

人群行为分析研究进展

陈昌红 朱秀昌

(南京邮电大学通信与信息工程学院 南京 210003)

摘要 人群行为分析是计算机视觉领域一个新兴的研究方向,在智能视频监控、虚拟现实、视频检索等方面有着广泛的应用前景和巨大的经济价值。主要从数据库和分析方法两个方面对人群行为分析的研究进展进行综述。人群行为数据库主要包括行为数据库和监控数据库两种,对这两类数据库中比较典型的库进行了总结,并从人群行为的分解方法和识别方法两方面对人群行为分析方法进行了详细总结。将分解方法分为3种,分别指出了各种方法的优缺点。将人群行为的识别方法分为基于统计的方法和基于描述的方法两种,对它们进行了详细比较,并指出了识别方法的发展趋势。总结了人群行为分析的潜在问题,并展望了其发展前景。

关键词 人群行为分析,数据库,人群行为分解,人群行为识别

中图分类号 TP391.41 **文献标识码** A

Research Envelopment on Human Group Activities Analysis

CHEN Chang-hong ZHU Xiu-chang

(College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract Human group activities analysis has become a new research interest of computer vision. It has broad application prospects and great economic value in many aspects, such as intelligent video surveillance, virtual reality, video retrieval and so on. This paper made a survey on the human group activities analysis from group databases and analysis algorithms. Firstly, the group databases mainly consist of behavior databases and surveillance databases. This paper summarized the typical databases of these two classes. Secondly, we summed up the human group activities analysis algorithms from decomposition methods and recognition methods of group activities. The decomposition methods were divided into three kinds and their strengths and weaknesses were pointed out, separately. The recognition algorithms of group activities were divided into statistical approaches and description-based ones. We compared them in detail and pointed out the development tendency. At last, we summarized the potential problems of human group activities analysis and looked forward to its future.

Keywords Human group activities analysis, Databases, Decomposition methods, Human group activities recognition

1 引言

由于在智能监控、虚拟现实、视频检索等方面有着广泛的应用前景和巨大的经济价值,人体行为分析激发了广大研究者和相关企业的浓厚兴趣,成为了以美国、英国为代表的许多国家的重点项目的研究重点。经过几十年的发展,目前人体行为的视觉分析已经取得了令人瞩目的成果,尤其是在个体行为分析和识别方面已经达到了较高的水平^[1-4]。在此基础上,高层次的事件监测和人群行为分析得到了一定的发展,虽然目前相关研究还比较少,但已呈燎原之势,成为了新的研究热点。

人群行为是指多个人行为的集合(从广义上来讲,两个人之间行为的交互也可以归为人群行为),可分为“集体性”的和“多元性”的两种。“集体性”的人群行为指所有成员具有相同的运动趋势,比如游行等,这类行为是比较简单的,用其中一

个成员的行为就可表征整个群体的行为^[5,6]。“多元性”的人群行为指各成员的行为不尽相同,有行为变化或子群间交互等,这类行为比较复杂,也更常见,对这类行为的研究更具有普适性和应用价值,这也是群行为研究中最活跃的部分。本文将重点对“多元性”的人群行为进行分析进行总结。虽然已经有很多行为分析方面的综述文献^[7-9],但这些文章只对个体行为分析的方法进行了总结和分类。在 Aggarwal 等的文献^[10]中对人群行为分析的方法有所涉及,但由于文章覆盖面较广,没有对这方面的方法进行系统的分析。本文将对数据库、人群行为分析的主要方法、面临的问题和发展趋势进行全面总结。

2 人群行为数据库

目前,比较常用的人群行为的数据库可以分为3类,一类是用于学术研究的行为数据库,它们包含多组不同参与者的

到稿日期:2011-12-15 返修日期:2012-03-21 本文受国家自然科学基金项目(61172118, 61001152),江苏省高校自然科学基金项目(11KJB510012),江苏省自然科学基金(BK2010523),南京邮电大学校科研基金(NY210073)资助。

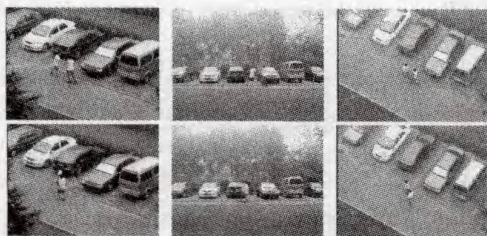
陈昌红 女,博士,讲师,主要研究方向为视频分析、模式识别, E-mail: chenchrh@njupt.edu.cn; 朱秀昌 男,教授,博士生导师,主要研究方向为多媒体信息处理。

相同交互行为,比如握手、拥抱、跟踪、打架等,这类数据库以中科院的 CASIA 行为数据库^[11]和 UT-interaction 数据库^[12]为代表;另一类是面向应用的监控数据库,这类数据库从实际场合拍摄得到,比如机场、地铁站等,以 PETS 数据库为代表;第三类是从电影或体育比赛视频中收集得到的数据库,以 HOHA 数据库^[19]和足球队数据库为代表。下面对这 3 类数据库做详细介绍。

2.1 行为数据库

行为数据库主要是为了验证算法在简单交互动作上的识别性能而拍摄的,不同的个体组合会有相同的行为。这类数据库以中科院的 CASIA 行为数据库^[11]和 UT-interaction 数据库^[12]为代表,主要还是针对两个个体间的交互行为。

CASIA 行为数据库^[11]有两个行为个体,包括侧面(angle view)、水平(horizontal view)和垂直(top down view) 3 个视角下的 7 种交互行为视频,即抢劫、打架、跟踪、跟踪-同行、相遇-分离、相遇-同行和赶超,同一种行为有 4 组个体组合执行,视频帧率为 25fps,分辨率为 320×240,视频长度在 5s 到 30s 之间。图 1 显示了打架、相遇-同行两种行为在 3 个视角下的图像示例。



(a) 侧面视角 (b) 水平视角 (c) 垂直视角
三列分别代表三种视角,第一行为打架,第二行为相遇-同行。

图 1 CASIA 行为数据库图像示例

UT-interaction 数据库^[12]包含 6 类交互行为,分别为握手、用手指对方、拥抱、推搡、踢人、拳击,共包含 20 个长度约为 1 分钟的视频序列,每个视频至少包含一次上述交互行为,平均含有 8 个动作。视频包含多于 15 种不同穿着的参与者,帧率为 30fps,分辨率为 720×380。根据场合不同,将上述视频分为两类,每类包含 10 个视频,一类发生在停车场,另一类发生在草坪。图 2 显示了在两种场合下推搡和拥抱两种行为的图像。背景分为静止和有干扰两种,行为的执行者有一组和两组之分,有干扰和两组执行者的称为高级交互行为(high-level human interaction),如图 3 所示。与 CASIA 行为数据库相比,UT-interaction 数据库背景比较简单,拍摄主体很清晰。



(a) 推搡 (b) 拥抱
第一行发生在停车场,第二行发生在草坪。

图 2 UT-interaction 数据库 punch, hug 两种行为的图像示例



图 3 UT-interaction 数据库高级交互行为的图像示例

2.2 监控数据库

监控数据库由实际采集到的真实监控视频组成,因此行为没有重复性,即不存在多个个体组合有相同行为的情况,缺乏普适性。这类数据库是为满足监控系统的公共安全需要的而构建,以 PETS 数据库为代表。近年来 PETS 专题讨论会提供了许多数据库。除此之外也有一些其它类似的数据库,比如 i-Lids 数据库^[13]、BEHAVE 数据库^[14]等。

比较典型的 PETS 行为数据库有 2004 年的 CAVIAR 数据库^[15]、2006 年的 ISCAPS 数据库^[16]和 2007 年的 REASON 数据库^[17]。CAVIAR 数据库包含 6 类行为,每类由以下一种或多种动作组成:步行、随便看、休息-晕倒-昏迷、遗留包裹、相遇-同行-分离和打架,每类行为包含 3 到 6 个视频,共有 26 个视频。视频分辨率为 384×288,帧率为 25fps。该数据库提供了背景图,视频在商场中拍摄得到,如图 4(a)所示。ISCAPS 数据库包含 7 个长视频,每个视频有 4 个视角,主要拍摄在火车站丢包的情况,参与主体有 1 个或 2 个人。所有镜头都是高清的,分辨率为 768×576,帧率为 25fps,如图 4(b)所示。REASON 数据库与 ISCAPS 数据库结构类似,分辨率和帧率相同,它是在机场大厅用 4 个摄像机拍摄得到的 8 个行为序列,主要涉及人和包的交互行为,2 个序列为“逗留”,4 个序列为“遗忘包裹”,另外 2 个序列为“丢弃包裹”,如图 4(c)所示。



(a) CAVIAR 图像 (b) ISCAPS 图像 (c) REASON 图像

图 4 PETS 典型数据库图像示例

与 PETS 行为数据库类似,i-Lids 数据库^[13]也主要涉及遗留包的问题。在地铁站台拍摄得到人、物体,以及移动的地铁和上下车的行为。包含 3 个训练视频和 6 个测试视频。分辨率为 720×576,帧率为 25fps。

BEHAVE 数据库^[14]包含两个视角下人与人之间的多种交互行为,如集合、靠近、同行、分离、忽视、跟踪、追赶等。分辨率为 640×480,帧率为 25fps。图 5 所示为经 VIPER^[18]标定后的打架和同行的图像示例。



图 5 BEHAVE 数据库中打架和同行的图像示例

除上述两种类型的数据库外,还有从电影或体育比赛视频中收集得到的数据库,比如 HOHA 数据库^[19]和足球队数据库等,在此不做详细介绍。

3 人群行为分析方法

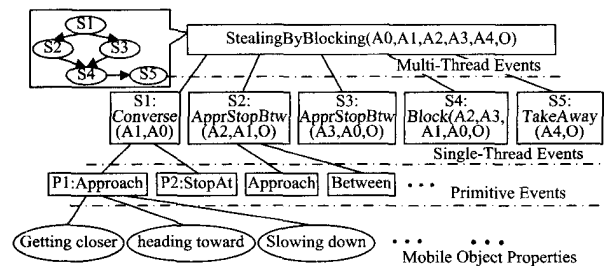
人群行为的分析和识别方法一般建立在对简单行为识别

结果的基础上,即通常采用分层算法(Hierarchical approaches)。因此,除了分析方法外,群行为的分解方法也是非常重要的。下面将从群行为的分解和识别方法两方面进行了解。

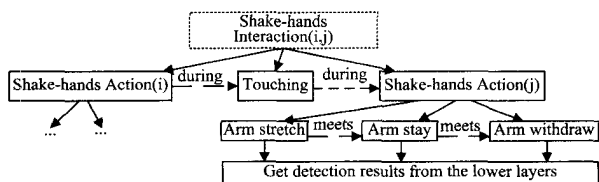
3.1 人群行为的分解方法

人群行为包含多个个体的行为,群行为的分解方法作为分析方法的重要组成部分,直接影响分析方法的表现。目前常用的群行为分解方法有3种:

(1) 将群行为看作事件,将事件抽象为原始事件、单线事件和多线事件3个层次,多线事件由一系列的逻辑上、空间上或时间上关联的单线事件组成。挖掘各类事件的时间和空间关系。该方法以文献[20,21]为代表,图6(a)即为对“Stealing By Blocking”的分解,其中靠近、停止等为原始事件,而原始事件又由一些子动作,比如接近、转头、减速等组成。单线事件描述了个体间的行为,如A1靠近A0并放下箱子进行交谈,A2靠近并停在A0和箱子之间,A3靠近并停在A0和箱子之间,A2和A3挡住视线,A4接近并偷走箱子,这些单线事件就构成了多线事件。文献[22-25]也采用了类似的分解方法,首先估计每一图像帧中人体的各个部分的状态,定义姿态(pose)为某个身体部分的状态的抽象值,姿势(gesture)为身体某部分的子动作。用状态作为参数,提取每帧的姿态。根据给定的姿态,在姿势层生成姿势序列。在更高层中,用子动作间的时间间隔及它们之间的关系来表示人群行为,图6(b)即为文献[25]中握手行为的识别过程。这种分解方法条理清楚,与人们对事件的认识过程相吻合,但过程比较繁琐,需要对子行为或动作准确辨识。



(a) “Stealing By Blocking”事件的分解^[20]



(b) 握手行为的识别过程^[25]

图6 群行为分解示意图

(2) 将人群根据行为的结构聚类为不同的子群,通过对子群行为及其关系的描述达到描述群行为的目的。Lin等^[26,27]将群行为分为对称行为和非对称行为,对称行为指群中任意两个个体都做对称运动,比如“一起步行”,而“跟踪”则属于非对称运动,因为“i跟踪j”和“j跟踪i”是不同的。作者首先将所有的人群行为聚类为非重叠的对称子群,将这些子群进行分类,对每种子群进行统一的表征,如图7中实心黑点所示,然后对不同于子群间的非对称接触通过计算动作的相关性进行检测,从而实现群行为的检测和识别。这种分解方法充分考虑到行为的特点,不需要准确辨识原始动作,可以容

忍视频中一定的模糊度,适合监控视频中群行为的分解。

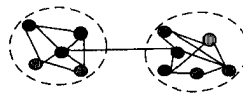


图7 文献[26,27]中的群行为分解示意图

(3) 监测并跟踪得到每个个体的路径,将人群行为分为3种类型进行描述:个体、两个体和群体,然后用“词袋”(bag-of-word)模型对行为进行表征。Ni等^[28]将人群行为的因果关系分为3类:自因果、配对因果和群因果,通过分段的行为路径来计算因果特征,每种因果关系都被表示为1个粒子滤波器,并分别构建“词袋”进行表征。Cheng等^[29]进一步发展了这种方法,仍将人群行为分为3类,采用高斯过程退化的方法对每类路径进行分段和提取特征向量,用“词袋”的方法整合众多特征,得到归一化的直方图来表示人群行为。从严格意义上来讲,该方法并不是群行为的分解方法,只是一种群行为的表示方法。该方法不考虑群行为的内部结构及个体间的关系,仅从路径入手进行分析,非常简洁,容易实现。

上述3种对人群行为的分解方法各有其适用的范围。第一种方法虽然需要对子行为或动作进行精确辨识,复杂度较高,通用性不强,但对于对识别结果的精确度要求较高的群行为的识别是有很有效的。第二种和第三种方法在降低复杂度的同时,也降低了准确度,适合在不考虑细节的前提下,对群行为进行辨识,比如靠近、分离、打架等,更适用于对海量监控视频的分析。第二种方法考虑了群行为的结构和子行为的特点,第三种方法紧紧依赖行为个体的路径,可行性都比较强。第二种方法不仅可以识别群行为,也可以进行子行为的辨识,准确度更高一些,尤其在个体路径难以准确提取的情况下。文献[30]考虑得最全面,将群行为分为群行为、群-个体交互行为、群-群交互行为、群内交互行为以及它们的组合。通过成员行为的时间间隔、子行为间的时间结构关系和成员间必要的空间约束3个方面来表征人群行为,这种方法充分考虑了人群行为的时间、空间和逻辑结构,在提高准确性的同时,复杂度也有所提升。

群行为的分解方法和识别方法密切相关,下面对群行为的识别方法进行总结。

3.2 人群行为的识别方法

现有群行为的识别方法可分为基于统计的方法(statistical approaches)和基于描述的方法(description-based approaches)两种。

基于统计的方法采用基于状态的统计模型,比如隐马尔可夫模型(Hidden Markov Model, HMM)、动态贝叶斯网(Dynamic Bayesian Networks, DBN)等具有多层结构(一般为两层)的概率图模型方法,这类方法在低层首先根据特征序列识别子动作,然后将这些子动作看作统计模型第二层的观测。文献[30,31]针对会议环境中的人群行为,分别采用双层HMM和DBN方法识别群行为。Zhang^[31]提取每个个体的视频和音频特征,训练I-HMM模型,识别个体行为,然后在此基础上结合群视频和音频特征,建立G-HMM模型,识别群行为。Dai等^[32]采用事件驱动的多层动态贝叶斯网检测事件,然后对每种行为进行定义,提出基于事件的动态上下文模型进行识别。Gupta等^[33]针对足球比赛视频提出了与或故事模型(AND-OR storyline model),用与或图中的边界表示

时空关系。Park 和 Aggarwal^[34,35]提出了分层贝叶斯网的方法来识别两个人的交互行为,身体各部分的姿势通过低层贝叶斯网来估计,整个身体的行为由高层贝叶斯网估计,姿势的变化通过动态贝叶斯网来描述。识别通过加入时空约束的高层语义描述实现。基于统计的方法非常适合行为序列的识别,只要有足够的训练数据,即使在有噪输入的情况下也可提供可靠的概率框架,但很难表征和识别有复杂时序结构的行为,比如同时发生的子行为。另外,行为越复杂,所需要的序列数据就越多,这也使得这类方法很难应用于复杂行为。

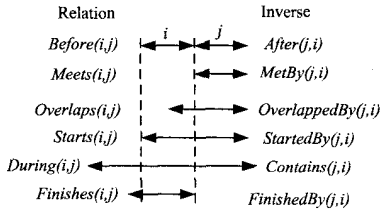


图 8 不同时间间隔之间的关系^[36]

基于描述的方法能够保持行为的空间和时间结构,它将人群运动看作满足一定关系的子运动,因此对行为的识别通过搜索满足定义的子行为来完成。基于描述的方法可以处理同时发生的子运动,通过描述子运动在时间和空间上的逻辑关系来表征复杂运动,因此人群行为分析就转化成了搜索满足一定关系的子行为。在基于描述的方法中,时间间隔一般与发生的子行为联系在一起,以说明子行为间的时间关系,空间和时间关系的描述对复杂事件是非常重要的。Allen 的时间确定法(temporal predicate)^[36]已经被广泛应用于这种方法来明确表达事件间的关系,如图 8 所示。这些关系不仅仅包含了两个时间间隔的时序关系,而且表达了同时发生的事件的复杂组合关系。Nevatia 等^[20]在将事件分为原始事件、单线事件和多线事件 3 个层次的基础上,结合 Allen 的时间确定法,通过设定数据类型、句法和参数来构建事件识别语言(Event Recognition Language, ERL),用以描述和识别各类事件。Hongeng 等^[37]也将群行为看作多线事件,采用随机有限的状态表示单线事件,这些状态可以通过对运动部分的路径和形状的贝叶斯分析的方法来识别。多线事件用逻辑和事件约束下的多个单线事件构成的事件图(event graph)表示,如图 9 所示。多线事件通过单线事件在时间逻辑的限制下产生群行为的概率得到识别。文献[28, 29, 31, 32]将群行为序列分解为身体部分、姿势和姿态,最后在语义层采用上下文无关法(context-free grammar, CFG)来描述行为的空间、时间和逻辑关系,结合低层对姿势和姿态的识别结果对群行为进行识别。与上述将逻辑推理融入到概率框架中的方法^[20, 28, 29, 31, 32, 36, 37]不同,文献[30]提出了一种随机表征和识别复杂人群行为的方法,即在表征过程中,寻找满足最大后验概率的一组成员,在识别过程中,采用马尔科夫链蒙特卡罗理论(Markov chain Monte Carlo, MCMC)进行概率分布抽样,从而达到检测活动人群的目的。基于描述的方法适合表示和识别具有复杂结构的人群行为,不仅可以处理顺序发生的行为,而且能够描述同时发生的子行为。这种行为主要的缺点在于无法弥补低层子行为的检测或识别的错误,过分依赖低层的检测结果。

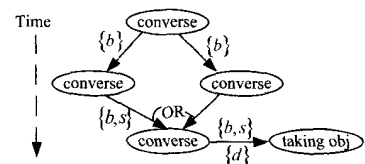


图 9 事件图示例

综上所述,基于统计的方法适合行为序列的识别,在有噪输入的情况下也可提供可靠的概率框架,准确性高,但需要较多的训练数据,且对具有复杂时序结构的行为无能为力。基于描述的方法充分考虑了子行为发生的时间间隔及其逻辑关系,注重行为的语义,使得对人群运动的分析更符合相关规则,可处理同时发生的子行为。然而,该方法以子行为的精确检测和识别为基础,仍需依赖基于统计的方法或其它方法对子行为进行检测。因此,基于统计的方法和基于描述的方法相结合能够克服两种方法自身的弱点,发挥两者的优势。基于统计的方法能够准确识别低层子行为,为高层基于描述的方法的有效性提供保证,而基于统计的方法无需处理复杂的群行为,只需对子行为进行识别,在训练数据较少的情况下,仍能在一定程度上保证准确性。

4 面临的问题及发展趋势

人群行为分析还处在初级发展阶段,由于人群行为复杂性高,变化较多,目前的研究还只能解决一些特定的简单的群行为,还不具有普适性,至今还很少见到人群行为分析在监控系统中的应用。目前人群行为分析面临的主要问题表现在以下几个方面:

(1) 诸多干扰因素的影响。人群行为的发生场景一般比较复杂,受到拍摄距离、光照、遮挡及其它因素的影响,尤其是在监控场景下,个体较小,受到干扰的影响较大。

(2) 复杂的行为结构的处理问题。虽然现有研究对某些人群行为的辨识进行了成功的尝试,但是这些行为一般比较简单,比如握手、跟随、赶超、打架等等。对于结构复杂的群行为还很少涉及,尤其对于连续发生的一系列群行为,它们的相关性需要人为干预。

(3) 以基本行为的准确辨识为基础,在很多情况下很难保证其可靠性。群行为的辨识是以个体基本行为的辨识为基础的,而在诸多干扰因素的影响下,基本行为辨识的准确度很难保证,这将直接导致群行为辨识的不可靠。

(4) 目前的人群行为辨识方法都是基于学习的方法,在很多情况下,训练样本的获取比较困难。另外,由于群行为的参与个体数目不定,表现形式多样,很难通过其他行为的学习获得对当前行为的准确辨识信息。

尽管目前人群行为分析还处于婴儿期,学者们提出的算法只能解决部分问题,即使在学术层面的研究也还不太完善,但人群行为分析的发展前景是光明的,在社会安全中有重要的应用价值。对干扰因素不敏感、对学习样本不苛刻、对基本行为的辨识结果依赖度低、能够自动对行为间的关系进行自动关联的人群行为分析方法将会成为发展趋势,并且将会在社会安全领域发挥重要作用。

参考文献

[1] Singh M, Basu A, Mandal M K. Human Activity Recognition

- Based on Silhouette Directionality [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008, 18(9): 1280-1289
- [2] Wang Y, Mori G. Human action recognition by semi-latent topic models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(10): 1762-1764
- [3] Monti F, Regazzoni C S. Human action recognition using the motion of interest points [C] // *Proceedings of International Conference on Image Processing*. 2010: 709-712
- [4] 陈昌红, 梁继民, 赵恒, 等. 步态表征和步态融合方法新进展[J]. *计算机科学*, 2010, 37(8): 15-20
- [5] Vaswani N, Chowdhury R A, Chellappa R. Activity recognition using the dynamics of the configuration of interacting objects [C] // *Proceedings of International Conference on Computer Vision and Pattern Recognition*. 2003, 2: 633-640
- [6] Khan S M, Shah M. Detecting group activities using rigidity of formation [C] // *Proceedings of ACM International Conference on Multimedia*. 2005: 403-406
- [7] Aggarwal J K, Cai Q. Human motion analysis: A review [J]. *Computer Vision and Image Understanding*, 1999, 73(3): 428-440
- [8] Kruger V, Kragic D, Ude A, et al. The meaning of action: a review on action recognition and mapping [J]. *Advanced Robotics*, 2007, 21(13): 1473-1501
- [9] Turaga P, Chellappa R, Subrahmanian V S, et al. Machine recognition of human activities: A survey [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2008, 18(11): 1473-1488
- [10] Aggarwal J K, Ryoo M S. Human activity analysis: A Review [J]. *ACM Computing Surveys*, 2011, 43(3): 1-43
- [11] CASIA gait database [DB/OL]. <http://www.cbsr.ia.ac.cn/english/Action%20Databases%20EN.asp>, 2011-10-10
- [12] Ryoo M S, Aggarwal J K. UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA) [EB/OL]. http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html, 2011-10-10
- [13] I-Lids bag and vehicle detection challenge [DB/OL]. http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html, 2011-10-10
- [14] BEHAVE Interactions Test Case Scenarios [DB/OL]. <http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/index.html>, 2011-10-10
- [15] CAVIAR test case scenarios [DB/OL]. <http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/>, 2011-10-10
- [16] ISCAPS dataset [DB/OL]. <http://www.cvg.rdg.ac.uk/PETS2006/data.html>, 2011-10-10
- [17] REASON dataset [DB/OL]. <http://www.cvg.rdg.ac.uk/PETS2007/data.html>, 2011-10-10
- [18] Mariano V Y, Min J, Park J H, et al. Performance Evaluation of Object Detection Algorithms [C] // *Proceedings of International Conference on Pattern Recognition*. 2002: 965-969
- [19] Laptev I, Marszałek M, Schmid C, et al. Learning realistic human actions from movies [C] // *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 2008: 1-8
- [20] Nevatia R, Zhao T, Hongeng S. Hierarchical language-based representation of events in video streams [C] // *Proceedings of IEEE Workshop on Event Mining*. 2003, 4: 39-46
- [21] Hongeng S, Nevatia R, Bremond F. Video based event recognition: activity representation and probabilistic recognition methods [J]. *Computer Vision and Image Understanding*, 2004, 96: 129-162
- [22] Ryoo M S, Aggarwal J K. Recognition of composite human activities through context-free grammar based representation [C] // *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. 2006: 1709-1718
- [23] Ryoo M S, Aggarwal J K. Spatio-Temporal Relationship Match: Video Structure Comparison for Recognition of Complex Human Activities [C] // *Proceedings of IEEE Conference on Computer Vision*. 2009: 1593-1600
- [24] Ryoo M S, Aggarwal J K. Recognition of high-level group activities based on activities of individual members [C] // *Proceedings of IEEE Workshop on Motion and Video Computing (WMVC)*. 2008: 1-8
- [25] Ryoo M S, Aggarwal J K. Semantic representation and recognition of continued and recursive human activities [J]. *International Journal of Computer Vision*, 2009, 32(1): 1-24
- [26] Lin W, Sun M, Poovendran R, et al. Group event detection with a varying number of group members for video surveillance [J]. *IEEE Trans. Circuits and Systems for Video Technology*, 2010, 20(8): 1057-1067
- [27] Lin W, Sun M, Poovendran R, et al. Group event detection for video surveillance [C] // *Proceedings of International Symposium on Circuits and Systems*. 2009: 2830-2833
- [28] Ni B, Yan S, Kassim A. Recognizing Human Group Activities with Localized Causalities [C] // *Proceedings of International Conference on Computer Vision Pattern Recognition*. 2009: 1470-1477
- [29] Cheng Z, Qin L, Huang Q, et al. Group Activity Recognition by Gaussian Processes Estimation [C] // *Proceedings of International Conference on Pattern Recognition*. 2010: 3228-3331
- [30] Ryoo M S, Aggarwal J K. Stochastic representation and recognition of high-level group activities [J]. *International Journal of Computer Vision*, 2011, 93(2): 183-200
- [31] Zhang D, Gatica-Perez D, Bengio S, et al. Modeling individual and group actions in meetings with layered HMMs [J]. *IEEE Transactions on Multimedia*, 2006, 8(3): 509-520
- [32] Dai P, Di H, Dong L, et al. Group interaction analysis in dynamic context [J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2008, 38(1): 275-282
- [33] Gupta A, Srinivasan P, Shi J, et al. Understanding videos, constructing plots learning a visually grounded storyline model from annotated videos [C] // *Proceedings of International Conference on Computer Vision and Pattern Recognition*. 2009: 2012-2019
- [34] Park S, Aggarwal J K. Semantic-level understanding of human actions and interactions using event hierarchy [C] // *Proceedings of International Conference on Computer Vision and Pattern Recognition Workshop*. 2004: 12-20
- [35] Park S, Aggarwal J K. A hierarchical Bayesian network for event recognition of human actions and interactions [J]. *Multimedia Systems*, 2004, 10(2): 164-179
- [36] Allen J F, Ferguson G. Actions and events in interval temporal logic [J]. *Journal of logic and computation*, 1994, 4(5): 531-579
- [37] Hongeng S, Nevatia R, Bremond F. Video-Based Event Recognition: Activity Representation and Probabilistic Recognition Methods [J]. *Computer Vision and Image Understanding*, 2004, 96(2): 129-162