研究方向为软件工程。

HOG<sup>[17]</sup>、HOF<sup>[18]</sup>和LPT<sup>[19]</sup>,但对于群体骚乱行为,由于群体成员多,成员行为不一致,采用这些特征描述子描述行为时难以有效区分群体骚乱行为和其他正常的群体行为,虚警率较高。

本文依据群体骚乱行为发生时光流模值变化大的特性,提出一种基于光流模值变化频率的群体骚乱行为检测方法。首先计算每一帧图像上各个像素点的光流;然后计算相邻帧各像素点的光流变化量,并进行自适应分割,得到一个反映像素点光流变化大小的二值映射图;接着将图像划分为互不重叠的宏块,并统计视频片段上各宏块的平均光流变化频率;在此基础上,统计各宏块光流变化的频率直方图,组建成行为描述子;最后采用线性支持向量机进行特征训练与分类,实现群体骚乱行为的鲁棒检测。

## 1 光流与行为

光流是空间运动物体在观测成像面上的像素运动的瞬时速度,代表了局部运动的方向和模值,可以利用视频中的像素强度数据的时域变化和相关性确定各自像素位置的运动。因此,光流可以反映目标的运动状况,对描述行为有一定意义。Hayashi 等基于光流场的模值和方向特征实现了电梯内暴力行为的识别<sup>[20]</sup>。与普通行为相比,暴力行为发生时光流场的模值更大,方向分布更散乱。如图 1 所示,图 1(a1)为非暴力场景帧图像,图 1(a2)为暴力场景帧图像及光流特征,图 1(b1)、(b2)分别为图 1(a1)、(a2)对应的光流直方图(其中光流方向共 8 个,模值数量级为 4),图 1(c1)、(c2)分别为图 1(a1)、(a2)对应的光流变化标准差(纵坐标代表了光流方向的标准偏差,横坐标代表了模值的标准偏差)。从图中可以明显看出,当暴力行为发生时,图像中多数位置的光流模值较大,光流方向分布散乱;光流方向和模值的偏差都大于正常行为,且这些性质受摄像机的安装位置、角度以及环境变化的影响较小。Hayashi 等基于光流的这一特性实现了电梯内暴力行为的检测。

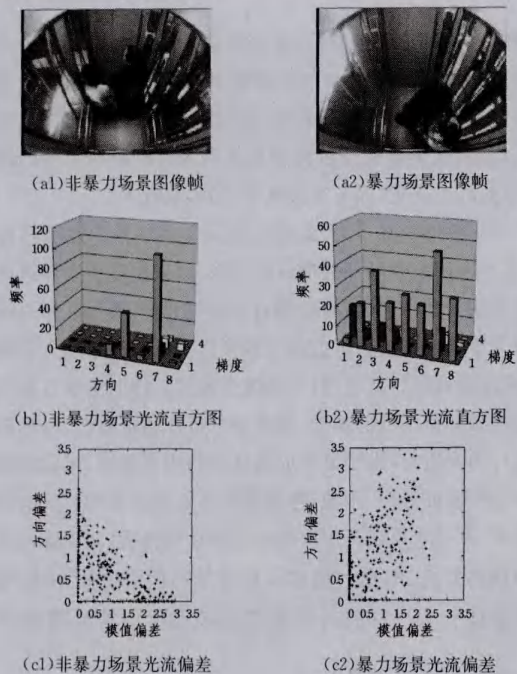


图 1 光流与行为的关系

群体骚乱行为与暴力行为不同,在群体发生骚乱行为之前,可能由于各成员的行为不一致导致光流的方向分布散乱,而当群体骚乱行为发生时,可能由于群体成员之间相互拥挤导致部分行为作用区域光流模值较小。而且群体行为中成员很多,不同成员与摄像机的距离不同,光流模值的大小难以有效反映运动的快慢。因此,光流的模值大小和方向分布特征难以有效区分群体骚乱行为和正常行为。但是,群体骚乱行为发生时行为作用区域的光流模值变化很大,而且模值变化量受摄像机角度、焦距以及目标距离摄像机远近等因素的影响较小。因此,本文基于光流模值变化特性来检测群体骚乱行为。

## 2 群体骚乱行为检测算法

依据群体骚乱事件发生时光流模值的突变特性,本文提出一种基于光流模值变化的群体骚乱行为检测算法,算法流程如图 2 所示。

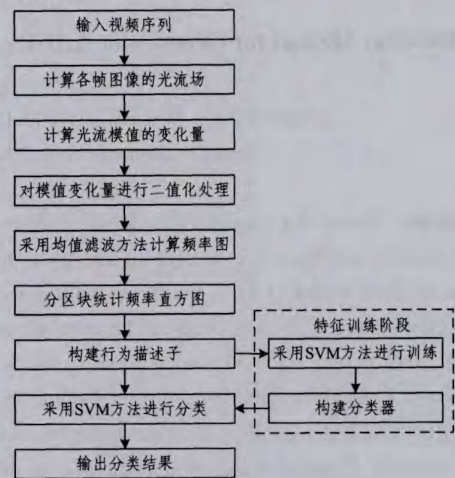


图 2 群体骚乱行为检测流程

具体步骤如下:

- Step1 计算视频片段中除第一帧之外的其他帧图像上各像素点的光流;
- Step2 除前两帧图像外,计算其他图像上各像素点处光流模值的变化量;
- Step3 除前两帧图像外,对于每一帧图像上各像素点的光流模值变化量,进行自适应二值化处理,得到二值映射图;
- Step4 对视频片段中的所有二值映射图进行均值滤波,得到反映光流模值变化频率的频率图;
- Step5 将均值映射图划分为互不重叠的宏块,计算各宏块的频率直方图,并进行归一化处理;
- Step6 将各宏块的归一化频率直方图组合,得到行为描述子;
- Step7 采用线性支持向量机对行为描述子进行分类,得到行为检测结果。

### 2.1 光流计算

光流计算的一般方法是采用光流约束方程<sup>[21]</sup>:

$$\Delta I \cdot V_p + I_t = 0 \quad (1)$$

其中,  $\Delta I = (I_x, I_y)$  表示图像中像素点  $p$  处的梯度,  $V_p = (u, v)$  表示像素  $p$  的光流,  $I_t$  表示前后两帧图像的灰度变化率。

光流计算的关键是求取图像各像素点的梯度,目前光流

计算采用的梯度计算方法有多种,如 Horn 算子、Prewitt 算子、Sobel 算子、Barron 算子等。其中经典的是 Horn 算子,该算子通过求取相邻像素点之间的灰度一阶差分计算像素点的梯度,如式(2)所示。

$$\begin{cases} I_x = \frac{1}{4} \{ I_{i,j+1,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i+1,j,k} + \\ I_{i,j+1,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i+1,j,k+1} \} \\ I_y = \frac{1}{4} \{ I_{i+1,j,k} - I_{i,j,k} + I_{i+1,j+1,k} - I_{i,j+1,k} + \\ I_{i+1,j,k+1} - I_{i,j,k+1} + I_{i+1,j+1,k+1} - I_{i,j+1,k+1} \} \\ I_t = \frac{1}{4} \{ I_{i,j,k+1} - I_{i,j,k} + I_{i+1,j,k+1} - I_{i+1,j,k} + \\ I_{i,j+1,k+1} - I_{i,j+1,k} + I_{i+1,j+1,k+1} - I_{i+1,j+1,k} \} \end{cases} \quad (2)$$

其中,  $I_{i,j,k}$  表示视频片段第  $k$  帧图像上像素点  $(i, j)$  处的灰度值。

Horn 等提出了光流的平滑性约束条件。假设图像上任一像素点的光流并不是独立的,光流在整个图像范围内是平滑变化的。基于这一假设,Horn 等采用正则化方法,通过求取式(3)的最小值来计算光流。

$$\epsilon^2 = \iint (a^2 \epsilon_c^2 + \epsilon_b^2) dx dy \quad (3)$$

其中,  $\epsilon_b$  表示由图像亮度保持不变假设引起的误差,  $\epsilon_c$  表示由光流平滑性假设引起的误差,  $\alpha$  表示  $\epsilon_b$  与  $\epsilon_c$  之间的相对权重。  $\epsilon_b$  和  $\epsilon_c$  的求解公式如下:

$$\epsilon_b = I_x u + I_y v + I_t \quad (4)$$

$$\epsilon_c^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \quad (5)$$

## 2.2 行为描述子

对于输入的视频片段,按照 2.1 节描述的光流计算方法,计算除第一帧之外的其他帧图像上各像素点的光流。设第  $t$  帧图像上像素点  $(x, y)$  处的光流为  $(u_{x,y,t}, v_{x,y,t})$ ,则光流的模值为

$$m_{x,y,t} = \sqrt{u_{x,y,t}^2 + v_{x,y,t}^2} \quad (6)$$

第  $t$  帧图像上像素点  $(x, y)$  处的光流模值变化量为

$$\Delta m_{x,y,t} = |m_{x,y,t} - m_{x,y,t-1}| \quad (7)$$

采用自适应二值化处理,得到光流模值变化量的二值映射图为

$$b_{x,y,t} = \begin{cases} 1, & \text{如果 } \Delta m_{x,y,t} \geq T_1 \text{ 且 } m_{x,y,t} \geq T_2 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

其中,  $T_1$  和  $T_2$  为自适应阈值,求解方法为:设图像的宽度和高度分别为  $W$  和  $H$ ,则  $T_1$  和  $T_2$  分别按式(9)和式(10)自适应取值。

$$T_1 = \frac{1}{W \times H} \sum_{y=0}^{H-1} \sum_{x=0}^{W-1} \Delta m_{x,y,t} \quad (9)$$

$$T_2 = \begin{cases} 1, & y < 0.1H \\ 9 \times \frac{y - 0.1H}{0.8H} + 1, & 0.1H \leq y < 0.9H \\ 10, & \text{otherwise} \end{cases} \quad (10)$$

这样,除前两帧图像外,对视频片段的其他帧图像都可以求取一个反映光流模值变化的二值映射图  $b_t$ 。

对视频片段中所有二值映射图进行均值滤波处理,可以得到表征场景中光流模值变化频率的频率图:

$$\bar{b}_{x,y} = \frac{1}{L} \sum_t b_{x,y,t} \quad (11)$$

其中,  $L$  表示视频片段中二值映射图的数量。

在同一场景下,可能在不同位置同时存在着群体骚乱行为和正常行为。这样,采用整幅图像的光流模值变化频率不易有效区分群体骚乱行为和正常行为。为此,本文将图像划分为  $M \times N$  个互不重叠的宏块,在每个宏块上独立统计光流模值变化频率的直方图。

在实际计算时,为了增大光流模值变化频率的动态范围和避免出现小数,式(11)改写为:

$$\bar{b}_{x,y} = \text{floor}\left(\frac{100}{L} \sum_t b_{x,y,t}\right) \quad (12)$$

其中,函数  $\text{floor}(x)$  表示取与  $x$  距离最近且小于  $x$  的整数,这里并不会增加误差。

记宏块内光流模值变化频率的级数为  $X$ ,也即其动态范围为  $[0, X-1]$ ,则宏块内光流模值变化频率为  $i$  的概率为

$$P(i) = \frac{\sum_x \sum_y (\bar{b}_{x,y} = i)}{M \times N} \quad (13)$$

用  $\text{Hist}[X]$  表示宏块的光流模值变化频率直方图,其求解方式为

Step1 初始化,令

$$\text{Hist}[i] = 0, i = 0, 1, \dots, X-1 \quad (14)$$

Step2 遍历宏块中所有像素,使

$$\text{Hist}[\bar{b}_{x,y}] = \text{Hist}[\bar{b}_{x,y}] + 1 \quad (15)$$

Step3 归一化,使

$$\text{Hist}(i) = \frac{\text{Hist}(i)}{M \times N}, i = 0, 1, \dots, X-1 \quad (16)$$

最后,由  $M \times N$  个光流模值变化频率直方图组成行为描述子,记为 MCFH (Magnitude Change Frequency Histogram):

$$\text{MCFH} = [\text{Hist}_1 \text{ Hist}_2 \dots \text{Hist}_{M \times N}]^T \quad (17)$$

为了保证直方图组合时各直方图的维数相等,需要将不同宏块的光流模值变化直方图的级数归一化到相同大小。本文采用简单的线性变换方法进行归一化处理,设归一化后的级数为  $Z$ ,则归一化后的光流模值变化频率为

$$\tilde{b}_{x,y} = \text{floor}\left(\frac{Z}{X} \bar{b}_{x,y}\right) \quad (18)$$

## 2.3 行为分类

对于给定的视频片段,按 2.2 节所述的方法提取行为描述子 MCFH。统计发现,对于含群体骚乱行为的视频片段和正常的视频片段,提取的 MCFH 向量之间具有线性可分性,可采用简单高效的线性支持向量机进行分类<sup>[22]</sup>,决策函数为

$$f(x) = \text{sgn}\left\{\sum_{i=1}^n a_i y_i K(x_i, x) + b\right\} \quad (19)$$

核函数采用线性核函数,为

$$K(x_i, x) = x_i \cdot x \quad (20)$$

其中,  $\{x_i, y_i\}$  表示参与训练的样本集,  $i = 1, 2, \dots, n$  为参与训练的样本数。具体地,  $x_i$  为从第  $i$  个视频片段提取的行为描述子 MCFH;  $y_i \in \{1, -1\}$  表示样本类别,  $y_i = 1$  表示该视频片段包含群体骚乱行为,  $y_i = -1$  表示该视频片段不含群体骚乱行为。通过训练过程,可以求取 Lagrange 乘子参数  $a_i$  和偏移量参数  $b$ ,构建分类器。在行为检测阶段,对于待检测的视频片段,提取行为描述子 MCFH(用  $x$  表示),并将其代入式(19),若  $f(x) = 1$ ,则表示输入视频片段中包含群体骚乱行为,否则表示视频片段中无群体骚乱行为。

### 3 实验结果与分析

#### 3.1 实验数据与评价指标

目前国际上还没有公开的群体骚乱行为数据库,因此我们构建了测试数据库。为了尽可能保证测试视频与真实监控场景相符,本文的测试视频全部通过收集整理百度视频和 YouTube 提供的监控视频而得到。在对测试视频的搜集过程中,不限制监控场景和环境,也不限制摄像机的拍摄角度、焦距等参数,以便测试算法的普适性。在测试视频整理过程中,先人工标记包含群体骚乱行为的视频以及视频中群体骚乱行为的起止时间,然后人工裁剪视频片段,最后采用相同的视频分辨率、帧率和编码格式对视频片段进行重新编码,存储为 AVI 文件,形成测试数据库,数据库的相关信息如下。

- 视频分辨率:640×480
- 编码格式:H. 264
- 帧率:25fps
- 视频片段总数:1634
- 正样本总数:642
- 负样本总数:992
- 最短视频片段时长:1. 6s
- 最长视频片段时长:10. 3s
- 平均视频片段时长:6. 6s

在算法评测阶段,采用行为识别领域常用的 4 个评价指标<sup>[23]</sup>,即

- 虚警率(False Alarm Rate, FAR):负样本被误认为正样本的数量与样本总数的比值,值越小越好;
- 漏警率(Missed Alarm Rate, MAR):正样本被误认为负样本的数量与样本总数的比值,值越小越好;
- 识别率(Genuine Acceptance Rate, GAR):正确识别的样本数量与样本总数的比值,值越大越好;
- 平均处理时间(Average Detection Time, ADT):从输入视频片段到输出识别结果所耗费的时间与视频片段中帧数的比值,值越小越好。

#### 3.2 参数分析

本文所提出的群体骚乱行为检测算法主要有以下两个参数。

- $Z$ :归一化后的光流模值变化频率直方图的级数;
- $M \times N$ :宏块尺寸。

参数  $Z$  取值越大,频率直方图的统计越精确,但运算复杂度越大,且算法的漏警率可能偏高; $Z$  取值越小,运算复杂度越小,但虚警率可能偏高。图 3 统计了本文数据库各视频片段中整幅图像的光流模值变化频率的级数分布。

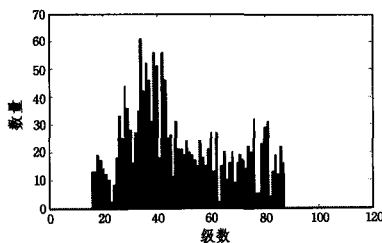


图 3 光流模值变化频率级数分布

在本文实验中,参数  $Z$  的取值为所统计的级数分布直方图的均值与方差之和,即  $Z=58$ 。

为了便于计算,常令宏块的宽度和高度相等。图 4 显示了宏块宽度与算法识别率的关系(实验过程中随机从正、负样本数据库中各选取一半数量样本进行训练,剩余样本进行测试)。从图中可以看出,当  $M=N=4$  时算法的识别率最高。

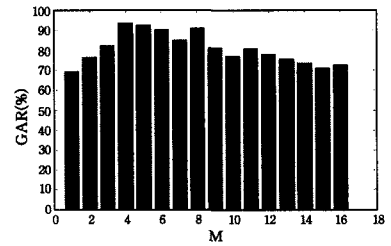


图 4 宏块尺寸与识别率的关系

#### 3.3 实验结果对比

行为描述子 MCFH 是本文算法的核心。为了对比 MCFH 识别群体骚乱行为的性能,在实验中将 MCFH 与经典的 HOG<sup>[17]</sup>、HOF<sup>[18]</sup>和 LPT<sup>[19]</sup>描述子进行对比,分类方法都采用线性支持向量机。表 1 给出了不同方法在本文数据库下的性能指标。其中,实验运行平台是 Intel 酷睿 i7 计算机,Windows 7 操作系统、内存 4GB,时间单位为毫秒。

表 1 算法识别性能对比

算法	FAR	MAR	GAR	ADT
HOG	20.67%	11.99%	82.74%	216
HOF	18.45%	30.84%	82.80%	287
LPT	29.94%	14.80%	76.01%	166
MCFH	7.36%	6.07%	93.15%	233

由表 1 可见,本文方法的虚警率、漏警率和识别率指标都高于其他方法,尤其是虚警率指标优势明显;而且本文方法的运算效率与 HOG 相当,尽管比 LBT 方法运算效率低,但识别性能比 LBT 方法高 10% 以上。

**结束语** 本文提出了一种基于光流模值变化频率的群体骚乱行为检测方法,其可以实现监控视频中群体骚乱行为的智能检测。首先,分析了光流与行为的关系,阐明了群体骚乱行为发生时光流模值快速变化的特性;接下来,讨论了基于光流模值变化频率检测群体骚乱行为的详细过程,主要介绍了光流计算方法、行为描述子构建方法和行为分类方法;最后,通过仿真实验,讨论了算法中相关参数对算法性能的影响,并对比了本文算法与现有行为识别领域的经典算法在检测群体骚乱行为方面的性能。实验表明,本文提出的基于光流模值变化频率的群体骚乱行为检测方法的虚警率和漏警率低、识别率高,可广泛用于智能视频监控领域。

#### 参考文献

- [1] Austin O T A. Human activity analysis: a review[J]. ACM Computing Surveys, 2011, 43(3): 194-218
- [2] Thonnat M. Group behavior recognition with multiple cameras [C]//IEEE Winter Conference on Applications of Computer Vision. 2002: 1-8
- [3] Hu Zhi-lan, Jiang fan, Wang Gui-jin, et al. Anomaly detection based on motion direction[J]. Acta Automatica Sinica, 2008, 34(11): 1348-1357 (in Chinese)
- [4] 胡芝兰,江帆,王贵锦,等. 基于运动方向的异常行为检测[J]. 自动化学报, 2008, 34(11): 1348-1357
- [4] Liu J, Kuipers B, Savarese S. Recognizing human actions by at-

- tributes[J]. University of Michigan, 2011, 42(7): 3337-3344
- [5] Han Lei, Li Jun-feng, Jia Yun-de. Human Interaction Recognition Using Spatio-Temporal Words[J]. Chinese Journal of Computers, 2010, 33(4): 776-784 (in Chinese)  
韩磊, 李君峰, 贾云得. 基于时空单词的两人交互行为识别方法[J]. 计算机学报, 2010, 33(4): 776-784
- [6] Nibbles J C, Feifei L. A hierarchical model of shape and appearance for human action classification[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2007: 1-8
- [7] Yuan He-jin, Wang Cui-ru. Human action recognition using Markov random walk based semi-supervised learning [J]. Journal of Computer-Aided Design & Computer Graphics, 2011, 23(10): 1749-1757 (in Chinese)  
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
- [8] Wang Chuan-xu, Liu Yun, Li Wan-qing. Research of Unsupervised Posture Modeling and Action Recognition Based on Spatial-Temporal Interesting Points [J]. Acta Electronica Sinica, 2011, 39(8): 1751-1756 (in Chinese)  
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
- [9] Mohamed B, Kaaniche, Franos. Recognizing Gestures by Learning Local Motion Signatures of HOG Descriptors[J]. Ann Anal and Mahn Nllgn Ranaon on, 2012, 34(11): 2247-2258
- [10] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features [C]//2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. 2005: 65-72
- [11] Du You-tian, Chen Feng, Xu Wen-li. Approach to Human Activity Multi-scale Analysis and Recognition Based on Multi-layer Dynamic Bayesian Network [J]. Acta Automatica Sinica, 2009, 35(3): 225-232 (in Chinese)  
杜友田, 陈峰, 徐文立. 基于多层动态贝叶斯网络的人的行为多尺度分析及识别方法[J]. 自动化学报, 2009, 35(3): 225-232
- [12] Saad A, Mubarak S. Human action recognition in videos using kinematic features and multiple instance learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(2): 288-303
- [13] Rodriguez M, Ali S, Kanade T. Tracking in unstructured crowded scenes [C]//2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 1389-1396
- [14] Basri R, Hassner T, Zelnik-Manor L. Approximate nearest subspace search [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(2): 266-278
- [15] Ke Y, Sukthankar R, Hebert M. Event Detection in Crowded Videos [C]//IEEE 11th International Conference on Computer Vision, 2007 (ICCV 2007). IEEE, 2007: 1-8
- [16] Kellokumpu V, Zhao G, Pietikäinen M. Human Activity Recognition Using a Dynamic Texture Based Method [J]. In BMVC, 2008
- [17] Bernejo N E, Deniz S O, Bueno G G, et al. Violence detection in video using computer vision techniques [C]//International Conference on Computer Analysis of Images and Patterns. 2011: 332-339
- [18] Orit K G, Tal H, Lior W, et al. The Action Similarity Labeling Challenge [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(3): 615-621
- [19] Yeffet L, Wolf L. Local Trinary Patterns for Human Action Recognition [C]//IEEE International Conference on Computer Vision, 2009: 492-497
- [20] Hayashi K, Seki M, Hirai T, et al. Real-time violent action detector for elevator [J]. Optomechatronic Machine Vision, 2005: 60510R-60510R-8
- [21] Horn B K P, Schunck B G. Determining Optical Flow [J]. Artificial Intelligence, 1980, 17(81): 185-203
- [22] Kim H, Pang S, Je H, et al. Constructing support vector machine ensemble [J]. Pattern Recognition, 2003, 36(12): 2757-2767
- [23] Wang Yuan-yuan, Wang Bin. The Conditional Random Fields Methods for Human Action Recognition [J]. Journal of Chongqing University of Technology (Natural Science), 2013, 27(6): 93-99, 105 (in Chinese)  
王媛媛, 王斌. 人体行为识别的条件随机场方法 [J]. 重庆理工大学学报 (自然科学), 2013, 27(6): 93-99, 105

(上接第 273 页)

- [8] Berg T L, Berg A C, Shih J. Automatic Attribute Discovery and Characterization from Noisy Web Data [C]//Proceedings of European Conference on Computer Vision. 2010: 663-676
- [9] Kiapour H, Yamaguchi K, Berg A C, et al. Hipster Wars: Discovering Elements of Fashion Styles [C]//Proceedings of European Conference on Computer Vision, 2014: 472-488
- [10] Rebecca. Domain-Independent Captioning of Domain-Specific Images [C]//Proceedings of North American Association for Computational Linguistics. 2013: 69-76
- [11] Kiros R, Zemel R S, Salakhutdinov R. Multimodal Neural Language Models [C]//Proceedings of International Conference on Machine Learning. 2014
- [12] Bo L, Ren X, Fox D. Kernel Descriptors for Visual Recognition [C]//Proceedings of Advances in Neural Information Processing Systems, 2010: 1734-1742
- [13] Bo L, Ren X, Fox D. Efficient Match Kernels between Sets of Features for Visual Recognition [C]//Proceedings of Advances in Neural Information Processing Systems. 2009: 135-143
- [14] Sivaram G, Hermansky H. Sparse Multilayer Perceptron for Phoneme Recognition [J]. IEEE Trans. Audio, Speech, & Language Proc., 2012, 20(1): 23-29
- [15] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2010: 3360-3367
- [16] Papineni K, Roukos S, Ward T, et al. Bleu: a method for automatic evaluation of machine translation [C]//Proceedings of the Annual Meeting on Association for Computational Linguistics. 2002: 311-318
- [17] Gao Sheng-hua, Tsang W-H, Chia L T. Sparse Representation With Kernels [J]. IEEE Transactions on Image Processing, 2013, 22(2): 423-434
- [18] Maas A L, Daly R E, Pham P T, et al. Learning Word Vectors for Sentiment Analysis [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies. 2011: 142-150