行为分析算法综述

谷军霞 丁晓青 王生进
(清华大学电子工程系智能技术与系统国家重点实验室,北京 100084)

摘 要 行为分析有着广泛的应用背景,如智能监控、人机交互、运动员辅助训练、视频编码等等。近年来,在这些应用的驱动之下,行为分析已经成为图像分析、心理学、神经生理学等相关领域的研究热点。本文概述了图像领域行为分析相关研究的发展历史、研究现状及目前存在的主要问题。行为分析的相关研究起始于 20世纪的 70年代,80年代有了初步的进展,90年代是行为分析的逐步发展阶段,在这个时期提出了一些影响较大的研究方法。2000年之后,由于智能监控等方面的迫切需求,行为分析的描述方法和识别算法以及行为理解都取得了快速而深入的发展。行为分析最基本的两个问题是行为的描述和识别,行为的描述方法可分为两类:一类是基于低层图像信息的方法,一类是基于高层人体结构的方法。行为的识别算法也可分为两类:一类是基于模板匹配的算法,一类是基于状态空间的算法。本文基于行为描述和行为识别这两个基本问题,综述了目前行为分析主要研究算法,并比较了各类算法的优缺点。本文在研究了各类算法的发展历史和现状的基础上,总结了行为分析目前存在的主要问题及可能的发展方向。

关键词 行为分析 行为描述 行为识别 状态空间算法 模板匹配算法
中图法分类号: TP 391.4 **文献标识码**: A **文章编号**: 1006-8961(2009)03-0377-11

A Survey of Activity Analysis Algorithms

GU Jun-xia, DING Xiao-qing, WANG Sheng-jin
(*State Key Laboratory of Intelligent Technology and Systems, Department of Electronic Engineering, Tsinghua University, Beijing* 100084)

Abstract Human activity analysis is receiving increasing attention from computer vision researchers. This interest is motivated by a wide spectrum of applications, such as surveillance, man-machine interfaces, video coding, and so on. It has been a hot research in image analysis, psychology, and neurophysiology. This paper gives an overview of the various tasks involved in image analysis field. We focus on three major areas: (1) development history of the activity analysis, (2) important and novel ideas, (3) open problems for future research. Research about activity analysis was originated 30 years ago. In the recent years, increasing attention has been paid to this field. The two basic problems of activity analysis are the representation and the recognition of activity. This paper reviews the existing algorithms based on the two basic problems. The representation of the activity can be classified into two classes: the methods based on low-level image information and the methods based on the high level human model. And there are two kinds of activity recognition algorithms: template matching methods and state space methods. Finally, some research challenges and future directions are discussed.

Keywords activity analysis, action representation, action recognition, state space method, template matching method

1 引 言

行为分析有着广泛的应用前景,已经成为多个领域的研究热点。本文概述了图像领域中行为分析相关研究的发展历史、研究现状及目前存在的主要问题。关于行为分析算法的综述在文献[1]~[5]等中都有涉及,但是这些文章只是将行为分析作为

基金项目:国家高技术研究发展计划(863)项目(2006AA01Z115);国家自然科学基金项目(60472002)
收稿日期: 2007-02-15;**改回日期**: 2007-12-12

第一作者简介:谷军霞(1981~),女。清华大学电子工程系直博研究生。研究领域包括模式识别,行为分析。E-mail: gujx04@mails.tsinghua.edu.cn

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

运动分析的一部分来介绍,不够详尽和系统,而目前行为分析已成为一个独立的研究课题,有必要对行为分析算法进行独立详细系统的综述。

本文从行为分析的历史发展过程出发,基于行为的描述与识别这两个基本问题,综述了目前的主要研究算法。行为描述方法可分为两类:一是基于低层图像信息的方法;二是基于高层人体结构的方法。基于低层图像信息的方法可以快速鲁棒地获取特征,但是一般只能描述简单的行为。基于高层人体结构的方法可以更精细地描述行为,但是特征获取比较困难,往往要依赖于人体姿势估计的准确性。行为识别算法也可分为两类:一是基于模板匹配的算法,二是基于状态空间的算法。基于模板匹配的算法计算量较少,但是对行为时间间隔敏感;基于状态空间的算法可以避免行为时间间隔建模的问题,但是模型训练复杂。本文在研究了各类算法的发展历史和现状的基础上,阐述了目前行为分析研究存在的主要问题及可能的研究方向。

2 行为分析的应用背景

行为分析有着广泛的应用背景,如智能监控、虚拟现实、人机交互、视频编码、运动分析等等。近几年由于公共安全的需要,智能监控方面的需求迅速增加。如 1997年美国国防高级研究项目署设立了以卡内基梅隆大学为首、麻省理工学院等高校参与的视觉监控重大项目 VSAM (visual surveillance and monitoring),主要研究用于战场及普通民用场景监控的自动视频理解技术[5]。国内近几年也在城市的重要位置安装了监控摄像头,在多起犯罪案件中,视频监控录像都提供了很重要的破案线索。但是目前公共场所装有的摄像头大都只能记录当时的场景,作为事后调查的依据,而不能做到实时自动报警。行为分析的研究正可以满足智能监控中自动实时报警的迫切需求。W4监控系统包括 4个因素:when、where、who、and what,在不同的监控需求下可以利用不同的因素或者综合利用若干个因素,有些场景下利用身份 (who)就可以监控,有些场景下利用位置 (where)就可以监控,但是在对人的身份和位置都不能限制的公共场所,就必须进行行为的分析识别,即对“what”进行研究。

3 行为分析的发展历史

行为分析的相关研究起源于 20世纪的 70年代,图 1是文献[3]～[4]总结的从 20世纪 80年代到 2006年前半年为止的一些相关的重要期刊和会议 (如 Computer Vision and Image Understanding、IEEE Transaction on Pattern Analysis and Machine Intelligence、IEEE Conference on Computer Vision and Pattern Recognition、IEEE International Conference on Computer Vision、International Conference on Pattern Recognition、European Conference on Computer Vision等)上发表的有关跟踪、人体姿势估计、行为识别方面的文章数目。由图 1可见近年来行为识别的文章数目与 20世纪相比有了显著的增长,这也从一个侧面反映了行为识别相关研究的发展过程。

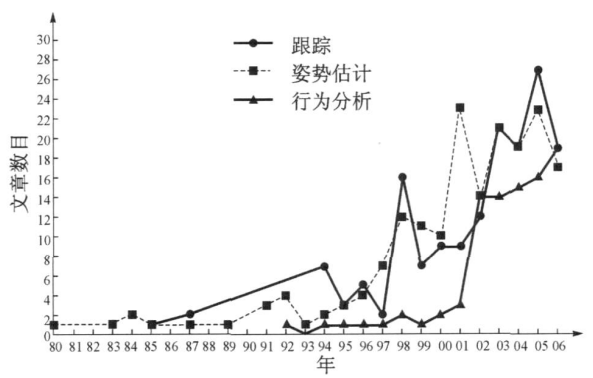


图 1 行为分析相关技术研究发展概况

Fig. 1 Development of the activity analysis

在本文中,将行为分析的发展历史分为 3个阶段来研究,第 1个是 20世纪 70到 80年代行为分析的初步研究阶段,第 2个是 20世纪 90年代行为分析的逐步发展阶段,第 3个是近年来行为分析的快速发展阶段。表 1列举了行为分析在各个发展阶段取得的重要研究成果。由此表可以看到,从 20世纪 70年代起人们就开始了行为分析的研究,到目前为止,已经从简化的运动识别发展到复杂的接近真实场景的行为分析。20世纪 70到 80年代是行为分析的初步研究阶段,在这个时期研究的重点是人对运动信息理解的原理,如 1975年 Johansson通过试验证明仅仅用人体上若干点的运动就可以描述人的行为[6],这种描述行为的方法对后来基于人体结构的行为描述算法起到了很重要的指导作用。20世

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

纪 80年代末,Nagel提出行为的四层定义[7],进一步加深了人们对行为在图像分析领域的理解。20世纪 90年代,行为分析的相关研究取得逐步发展,这个时期提出了一些影响较大的研究思路,如 Yamato将隐马尔可夫模型(HMM)引入行为分析[8],开始了基于状态空间的行为识别方法;之后 Davis等人提出了描述行为的两个模板:运动能量图(MEI)和运动历史图(MHI)[9],开启了模板匹配的行为分析算法。这个时期的研究往往限于视角固定、环境简单、行为规则的情况。2000年之后,行为分析各方面的研究都得到了快速而深入的发展,行为识别的视角问题取得了一定的研究成果[10-14]。但是目前的研究还主要集中在有限类别简单规则的行为识别或者特定场景中的异常行为检测上[15-17],很少有算法涉及到复杂行为的高层次理解。行为分析要在真实场景中应用,仍然存在很多严峻的问题。

表 1　行为分析相关研究的发展历史

Tab.1　Development history of the activity analysis

时间	作者	主　要　贡　献
		20世纪 70～80年代:行为分析初步研究阶段
1975	Johansson[6]	人体上若干关节点的运动可以描述行为。
1980	Rourke[18]	将人体 3维模型引入单视角图像的人体姿态估计,开始了人体姿势估计算法的研究。
1988	Nagel[7]	提出对行为的分层次定义,进一步加深了人们对行为的理解。该文中将行为分为 4层: change-event-verb-history。
		20世纪 90年代:行为分析逐步发展阶段
1992	Yamato[8]	首次将 HMM引入行为分析,开始了基于状态空间的行为分析研究算法。
1994	Polana[19]	将时空模板的方法引入行为分析,开始了基于模板的行为分析研究算法。
1996	Gavrila[20]	基于 3维人体模型进行姿势估计,将时变模式的模板匹配算法——动态时间校正(DTW)引入行为识别算法,以解决不同行为样本的时间间隔不同的问题。
1997	Davis[9]	提出描述行为的两个模板:MEI和 MHI。
1997	Brand[21]	将 CHMM(coupled HMM)用于双手行为识别。
1998	Grimson[22]	提出了一种异常行为检测的算法,开始了监控中异常行为检测的研究。
1999	Haritaoglu[23]	实现了一个 W4的监控系统,可以分析 carrying objects行为。
		2000年 ～ :行为分析快速发展阶段
2001	Bobick[24]	基于行为描述模板 MEI和 MHI识别行为。
2002	Ren[25]	用 2维、3维混合人体模型描述行为,提出 PBCHMM(primitive-based CHMM)的双手行为识别算法。
2002	Ben-Arie[26]	提出基于行为索引进行识别的算法,避免了不同行为样本时间间隔不同的问题,主要适用于周期行为识别。
2003	Luo[27]	将动态贝叶斯网络(DBN)引入行为识别的算法。
2003	Shimozaki[28]	利用自组织神经网络 ECSO(energy competition self-organization network)和 R-ECSO(recurrent ECSO)分别识别行为的空间信息和时域信息。
2003	Masoud[29]	提出了用回归滤波描述行为的方法,并将整个行为过程表示为一个流形。
2004	Gritai[30]	用人体姿势描述行为,并将人体测量学引入行为识别的视角不变性研究。
2005	Li[31]	提出 PCRM(pixel change ratio map)的行为描述方法。
2005	Yilmaz[32]	提出 STV(spatiotemporal volume)的行为描述方法。
2005	Fanti[33]	利用人体混合模型描述行为。
2005	Robertson[34-35]	用 HMM识别行为,并涉及到行为理解。
2006	Parameswaran[11]	提出 2维视角不变空间,以解决行为识别的视角问题。
2006	Weinland[12]	提出了 MHV(motion history volumes)的行为描述方法,相当于 3维空间的 MHI。
2006	Yilmaz[14]	利用对极几何性质解决行为识别中的视角变化问题。
2006	Ryoo[36]	用人体姿势描述行为,并提出用 CFG(context-free grammar)方法理解行为。
2006	Wang[37]	用平均运动形状模板和平均运动能量模板描述行为。
2007	Li X[38]	提出用光流方向直方图描述行为的方法。
2007	Guerra-Filho[39]	提出了三层次的行为描述语言:kinetology, morphology, syntax,为行为理解提供工具。
2007	Abhinav[40]	将目标识别嵌入行为分析中,实现人与物的交互行为分析。

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

4 行为分析的研究现状

行为分析最基本的两个问题就是行为的描述和识别。这两个问题虽然从行为分析起始就存在,但是到目前为止仍然还没有一个成熟的解答。本文对于每种算法都从这两个方面进行讨论。

在以往的文献当中,表示行为的词很多,如: action, activity, behavior, event, movement等等。在不同的应用背景下,行为会被赋予不同的含义,甚至同一个词表示的含义也不尽相同。1988年, Nagel提出了对事件或行为的分层次定义[7],该文将行为分为四层定义: change-event-verb-history, 其中‘change 表示变化,即图像中不同于噪声的变化,‘event 指预先定义的动作,是描述更复杂行为的基元,‘verb 表示一种行为,‘history 是对整个行为的理解,在这个层次上不仅需要知道人的行为,还必须了解人所处的环境。2006年,Moeslund也采用了分层定义法[4]来定义行为,他将行为分为三个层次: motor primitives, action 和 activity。motor primitives 是指描述行为的基元实体; action是 motor primitives 的有序组合; activity是由许多 action经过逻辑组合而形成的。图 2是行为三层定义的示意图和打网球的例子。本文采用了 Moeslund的三层定义法。

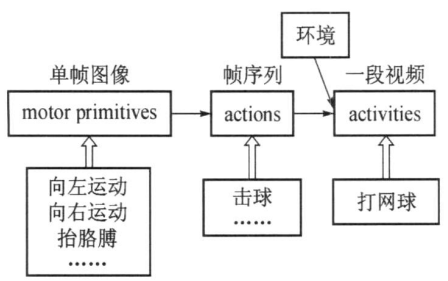


图 2 行为的分层次定义法

Fig. 2 Hierarchy definition of activity analysis

根据三层定义法,目前大部分行为分析的研究还处于 action阶段,即 action recognition。在 action 阶段,行为描述和识别都可以分为两类算法。表 2是一些典型算法按照行为描述识别的不同方法分类的结果。由此表可见,在行为识别初期(20世纪 90年代)所利用的特征大都是基于低层图像信息的,近年来随着姿势估计算法的发展,基于高层人体结构特征的行为描述算法在日益增多。

表 2 行为分析相关算法分类

Tab. 2 Algorithms of the activity analysis

行为描述	行为识别	
	模板匹配	状态空间
高层人体结构	[1996, Gavrila][20] [1999, Haritaoglu][23] [2002, Ben-Arie][26] [2004, Gritai][30] [2005, Fanti][33] [2006, Yilmaz][14]	[1997, Brand][21] [2002, Ren][25] [2006, Ryoo][36] [2006, 李][41]
低层图像信息	[1994, Polana][19] [1997, Davis][9] [1998, Grimson][22] [2000, 胡][42] [2001, Bobick][24] [2003, Masoud][29] [2003, Alexei][43] [2005, Yilmaz][44] [2006, Weinland][12] [2006, Wang][37] [2007, 冯][45]	[1992, Yamato][8] [2003, Luo][27] [2005, Robertson][34] [2007 Li][38]

在下文中将以行为分析的这两个基本问题为分类标准,依次介绍各种算法,并比较每类算法的优缺点。

4.1 行为描述方法

4.1.1 基于低层图像信息的行为描述方法

低层图像信息特征获取简单,这种描述行为的算法一直以来都是行为描述的一个重要方向。在行为描述中可利用的低层图像信息包括:前景目标、前景目标的运动速度、光流、运动轨迹信息、前景目标的轮廓等等。

1992年, Yamato利用运动网格特征序列描述行为[8],指出人的高层结构信息可以很精细地描述人的行为,但是在复杂背景下人体姿势很难精确提取,所以他采用了易提取的、鲁棒的前景目标来描述行为。该文中采用的运动网格特征如图 3所示[8],每个网格中黑像素个数的比例构成这帧图像的运动网格特征矢量,一个行为就用若干连续帧图像的运动网格特征矢量表示。这种描述行为的方法非常简单,但是对于视角变化和行为主体变化很敏感。

1994年, Polana在文章“How to get your man without finding his body parts?”[19]中对运动网格特征进行了改进,将运动幅度引入了行为描述之中。

之后随着光流法的发展,这种表示了运动速度

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

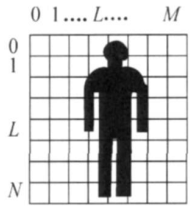


图 3　单帧图像的运动网格特征矢量

Fig. 3　Mesh feature of one frame

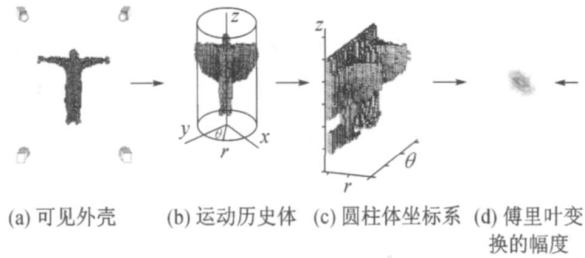
幅度和方向的方法成为描述行为中一种最常用的算法。众所周知,光流的噪声比较大,所以在光流应用到行为描述时进行了各种改进。2003年,Efros等人[43]针对远距离、分辨率低的行为主体提出用校正和模糊的方法提高光流法对噪声的鲁棒性。2007年,光流方向直方图[38,45]被应用于行为的描述,用这种统计的方法增强了光流特征的稳定性。另外光流特征也常和其他特征一起使用,如 Ahmad等人[46]将光流特征和运动区域的表面特征融合使用来描述行为。

1997年 Davis和 Bobick提出了一种新颖的描述行为的时空模板方法[9]:MEI和 MHI。MEI是一幅二值图像,它表示了在一个图像行为序列中发生运动的所有区域。MHI相当于一幅灰度图像,它描述了运动序列的时序信息,距离当前时刻越远的运动值越小。其中行为序列"跳舞"的 MEI和 MHI如图 4。这种行为描述方法用两幅模板图像表示了整个行为的过程。

图 4　行为"跳舞"的 MEI和 MHI描述

Fig. 4　MEI and MHI of 'dancing' sequence

2006年,Weinland提出用 MHV (motion history volumes)模板描述行为[12],这相当于 3维空间的 MHI模板。为了解决行为的视角变化问题,Weinland提出的算法并没有直接利用 MHV作为模板特征,而是提取了 Fourier变换特征。如图 5所示,首先由多个摄像头获取的多视角原始图像重建 3维目标,之后将 3维目标投影到圆柱坐标系,在圆柱坐标系内进行 Fourier变换以获得描述行为的 Fourier特征。



(a) 可见外壳　(b) 运动历史体　(c) 圆柱体坐标系　(d) 傅里叶变换的幅度

图 5　Fourier变换特征的获取

Fig. 5　Extraction of the Fourier transformation feature

在 Weinland的算法中,多视角的 3维重建技术被引入了行为特征的获取过程。这种方法有助于解决行为分析中的视角难题,这也是近几年行为分析的一个发展趋势。

前景目标的轮廓描述了行为者的形状,也是描述行为的一种有效方法。2005年,Yilmaz提出用 STV (spatio-temporal volume)描述行为[32],STV是指行为主体的轮廓随时间变化的过程,它既利用了行为的轮廓信息又包含了时间信息,如图 6所示。

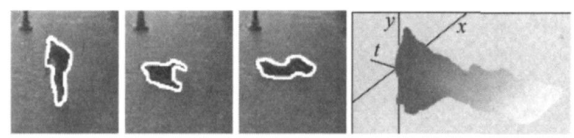


图 6　行为"躺下"的 STV描述

Fig. 6　STV of the 'lying' sequence

之后,Wang等人[37]提出用基于轮廓的平均运动形状 (MMS)和基于运动前景的平均运动能量 (AME)两个模板描述行为,如图 7所示。MMS用一个轮廓模板描述了整个行为序列过程中行为主体轮廓的变化。

(a) 跑步序列MMS特征(最后一个轮廓曲线为序列的MMS特征)

(b) 一个序列AME特征(最后一个图为序列的AME特征)

图 7　行为描述 MMS和 AME模板

Fig. 7　MMS and AME template of activity description

在一些特殊的应用场景中,轨迹也被作为行为描述的特征。例如在文献 [42]中,双手和脸的运动

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

轨迹特征被用来识别太极拳的套路。Rao等人[47]也将运动轨迹曲线用于手部的行为识别。Grimson用速度和轨迹描述行为进行异常行为检测[22];运动速度、轨迹和目标表面等特征也被综合利用进行特定行为的理解[34]。

自 20世纪 90年代开始,基于图像低层信息的行为描述方法取得了不断地发展,由最初简单的运动网格特征发展到描述序列行为的模板,由单视角 2维行为描述方法扩展到了多视角 3维行为描述方法。由上文中可以看到基于图像低层信息的特征获取简单,但它的应用一般限于有限类别的、规则的行为描述。

4.1.2 基于高层人体结构的行为描述方法

人的高层结构信息是指人身体结构所呈现的姿势,与低层图像信息相比,它可以更精细地描述人的行为。

根据提取特征过程中利用的人体模型不同,可以将这类描述行为的算法分为三种:基于人体点模型的方法、基于 2维人体模型的方法和基于 3维人体模型的方法。

人体点模型源于 1975年 Johansson的一个实验[6]。图 8是该文中提出的 12点人体模型。他的试验称为 LPD(light pints display),其过程为:首先在表演者身体上 12个关节装上小灯,然后表演者在黑屋子里运动,并用摄像机记录灯的运动轨迹,那么人仅仅依靠这些灯的运动就可以推断出表演者的行为。Johansson通过这个试验证明了仅仅用人体上若干点的运动就可以描述人的行为。

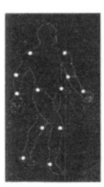


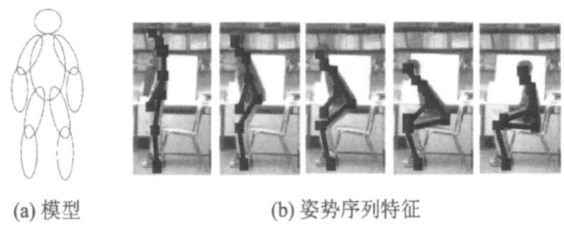
图 8 12点人体模型

Fig. 8 Human point model

这种描述行为的点模型方法对后来基于人体结构的行为描述算法起到了很重要的指导作用。但是由于需要自动估计关节点的位置,所以要受限于姿势估计算法的发展。到了近几年随着姿势估计算法的快速发展,人体点模型才被应用于行为的描述。

2004年,Gritai利用 13个点的人体模型描述人体姿势[30],其中每个点的坐标就是 1维特征。之后,Yilmaz提出了运动摄像头下的行为识别问题[14],该文也采用了 13点人体模型描述行为,一个行为就表示成人体上这些关键点的运动轨迹。在点模型中,有些算法也考虑了关节点之间的拓扑关系,由此构成图模型来描述人体结构[48]。基于人体点模型的方法有助于解决行为视角变化问题,不失为一种有效的行为描述方法。但是目前人体 3维关节点位置的自动估计问题还没有完全解决。

在单摄像头系统中,常用 2维人体模型描述行为。如 Ben-Arie使用了 2维椭圆人体模型,如图 9(a)所示[26]。人的行为则用 2维姿势序列表示,如图 9(b)所示的"坐"这个行为的姿势序列。这种行为描述算法会遇到遮挡和视角变化的难题,一般只能处理限制视角的行为描述。



(a) 模型 (b) 姿势序列特征

图 9 人体模型与姿势序列

Fig. 9 Homan model and pose sequence

早在 20世纪 80～90年代,研究者就提出了基于 3维人体模型的姿势估计算法。例如 1980年 Rourke[18]和 1996年 Gavrila[20]提出的基于 3维人体模型行为描述方法。这两种算法都是根据摄像机的投影矩阵将 3维模型投影到 2维图像平面,然后寻找与 2维图像最相似的姿势。Gavrila提出的 3维人体模型如图 10所示。该算法假设 3维人体模型具有 N维自由度,且 3维模型到 2维平面的投影矩阵已知,姿势估计问题就变成在 3维人体模型姿势空间中寻找与 2维平面投影图像最相似的姿势。对于种类繁多的行为来说,这种算法的搜索空间是巨大的,所以这类算法往往只适用于有限行为种类的情况。

图 10 人体模型(saying'Hi!')

Fig. 10 Human model(saying'Hi!')

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

近几年还提出了直接在目标的 3维重建空间进行姿势估计的算法,这类算法先由多视角图像进行目标的 3维重建,然后直接在 3维空间进行姿势估计。Kehl等人在 2005年的人体跟踪算法中使用了这种姿势估计的方法[49],其结果如图 11所示。

(a) 3维人体模型　(b) 3维目标重建　(c) 姿势估计结果

图 11　3维空间姿势估计

Fig. 11　Pose estimation in 3D

这种直接在 3维空间估计姿势的算法可以在一定程度上解决视角和遮挡问题,但是这种算法是以目标的 3维重建为基础的,所以 3维重建的性能将会影响到姿势估计的准确性。

由上述各种基于高层人体结构的行为描述方法可以看到,这类算法依赖于姿势估计的准确性。近年来,随着姿势估计算法的成熟,基于高层人体结构的行为描述算法必将日益发展。

基于低层图像信息和基于高层人体结构的行为描述方法各有优缺点。从特征的获取难度来看,基于低层图像信息的行为描述算法获取特征方式简单,一般可以由前景目标区域和目标的运动信息直接获得;基于高层人体结构特征的方法获取特征要依赖于人体姿势估计的准确性,而人体姿势估计本身就是一个还没有完全解决的问题。从对行为描述的精度来看,基于低层图像信息的方法是用一种粗糙的特征描述整个行为,而后者可以对行为的过程进行精细的有物理意义的描述,更有利于进行高层次的行为分析。

行为描述方法从 20世纪 70年代发展到现在,取得了很大的进步,目前面临的主要难点就是视角和遮挡问题,3维行为描述是解决这个问题的一个有效途径。

4.2　行为识别算法

行为是一个时变信号,行为的识别问题就是一个时变信号的分类问题。用于解决这类问题的常用方法有两种,一是模板匹配法,一是状态空间法。

4.2.1　基于模板匹配的行为识别算法

在行为识别中,基于模板匹配的算法可以分为帧对帧匹配方法和融合匹配方法。

帧对帧匹配方法是指直接用测试集的行为特征序列与参考集的行为特征序列逐帧匹配。这种算法常用动态时间规划 (DTW)来解决时间配准问题。DTW算法是基于动态规划思想的模板匹配算法,用于计算两个长度不同的模板之间的相似程度。DTW通常假设两个模式的端点已经准确地对齐从而把匹配问题转化为在有限网格上寻找从起点到终点的最优路径的问题。1978年,DTW由 Sakoe和 Chiba首次在语音识别中引入[50],而后成为语音识别的一种广泛应用的算法。1996年 Gavrila将 DTW引入行为识别[20],用于解决不同行为样本时间间隔不同的问题。2004年,Gritai提出行为识别算法中也采用了 DTW方法[30]。自组织神经网络也被用于逐帧匹配算法中。在文献[51]中,首先将视频切分为 15帧一段的子视频,然后用一个三层的神经网络分类器分析识别各个子视频。在文献[28]中提出分别用 ECSO和 R-ECSO两个神经网络识别行为的空间信息和时域信息。

融合匹配方法是指先将整个行为过程融合为一个整体模板或者若干个固定数目的模板,然后再利用这有限个模板进行匹配。Davis[9]将整个行为过程融合为两个 2维模板:MEI和 MHI,然后用马氏距离分类器进行行为识别;Wang[37]也将整个行为过程融合为两个 2维模板:MMS和 AME,然后用最近邻分类器识别行为;文献[52]用一个模板描述了整个行为,并用神经网络来识别行为模板;Masoud[29]将一个行为序列描述为一个流形进行识别;Yilmaz[14]将整个行为融合为一个 3维的 STV点集,然后利用对极几何原理[53]计算 3维点集之间匹配程度;Weinland[12]提出用 3维空间的 MHV模型描述一个行为,在 MHV模型上提取特征矢量之后,用 fisher分类器对行为进行分类。另外对于周期行为识别,Polana[19]将每个行为划分为固定的 T段识别,Ben-Arie[26]用关键帧表示每个行为,然后用模板匹配法识别这些固定帧数的行为。

4.2.2　基于状态空间的行为识别算法

HMM在基于状态空间的行为识别算法中应用最为广泛。1992年,Yamato首次将 HMM引入行为识别[8],2005年,Robertson在他的文章中也采用了 HMM识别行为[34],2006年,Ryoo将 HMM用姿势识别[36]。同时 HMM的各种改进方法也被用于行为识别,如 Brand将 CHMM (coupled hidden Markov model)用于双手行为的识别[21],之后 Ren将 CHMM

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

改进为 PCHMM (primitive-based CHMM)用于双手行为识别[25]。HMM是一种有效的时变信号的处理方法,它隐含了对时间的校正,并提供了学习机制和识别能力。但是如果对于每一类行为都建立一个模型,那么所需要的训练样本将是巨大的。

2003年,Luo将 DBN(动态贝叶斯网络)引入行为识别[27],DBN是沿时间轴展开的贝叶斯网络。该文还对 HMM和 DBN进行了比较:在一个时间切片上,HMM只能含有一个隐含节点和一个观测节点;而 DBN在一个时间切片上是一个贝叶斯网络,可以包含多个有因果关系的节点; HMM在一个时刻需要将所有的特征压缩到一个节点中,那么所需要的训练样本将是巨大的(相当于联合概率密度函数);而 DBN用多个节点描述,即用条件概率来形成联合概率,训练相对要简单;但是 DBN的设计要比 HMM复杂得多。文献[41]中采用了贝叶斯网络来识别 2维行为,该算法在贝叶斯网络分类器之前加入了一个视角分类器,将行为视角分为三类:正视角,斜视角,侧视角,在一定程度上解决 2维视角问题。

模板匹配的行为识别算法计算相对简单,但是对行为的时间间隔敏感,鲁棒性差;基于状态空间的行为识别算法可以避免对时间间隔建模的问题,但是需要的训练样本大,计算复杂。

4.3 行为理解算法

上述所介绍的行为描述和识别算法系统都是在 action recognition层次的一些研究结果,除此之外,近几年对行为语义层次上的理解也取得了一定的研究进展。如 2005年,Robertson提出在特定场景中理解行为[34-35],其中涉及到了行为定义的第 3个层次:activity analysis。该文将行为分为两层定义:简单行为识别和行为理解。例如:走(简单行为)—在人行横道上走路(时空行为)—穿越马路(行为的理解)。该算法首先由低层图像特征,如位置、运动速度、运动区域等来进行简单行为识别,然后利用 HMM进行特定场景下的行为理解。2006年,Ryoo提出了基于 CFG(context-free grammer)语言规则的行为理解算法[36],在逻辑层次上分析了两个人之间的 8种行为。该文将行为分析过程分为四部分依次进行:首先根据人体结构进行直立人的身体部件检测、之后利用贝叶斯网络进行人体姿态估计、再利用 HMM进行人体姿势估计、最后基于 CFG进行行为的逻辑推理。

行为语言是进行行为理解的一个有效工具。2007年 Guerra-Filho等人提出了三个层次的行为语言:kinetology, morphology 和 syntax[39]。Kinetology相当于语言中的字母,表示最基本的运动单元并将其符号化;morphology为词,表示了 kinetology的组合规则;多个 morphology就构成了句子 syntax。借助于行为语言可以描述行为,理解行为并重现行为。

行为理解是进行行为分析的最终目的,虽然目前行为理解算法还处于初步发展阶段,但是行为识别和行为语言的快速发展必然会推动行为理解的进一步发展。

5 行为分析存在的问题及发展方向

行为分析自 20世纪 70～80年代发展至今,在各个层次上都取得了很大的研究进展,但是行为分析算法并不成熟,存在很多严峻的问题有待解决。

(1)行为描述:视角、遮挡问题

在不同的应用场景下,学者们已经提出了很多行为描述的方法。但是行为极其复杂多变,低层图像信息显然难以精确描述行为,而高层人体结构特征要依赖于姿势估计的精度。如何更有效地描述行为仍然是制约行为分析发展的一个重要问题。

无论哪类描述方法,目前面临的主要难点都是视角和遮挡问题。这个问题的解决一般采用多视角技术或者 3维行为描述方法。对于多视角系统,多摄像机之间的选择和信息融合是一个关键的问题[5]。3维行为描述是指先由多视角图像重建 3维目标,然后在 3维空间中提取行为特征。3维重建对摄像机标定有较高要求,所需要运算量和存储量也较大。

(2)行为分析:高层次理解问题

目前对行为分析的研究往往只进行到第 2个层次,即 action recognition。处理的行为可以分为两类,一类是有限类别简单的规则行为,如:坐、走、跑、站立、蹲下等。另一类是在具体场景中处理特定的行为[23, 54-57],如丢包行为、取物行为等等,在这种场景下行为有严格的限制,行为描述一般采用运动速度或者轨迹。这两种背景下的行为分析距离实际应用都还有很大差距。行为分析的最终目的是要理解人的行为,这种理解不仅要识别人的基本行为,还要结合行为者所处的环境进行分析。

另外行为分析还存在前景目标自动检测、行为的自动切分、实时报警、公开数据库建设[58]及性能评测标准建立等等问题。

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

6 结 论

自 20世纪 70年代至今,行为分析已由原来对人体运动信息的简单分析发展到对规则行为和特殊行为的识别,进一步发展为行为的高层次理解。在行为分析中,最基本的两个问题就是行为的描述和识别。在 20多年间,对于行为的描述问题,提出了基于低层图像信息和高层人体结构的各种方法,但目前这些方法都面临视角和遮挡问题,而 3维行为描述成为解决这个问题的有效途径之一。对于行为识别问题,引入了时序模式识别的经典处理方法,如 DTW模板匹配算法、HMM状态空间法等等。但是复杂多变的行为类内方差往往很大(行为者、视角、环境等可变因素造成极大的类内差异),同时还带有模糊性,如何将时变模式识别方法有效地应用到行为识别中仍然是一个很重要的问题。

本文从行为分析的发展历史出发,基于行为分析的两个基本问题:行为描述问题和行为识别问题,综述了行为分析的各种算法,并在此基础上阐述了行为分析目前仍然存在的问题和发展方向。希望能对相关领域的研究人员有所裨益。

参考文献 (References)

1 Aggarwal J K, Cai Q. Human motion analysis: A review [J]. Computer Vision and Image Understanding, 1999, **73**(3): 428-440.

2 Gavrila D M. The visual analysis of human movement: A survey [J]. Computer Vision and Image Understanding, 1999, **73**(1): 82-98.

3 Moeslund Thomas B, Granum Erik. A survey of computer vision-based human motion capture [J]. Computer Vision and Image Understanding, 2001, **81**(3): 231-286.

4 Moeslund Thomas B, Hilton Adrian, Krüger Volker. A survey of advances in vision-based human motion capture and analysis [J]. Computer Vision and Image Understanding, 2006, **104**(3): 90-126.

5 Wang Liang, Hu Wei-ming, Tan Tie-niu. A survey of visual analysis of human motion [J]. Chinese Journal of Computers, 2002, **25**(3): 225-237. [王亮,胡卫明,谭铁牛. 人运动的视觉分析综述 [J]. 计算机学报, 2002, **25**(3): 225-237.]

6 Johansson G. Visual motion perception [J]. Scientific American, 1975, **232**(2): 76-88.

7 Nagel H H. From image sequences towards conceptual descriptions [J]. Image and Vision Computing, 1988, **6**(2): 59-74.

8 Yamato Junji, Ohya Jun, Ishii Kenichiro. Recognition human action in time-sequential images using hidden Markov model [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], Champaign, Urbana, USA, 1992: 379-385.

9 Davis James W, Bobick A F. The representation and recognition of human movement using temporal templates [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Juan, Puerto Rico, 1997: 928-934.

10 Parameswaran V, Chellappa R. View invariants for human action recognition [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], Madison, Wisconsin, USA, 2003: 613-619.

11 Parameswaran V, Chellappa R. View invariance for human action recognition [J]. International Journal of Computer Vision, 2006, **66**(1): 83-101.

12 Weinland Daniel, Ronfard Remi, Edmond Boyer. Free viewpoint action recognition using motion history volumes [J]. Computer Vision and Image Understanding, 2006, **104**(2-3): 249-257.

13 Rao C, Gritai A, Shah M, *et al*. View invariant alignment and matching of video sequences [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Nice, France, 2003: 939-945.

14 Yilmaz Alper, Shah Mubarak. Matching actions in presence of camera motion [J]. Computer Vision and Image Understanding, 2006, **104**(2-3): 221-231.

15 Zhong Hua, Shi Jian-bo, Visontai M. Detecting unusual activity in video [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], Washington DC, USA, 2004: 819-826.

16 Dong Zhang, Gatica-Perez D, Bengio S, *et al*. Semi-supervised adapted HMMs for unusual event detection [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, California, USA, 2005: 611-618.

17 Duong T, Bui H, Phung D, *et al*. Activity recognition and abnormality detection with the switching hidden semi-Markov Model [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, California, USA, 2005: 838-845.

18 Rourke J O, Badler N. Model-based image analysis of human motion using constraint propagation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1980, **2**(6): 522-536.

19 Polana R, Nelson R. Low-level recognition of human motion [A]. In: Proceedings of Motion of Non-Rigid and Articulated Objects [C], Austin, Texas, USA, 1994: 77-82.

20 Gavrila Dariu M. Vision-Based 3-D Tracking of Human in Action [D]. Maryland, USA: University of Maryland, 1996.

21 Brand M, Oliver N, Pentland A. Coupled hidden Markov models for complex action recognition [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Juan, Puerto Rico, 1997: 994-999.

22 Grimson W E L, Stauffer C, Romano R, *et al*. Using adaptive tacking to classify and monitor activities in a site [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], Santa Barbara, California, USA, 1998: 22-29.

23 Haritaoglu Ismail. W4: A Real-time System for Detection and

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Tracking of People and Motion Their Activities [D]. Maryland, USA: University of Maryland, 1999.

24 Bobick Aaron F, Davis James W. The recognition of human movement using temporal templates [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, **23**(3): 257-267.

25 Ren Hai-bing, Xu Guang-you. Human action recognition with primitive-based coupled-HMM [A]. In: Proceedings of International Conference on Pattern Recognition [C], Quebec City, Canada, 2002: 494-498.

26 Ben-Arie Jezekiel, Wang Zhi-qian, Pandit Puvin, *et al*. Human activity recognition using multidimensional indexing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, **24**(8): 1091-1104.

27 Luo Ying, Wu Tzong-der, Hwang Jenq-neng. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks [J]. Computer Vision and Image Understanding, 2003, **92**(2-3): 196-216.

28 Shimozaki M, Kuniyoshi Y. Integration of spatial and temporal contexts for action recognition by self organizing neural networks [A]. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems [C], Las Vegas, Nevada, USA, 2003: 2385-2391.

29 Masoud Osama, Papanikolopoulos Nikos. A method for human action recognition [J]. Image and Vision Computing, 2003, **21**(8): 729-743.

30 Gritai Alexei, Sheikh Yaser, Shah Mubarak. On the use of anthropometry in the invariant analysis of human actions [A]. In: Proceedings of International Conference on Pattern Recognition [C], Cambridge, UK, 2004: 923-926.

31 Li Hao-ran, Rajan Deepu, Chia Liang-tien. A new method histogram to index motion content in video segments [J]. Pattern Recognition Letters, 2005, **26**(9): 1221-1231.

32 Yilmaz Alper, Shah Mubarak. Actions sketch: A novel action representation [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, California, USA, 2005: 984-989.

33 Fanti Claudio, Zelnik-Manor Lihi, Perona Pietro. Hybrid models for human motion recognition [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, California, USA, 2005: 1166-1173.

34 Robertson N, Reid I. Behavior understanding in videos: a combined method. [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Beijing, 2005: 808-815.

35 Robertson N, Reid I. A general method for human activity recognition in video [J]. Computer Proceedings of Vision and Image Understanding, 2006, **104**(2-3): 232-248.

36 Ryoo M S, Aggarwal J K. Recognition of composite human activities through context-free grammar based representation [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], New York, USA, 2006: 1709-1718.

37 Wang Liang, Suter David. Informative shape representations for human action recognition [A]. In: Proceedings of International Conference on Pattern Recognition [C], Hong Kong, 2006: 1266-1269.

38 Li X. HMM based action recognition using oriented histograms of optical flow field [J]. Electronics Letters, 2007, **43**(10): 560-561.

39 Guerra-Filho G, Aloimonos Y. A language for human action [J]. Computer, 2007, **40**(5): 42-51.

40 Abhinav Gupta, Larry Davis. Objects in action: An approach for combining action understanding and object perception [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern [C], Minneapolis, Minnesota, USA, 2007: 1-8.

41 Li Yan-ting, Luo Yu-pin, Tang Guang-rong. Activity recognition method of multiple view angles from monocular videos [J]. Journal of Computer Applications, 2006, **26**(7): 1592-1594. [**李妍婷,罗予频,唐光荣. 单目视频中的多视角行为识别方法**[J]. **计算机应用**, 2006, **26**(7): 1592-1594].

42 Hu Chang-bo, Feng Tao, Ma Songde, *et al*. PCA based human activity recognition [J]. Journal of Image and Graphics, 2000, **5**(10): 24-30. [**胡长勃, 冯涛, 马颂德等. 基于主元分析法的行为识别**[J]. **中国图象图形学报**, 2000, **5**(10): 24-30].

43 Alexei A Efros, Alexander C Berg, Greg Mori, *et al*. Recognizing action at a distance [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Nice, France, 2003: 726-733.

44 Yilmaz Alper, Shah Mubarak. Recognizing human actions in videos acquired by un-calibrated moving cameras [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Beijing, 2005: 150-157.

45 Feng Bo, Zhao Chun-hui, Yang Tao, *et al*. Real-time human action recognition based on optical-flow feature and sequence alignment [J]. Application Research of Computers, 2007, **24**(3): 194-196. [**冯波,赵春晖,杨涛等. 基于光流特征与序列比对的实时行为识别**[J]. **计算机应用研究**, 2007, **24**(3): 194-196].

46 Ahmad Mohiuddin, Lee Seong-Whan. Human action recognition using multi-view image sequences features [A]. In: Proceedings of International Conference on Automatic Face and Gesture Recognition [C], Southampton, UK, 2006: 523-528.

47 Rao Cen, Yilmaz Alper, Shah Mubarak. View-invariant representation and recognition of actions [J]. International Journal of Computer Vision, 2002, **50**(2): 203-226.

48 Taycher Leonid, Fisher Iii John W, Darrell Trevor. Recovering articulated model topology from observed rigid motion [A]. In: Proceedings of Neural Information Processing Systems [C], Vancouver, Canada, 2002: 1311-1318.

49 Kehl R, Bray M, VanGool L. Full body tracking from multiple views using stochastic sampling [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], San Diego, California, USA, 2005: 129-136.

50 Sakoe H, Chiba S. Dynamic programming algorithm optimization for spoken word recognition [J]. IEEE Transactions on Acoustics, Speech and Signal Processing, 1978, **26**(1): 43-49.

51 Sacchi, C, Regazzoni C, Gera G, *et al*. Use of neural networks for

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

behaviour understanding in railway transport monitoring applications [A]. In: Proceedings of International Conference on Image Processing [C], Thessaloniki, Greece, 2001: 541-544.

52 Sharma A, Kumar D K, Kumar S, *et al*. Recognition of human actions using moment based features and artificial neural networks [A]. In: Proceedings of International Conference on Multimedia Modelling [C], Brisbane, Australia, 2004: 368.

53 Hartley Richard, Ziaaerman Andrew. Multiple View Geometry in Computer Vision [M]. Cambridge, MA, USA, Cambridge University Press, 2004: 239-259.

54 Tao Dacheng, Li Xuelong, Maybank, *et al*. Human carrying status in visual surveillance [A]. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition [C], New York, USA, 2006: 1670-1677.

55 Davis J, Taylor S. Analysis and recognition of walking movements [A]. In: Proceedings of International Conference on Pattern Recognition [C], Quebec City, Canada, 2002: 315-318.

56 Lv F, Song X, Wu B, *et al*. Left luggage detection using Bayesian inference [A]. In: Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance [C], New York, USA, 2006: 83-90.

57 Auvinet E, Grossmann E, Rougier C, *et al*. Left-luggage detection using homographies and simple heuristics [A]. In: Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance [C], New York, USA, 2006: 51-58.

58 Weinland Daniel. The Multiple-video Data Used Here are from INRIA Rhone-Alpes' Multiple-camera Platform Grimage and Perception Research Group [DB/OL]. https://charibdis.inrialpes.fr/html/sequences.php. 2006.

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net