

DOI: 10.3969/j.issn.1006-4729.2017.01.022

人体行为的识别及发展

张行健

(上海电力学院 电子与信息工程学院, 上海 200090)

摘要:人的动作行为识别是计算机视觉领域中的热点研究课题之一,广泛应用于安防监控系统、医疗诊断监护、人机交互等方面.从运动对象检测、行为特征提取和人体行为识别方法方面阐述了人体行为识别技术的现状和相关难点.与已有相关综述性文章相比,结合了新的研究成果,介绍了基于轨迹能量扩散图的组行为识别方法以及基于稠密光流轨迹和稀疏编码算法的行为识别算法.且指出了未来行为识别的发展趋势.

关键词:动作行为识别; 运动对象检测; 特征提取; 识别方法

中图分类号: TP391.41

文献标志码: A

文章编号: 1006-4729(2017)01-0102-05

Human Body Motion Recognition and Development

ZHANG Xingjian

(School of Electronics and Information Engineering, Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: Human motion recognition is one of the hot topics in the field of computer vision, which is used widely in security monitoring system, medical diagnosis and monitoring, human-computer interaction, etc. The current state and related difficulties of human motion recognition technology is presented within the motion object detection, feature extraction and human motion recognition method. Compared with the existing relevant review articles, the new research results are combined, such as group behavior recognition method based on trajectory energy diffusion map and behavior recognition algorithm based on dense flow trajectory and sparse coding algorithm. Furthermore, some research direction of motion identification on future work is given.

Key words: action recognition; motion object detection; feature extraction; recognition method

随着科学技术的迅速发展,人工智能已经逐渐从梦想走进现实.计算机视觉技术是人工智能的重要组成部分,在人工智能的发展过程中发挥了重要作用.在获取外部世界的信息中,视觉是人类了解事物本质信息的重要途径.计算机视觉技术就是研究如何让计算机像人类一样对通过摄像头获取到的外部信息进行分析 and 识别.因此,静态

物体识别、目标跟踪和人体行为识别是计算机视觉研究的重要方向.

人体行为识别的目的是分析理解视频中个人的动作、多人之间的交互行为,作为计算机视觉研究的热点,人体行为识别研究具有重要的学术意义.同时,人体行为识别技术在安防监控系统^[1-2]、医疗诊断监护^[3]、人机交互^[4]等领域具有应用价值.由于人

收稿日期: 2016-03-16

通讯作者简介:张行健(1991-),男,在读硕士,上海人.主要研究方向为深度学习和模式识别. E-mail: txhtxh1991@163.com.

体行为识别在许多方面都具有广阔的应用意义,国内外科研者对该领域进行了研究,如美国国防高级研究项目署^[5]设立了视频监控重大项目用于战争和普通家庭的视频动作分析,先后投用了 W4^[6]和 Pfunder^[7]等监控系统;中国科学院开发了一套面向体育训练的三维人体运动模拟与视频分析系统用于国家队集训等^[8]. 除此之外,一些国际权威期刊,如 PAMI (IEEE Transactions on Pattern Analysis and Machine Intelligence), IVC (Image and Vision Computing) 和 CVIU (Computer Vision and Image Understanding) 为研究者提供了交流机会,大大促进了行为识别技术的发展.

相比于静态物体识别研究^[9],行为识别更加重视如何发现运动目标在图像序列中的空时运动变化. 动作行为从二维空间到三维时空的转化大大增加了行为描述及后续识别方式的复杂性,但也为研究者提供了更宽广的空间来使用不同的技术方法. 因此,对于人的行为识别研究在计算机视觉领域中成了热门话题^[10-12]. 人体行为识别的基本过程是通过计算机检测行为动作数据,继而从图像序列中提取和理解动作特征,以实现行为识别分类的过程.

本文将从运动对象检测、行为特征提取和人体行为识别方法等方面综述人体行为识别技术的现状. 同时结合新的研究成果,展望今后行为识别的研究方向.

1 运动对象检测

对于样本的预处理就是对运动对象的检测,即把视频转换成图像帧,并检测出目标的动作信息. 有效的运动目标检测对行为特征提取、人体行为表示和行为识别理解等后续的研究十分重要. 近年来,在检测运动目标的技术方面已有了很大的突破,目前主要的方法为光流法^[13-14]、时间差分法^[15]和背景减除法^[16-17]. 同时,预处理过程还包括使用形态学操作对获取的运动目标信息进行处理来使结果更加精确.

1.1 光流法

光流法^[18]的思想最先由 GIBSON 提出,光流是像素点运动产生的瞬时速度场,光流法是通过计算光流场从静态图像中分割出场景,确定运动目标的速度和方向等信息. 首先,近似估计真实运

动场景的速度. 图像中某点 (x, y) 在 t 时刻的灰度值用 $I(x, y, t)$ 表示. 让图像灰度值保持不变,即:

$$\frac{dI(x, y, t)}{dt} = 0 \quad (1)$$

则可以得到光流约束方程:

$$I_x u + I_y v + I_t = 0 \quad (2)$$

式中: u, v ——沿着 x 轴和 y 轴方向的分量.

接着通过设光流在整个图像上的光滑性一致来求解光流约束的解.

该方法的优点是在运动的摄像头下,也能检测出运动对象,且对于静态或动态背景都有很好的效果. 缺点是当运动对象和背景之间存在噪声时,由于复杂度很高,无法实时检测运动目标. BARRON J 等人详细给出了光流场计算的求解过程^[18]. 黄士科等人用新的 3D-Sobel 算子计算像素间的时空梯度,提高了不同速度运动目标检测的有效性^[14].

1.2 帧间差分法

帧间差分法(时间差分法)^[19]是通过图像序列中相邻 2 帧或者 3 帧的差值,获取变化区域来确定运动对象,如图 1 所示. 其优点是在全局运动补偿下,能适应动态变化的背景;缺点是存在空洞和重像现象,并受噪声影响. COLLINS R 等人通过对相邻 3 帧中的像素进行逐个差分运算和背景减除,除去了帧之间差分所出现的空洞重像现象^[5].

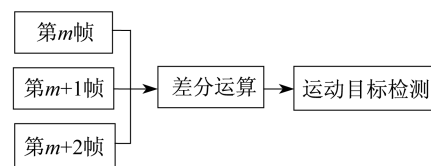


图 1 帧间差分检测示意

1.3 背景减除法

背景减除法^[20]是当前应用最为广泛的一种算法,它适用于摄像机停止的状况,其核心是对比背景图像和当前图像进行差分计算来确定运动目标区域. 它的关键是如何使背景图像很好地适合当下环境变化. 目前常用的算法有隐马尔可夫法,均值滤波法等. 其优点是能够保证对运动目标区域较完整的提取. 但在实践应用中,会受外部环境和背景干扰的影响. 因此,背景模型的准确选取及实时更新对于准确检测出运动目标有很大的影

响,现在主要采用高斯模型、混合高斯模型、非参数化模型等方法。

背景减除法一般包括预处理、建立背景模型、比较输入图像和背景图像的差异和后处理4个步骤^[21]。预处理是指对视频数据进行时空域的滤波,减少各种环境噪声,与此同时要符合背景减除算法的要求,将原视频的数据格式转换;建立背景模型是指创建某背景模型来代表当前背景,包括背景的初始化和实时更新,来消除光照条件的变化和环(如树枝的晃动和其他物体运动)带来的影响;输入图像和背景图像作比较是指减掉当前视频图像与背景图像之间的差异;后处理是指对上面步骤的前景区域作一些后处理,一般包括噪声消除,减少其他场景中的干扰,通过后处理,让所得的前景目标更加明确。

2 行为的特征选择及提取

人的行为特征的选择及提取是行为识别中的重要一环。其基本过程是从各种特征信息中找出符合研究阶段的特征,并通过相关技术提取出来。但是一般提取出的特征的维数相对较高,因而要对特征进行降维。根据人体行为的组成形式,目前人体行为特征提取方法可以划分为整体表示和局部表示两种方式。

2.1 整体表示方式

整体表示方式是吧视频帧数当作一个全体,检测出运动目标的位置,然后通过自上而下的方法提取到全局特征。全局特征的提取通常是在视频帧中采用背景减除法、时间差分法或者光流法检测到人体行为感兴趣区域(Region Of Interest, ROI),然后从中得到表示人体行为的特征,如轮廓信息、边缘信息等。该方法描述能力强,区分度好,包含了人体重心、长宽比例等重要信息,可以大大提高识别的效果。但是该方法提取精度不高且含有噪声,同时受摄像机角度和物体遮挡影响。WANG L 等人^[22]提出统计每块轮廓图的白色数目所占的比值作为特征模型,再利用隐马尔科夫模型(HMM)进行识别,如图2所示。DAVIS J W 和 BOBICK A F^[23]提出基于视频的轮廓提取方法,利用背景剪除法获取背景的轮廓特征,然后叠加所有帧的轮廓特征,对比前后帧之间轮廓的差异,建立了具有运动效果的运动能量图像(Motion Energy Image, MEI),如图3

所示。文献[24]也是利用视频中的人体轮廓序列特征来表示人体行为,该方法与文献[23]中提出的方法不同,是在人体轮廓特征上进行傅里叶计算获得傅里叶描述子,然后将其送入支持向量机(SVM)分类器训练并识别出姿态类别,最后将姿态标签序列一同交给HMM,使用HMM最终判断出视频中人体行为所属的类别。

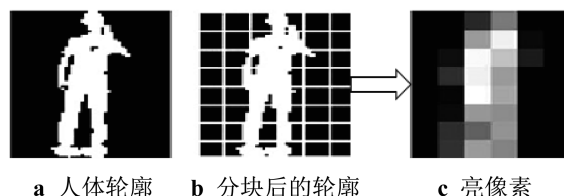


图2 轮廓网格像素统计

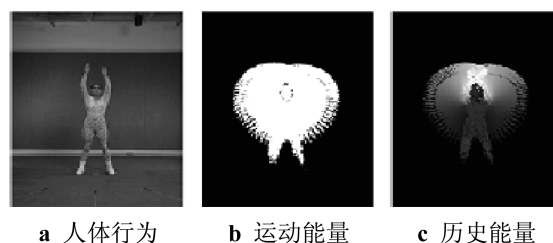


图3 人体行为模型

使用轮廓表示人体行为的方法虽然已经获得了很好的识别性能,但影响这一特征的两个关键问题仍旧没有得到很好的解决,即人体行为轮廓提取及帧序列中轮廓的跟踪问题。为了使该特征能精准地描述人体行为特性,研究者专门进行了研究并取得了一些成果。为了解决视频帧序列间的人体轮廓跟踪问题,文献[25]将相邻帧轮廓的形状及梯度像素点作为图的节点,并采用二分图分割算法将相邻帧之间轮廓点进行匹配,该方法的优点是精度好,缺点是需要人工干预,且不能解决遮挡等问题。文献[26]和文献[27]结合颜色、纹理等特征来对人体行为区域进行分割,文献[7]利用人体运动的特性来实现分割,但是这些都只能在有条件限制的场景才能得到很好的分割效果,并且可能需要人工干预才可以实现,因此在实际应用中很难直接利用轮廓特征方法来对人体行为进行识别。

2.2 局部表示方式

与整体表示对应的是局部表示方式,它是将视频段作为一个整体,通过自下而上的方法获得

局部特征. 开始在整个时空域内检测兴趣点, 然后计算出邻域内的特征点, 最后将所有视频段整合在一起形成总特征. 该表示方式与整体表示方式相比较, 不需要前景检测、ROI 兴趣区域的定位和分割, 所以不受前景检测、背景更新、以及 ROI 区域定位结果的影响, 并且对互相遮挡及噪声不敏感. 但是需要在一段视频中采集到关键的兴趣点, 因此摄像机运动产生的误差会对兴趣点采集造成一定的影响.

3 人体行为识别方法

3.1 基于轨迹能量扩散图的行为识别方法

目前, 在提取组行为的特征过程中, 多数方法忽视了轨迹的时间信息(如不能得出某一时刻检测对象的位置), 虽然一些算法, 如隐马尔克夫模型, 可以保留轨迹的时间信息, 但不能得到大量的训练信息^[28]. 而且因为人的运动变化产生了噪声, 对于识别效果有很大影响. 同时一些算法采用高斯估计或滤波法来处理噪声^[29], 但没有考虑保存轨迹的时间信息.

因此, 本文提出采用轨迹能量块图来保留组行为的轨迹时间信息, 用轨迹能量扩散图来解决因运动目标的运动变化产生的噪声. 同时, 结合自适应分层直方图(AHDH)和模糊形状模型(BSM)算法分别对轨迹能量等势线图提取特征. 最后, 基于典型相关分析的特征融合法获取融合特征向量, 选择最优的特征, 用支持向量机对特征分类, 实现行为识别. 采用轨迹能量扩散图的行为识别方法虽然解决了轨迹时间信息的丢失以及噪声问题, 但此方法只适合于简单的组行为分类识别, 并不能应用于更为复杂的组行为识别. 对于更为复杂的组行为识别是我们攻克的一个方向.

3.2 基于稠密光流轨迹和稀疏编码的行为识别方法

为了解决现有特征提取发放识别低的问题, 采用基于一种融合稠密光流轨迹和稀疏编码的无监督特征提取方法^[30-31], 然后利用词袋模型^[32]获得行为特征表示, 最后采用支持向量机进行动作特征分类, 大大提高了识别的准确率. 其主要任务是基于稠密光流轨迹提取原图像特征的情况下,

采用系数编码的无监督学习方法, 获得行为特征表示, 从而实现行为分类识别. 该方法的优点: 一是自动提取检测目标的特征, 避免了人工提取特征的复杂性; 二是稀疏的特征表示减小了后续运算的繁琐度. 但其也有不足之处: 优化稀疏编码需要消耗大量的内存和计算时间, 对电脑 CPU 的性能和内存有一定的要求. 后续的研究方向是优化系数编码计算行为特征的迭代, 以及采用并行算法加速迭代过程.

4 结 语

本文从对运动对象检测、行为特征提取和人体行为识别方法等方面阐述了人体行为识别技术的现状和相关难点.

当前研究对人体行为的检测问题进行了回避与简化, 特别在对其中关键人体行为动作及起始、终止时间的动作发生不能精确检测, 研究人员需要结合更多中高层的行为特征和底层特征, 把行为识别用于更具挑战性的真实场景中. 另外, 相比于其他视觉识别任务, 在人体行为识别上, 由于时间维度的加入, 行为特征的维度变得更高, 所需样本数量也更多, 深度学习没有完全取得显著提高. 因此, 结合递归神经网络(RNN)和深度学习的方法在时序数据中的前景将成为研究热点.

参考文献:

- [1] AYERS D, SHAH M. Monitoring human behavior from video taken in an office environment[J]. *Image and Vision Computing*, 2001, 19(12): 833-846.
- [2] LÓPEZ M T, FERNÁNDEZ-CABALLERO A, FERNÁNDEZ M A, et al. Visual surveillance by dynamic visual attention method[J]. *Pattern Recognition*, 2006, 39(11): 2 194-2 211.
- [3] 杜友田, 陈锋, 徐立文. 基于视觉的人的运动识别综述[J]. *电子学报*, 2007, 35(1): 84-90.
- [4] 黎洪松, 李达. 人体运动分析研究的若干新进展[J]. *模式识别与人工智能*, 2009, 22(1): 70-78.
- [5] COLLINS R, LIPTON A, KANADE T, et al. A system for video surveillance and monitoring: VSAM final report[R]. Carnegie Mellon University: CMU-R1-TR-00-12, 2000.
- [6] HARITA OGLU I, HARWOOD D, DAVIS L. W4: Real time surveillance of people and their activities[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 809-830.
- [7] WREN C, AZARBAYEJANI A, DARRELL T, et al. Pffinder: Real time tracking of the human body[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 780-

- 785.
- [8] 王兆其,张勇东,夏时洪. 体育训练三维人体动作模拟与视频分析系统[J]. 计算机研究与发展, 2005, 42(2): 344-352.
- [9] 黄凯奇,任伟强,谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6): 1-18.
- [10] TURAGA P, CHELLAPPA R, SUBRAHMANIAN V, *et al.* Machine recognition of human activities: a survey[J]. IEEE Trans on Circuits and Systems for Video Technology, 2008, 18(11): 1473-1488.
- [11] POPPE R. A survey on vision-based human action recognition[J]. Image and Vision Computing, 2010, 28(6): 976-990.
- [12] CANDAMO J, SHREVE M, GOLDFOG D, *et al.* Understanding transit scenes: a survey on human behavior recognition algorithms[J]. IEEE Trans on Intelligent Transportation System, 2010, 11(1): 206-224.
- [13] ADIV G. Determining three-dimensional motion and structure from optical flow generated by several moving objects[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1985, 7(4): 384-401.
- [14] 黄士科,陶琳,张天序. 一种改进的基于光流的运动目标检测算法[J]. 华中科技大学学报: 自然科学版, 2005, 33(5): 39-41.
- [15] KIM M, CHOI J G, KIM D. Automatic segmentation of moving objects in image sequences based on spatio-temporal information[J]. IEEE Trans on Circuits and Systems for Video Technology, 1998, 9(8): 1216-1226.
- [16] PAVLIDIS T, MORELLAS V, TSIAM P, *et al.* Urban surveillance systems from the laboratory to the commercial world[C]//Proceedings of the IEEE, 2001: 1478-1497.
- [17] ELGAMMAL A M, HARWOOD D, DAVIS L. Non-parametric model for background subtraction[C]//Proceedings of the 6th European Conference on Computer Vision, 2000: 751-767.
- [18] BARRON J, FLEET D, BEAUCHEMIN S. Performance of optical flow techniques[J]. International Journal of Computer Vision, 1994, 12(1): 42-77.
- [19] LIPTON A J, FUJIYOSHI H, PATIL R S. Moving target classification and tracking from real-time video[C]//Proceedings of the IEEE Workshop Application of Computer Vision, 1998: 8-14.
- [20] STAUFFER C, GRIMSON WEL. Learning patterns of activity using real-time tracking[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(8): 747-757.
- [21] 代科学,李国辉,涂丹,等. 监控视频运动目标减背景技术及研究现状和展望[J]. 中国图像图形学报, 2006(11): 7-10.
- [22] WANG L, SUTER D. Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model[C]//IEEE Conference on In Computer Vision and Pattern Recognition, 2007: 1-8.
- [23] DAVIS J W, BOBICK A F. The representation and recognition of human movement using temporal templates[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Puerto Rico; IEEE, 1997: 928-934.
- [24] KELLOKUMPU V, PIETIKAINEN M, HEIKKILA J. Human activity recognition using sequences of postures[C]//Proceedings of the IAPR Conference on Machine Cision Applications. Japan; IEEE, 2005: 570-573.
- [25] WINAWER J, HUK A C, BORODITSKY L. A motion after-effect from still photographs depicting motion[J]. Psychological Science, 2008, 19(3): 276-283.
- [26] XU G Y, HUANG F Y. Viewpoint insensitive action recognition using envelop shape[M]. Computer Vision. Japan: Springer, 2007: 477-486.
- [27] STEIN A, HOIEM D, HEBERT M. Learning to find object boundaries using motion cues[C]//IEEE 11th International Conference on Computer Vision. Brazil; IEEE, 2007: 1-8.
- [28] LIU X H, CHUA C S. Multi-agent activity recognition using observation decomposed hidden Markov model[J]. Image and Vision Computing, 2005, 24(2): 166-175.
- [29] CHENG Zhongwei, QIN Lei, HUANG Qingming, *et al.* Group activity recognition by Gaussian processes estimation[C]//Proc of Pattern Recognition. Istanbul; IEEE, 2010: 3221-3231.
- [30] OLSHAUSEN B A, FIELD D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images[J]. Nature, 1996, 381: 607-609.
- [31] OLSHAUSEN B A, FIELD D J. Sparse coding with an over-complete basis set: a strategy employed by V1[J]. Vision Research, 1997, 37(23): 3311-3325.
- [32] NIEBLES J C, WANG H, LI F. Unsupervised learning of human action categories using spatial-temporal words[J]. International Journal of Computer Vision, 2008, 79(3): 299-318.

(编辑 桂金星)