

人体动作姿态识别综述

黄国范,李亚

(南阳师范学院 软件学院,河南 南阳 473061)

摘要: 人体动作姿态的识别是当前的研究热点,该文主要从人体动作姿态的分类和人体动作姿态识别的方法两个方面进行了介绍,并重点阐述了每种识别方法的研究进展情况及其优缺点,最后对当前研究的难点问题以及未来的发展趋势也进行了较为深刻的阐述。

关键词: 人体动作姿态识别; 人工智能; 隐马尔可夫模型; 动态贝叶斯网络; 模板匹配

中图分类号: TP391 **文献标识码:** A **文章编号:** 1009-3044(2013)01-0133-03

A Survey of Human Action and Pose Recognition

HUANG Guo-fan, LI Ya

(College of Software, Nanyang Normal University, Nanyang 473061, China)

Abstract: The recognition of human action and pose is receiving increasing attention. This paper introduces two aspects: the classification of human action and pose and recognition methods of human action and pose, and focuses on the the research progress and their advantages and disadvantages of each method. At the end, some detailed discussions on research challenges and future directions in human action and pose recognition are also provided.

Key words: human action and pose recognition; artificial intelligence; HMM; DBN; template matching

人体动作姿态识别是计算机视觉研究领域中最具挑战的研究方向,是当前的研究热点。对人体动作姿态进行自动识别将带来一种全新的交互方式,通过身体语言即人体的姿态和动作来传达用户的意思,如在机场、工厂等喧闹的环境下,采用手势、动作姿态识别等人机交互技术能够提供比语音识别更加准确的信息输入。总之,在智能监控、虚拟现实、感知用户接口以及基于内容的视频检索等领域,人体动作姿态的识别均具有广泛的应用前景。该文首先简单介绍了人体动作姿态的分类,然后对人体动作姿态识别的方法进行了分类介绍,并对每种识别方法的研究进展情况及其优缺点进行了重点介绍,最后阐述了人体动作姿态识别当前的研究难点问题以及未来的发展趋势,希望能对相关领域的研究人员有所裨益。

1 人体动作姿态的分类

人体的运动涉及很多部位,包括人整个身体的运动或某个部位的运动,如手部运动、腿部运动或者头部运动,甚至人的表情、手势等。这些运动分别具有自己不同的特点和不同的复杂度,所用的研究方法也都不同。而针对人体简单动作姿态的识别是当前研究的一个热点,也是本文讨论的重点。人体运动的识别可分为两类:姿态识别和动作过程识别。

姿态识别的对象是静态系统,主要是识别人体整体或者某一部位的姿势。根据识别对象的不同,可分为手形识别、体势识别、头部姿态识别等。

动作过程识别是识别人体运动的动态过程,比如人体的动作识别、步法识别以及手势识别等。

2 人体动作姿态识别的方法

人体动作姿态的识别方法可分为三类:基于统计的方法、基于模板的方法和基于语法的方法。具体如图1所示。

2.1 基于统计的方法

基于统计的方法是动态识别系统中效果最好的方法,目前主要是隐马尔可夫模型(HMMs)和动态贝叶斯网络(DBN)两种方法。

2.1.1 隐马尔可夫模型

隐马尔可夫模型是目前应用最广的一种方法。HMMs由状态和观测两部分组成,是一种基于转移概率和传输概率的随机模型,系统当前所处状态的概率只与前一个时刻的状态有关,与其它历史状态条件无关。在HMMs的识别过程中,首先提取出特征向量

收稿日期:2012-11-25

作者简介:黄国范(1981-),女,硕士,助教,河南人,研究方向为图象处理;李亚(1984-),女,硕士,助教,河南人,研究方向为数据挖掘。

本栏目责任编辑:唐一东

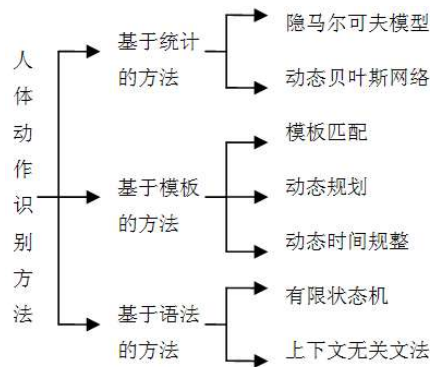


图1 人体动作姿态识别方法分类

序列,然后通过学习算法进行模型参数训练,最后对未知的运动序列进行识别分类。

Yamato 等人^[1]首次用HMMs进行人体动作姿态的识别,之后又出现了HMMs的各种改进模型并成为人体动作姿态识别的主流方法^[2-4]。Brand 等人^[2]提出了耦合隐马尔可夫模型(CHMM),并对手语进行了识别。Nguyen^[3]提出了分层隐马尔可夫模型(HHMMs),HHMMs具有多层隐马尔可夫模型的结构,能够清楚地表达出人体运动中的行为细节。Duong 等人^[4]提出了S-HSMM(Switching Hidden-semi Markov Model)模型,S-HSMM的识别性能高于HHMMs,但是算法复杂度较高。

上述各种隐马尔可夫模型有两个缺点:第一,由于HMMs只适合于描述单个运动物体的系统,随着运动物体数目的增加,系统的状态数目以及HMM模型中的状态转移矩阵和观测模型的参数都呈指数增长,极大地增加了系统的计算复杂度,这样,HMMs的应用范围就有很大的局限性;第二,HMMs模型无法进行信息的有效融合,由此导致参数误差的相互叠加,进而无法保证模型参数的精确度。

2.1.2 动态贝叶斯网络

动态贝叶斯网络(DBN)是近年发展起来的统计模型,能够学习变量间的概率依存关系及其随时间变化的规律,具有很好的可扩展性和可解释性,对于多信息融合推理、多物体动态系统识别非常有效,是人体动作姿态识别今后发展的方向。

在动态识别过程中,DBN可以针对具体问题设计出相应的网络结构,而且可以将各种信息进行融合推理,同时为了使信息的配置更加优化,还可以对各种不同信息根据其来源、置信度等设置不同的权重。因此不少学者已经开始了这个方向的研究,将其用于人体动作姿态的识别,并取得了一定的成就^[5-6]。而且,也有部分学者将HMM和动态贝叶斯网络两者结合起来,高效地进行人体动作姿态的建模和识别^[7-8]。

但是,目前动态贝叶斯网络在人体动作姿态识别中的应用并不是很多,也还不够深入,主要原因有:

1)动态贝叶斯网络主要应用于识别多个物体的动态系统,而这样的系统需要大量的特征参数,所以动态贝叶斯网络的方法计算量较大,复杂度较高。

2)对目前主流的简单环境中人体规范动作姿态的识别,HMMs等方法已经可以胜任,达到较好的效果。

2.2 基于模板的方法

基于模板的方法主要是模板匹配法、动态时空规整法和动态规划法。

模板匹配法的思想是首先对人体动作姿态进行训练建立模板库,然后将待识别的动作姿态与模板进行匹配,计算二者之间的相似度,如文献[9]。模板匹配法的优点是实现起来比较简单、计算复杂度较低,缺点是对噪声和动作姿态持续时间的变化比较敏感。

动态时空规整法是一种非线性时间规整方法,它是将待识别的人体动作姿态模板的时间轴非线性地映射到训练模板的时间轴上,从而能够使两者的距离最小,如文献[10]。动态时空规整法较好地解决了人体动作姿态在时间尺度上的不确定性。

在动态规划算法的识别过程中,待识别样本模板中的每个时刻特征可以与特征模板中的任意时刻特征进行匹配,二者都无须进行时间规整,但要求必须顺序进行,如文献[11]。动态规划算法的主要缺点是计算量会随着训练样本数目的增加而增加。

2.3 基于语法的方法

基于语法的方法主要是有限状态机^[12]以及上下文无关的文法^[13]等。当前已有不少学者开始了这个方面的研究,用自然语言对人体动作姿态进行描述是未来的发展趋势。

但是,对图像序列中的人体动作姿态进行描述是非常复杂的,而且对事件、行为、状态这些概念还存在着不同的理解,如何对场景内容进行充分有效的表达还是非常困难的。目前对人体动作姿态的描述还仅限于简单的语义解释,对复杂场景中人体动作姿态的语义描述工作还相当的艰巨。

3 难点与发展趋势

近年来,人体动作姿态的识别研究已经取得了非常重要的成果,展现出了良好的发展前景。但下述几个方面仍是今后研究的难点问题:

人体动作类别的限制。目前人体动作姿态识别的研究主要还是集中在简单的人体动作姿态上,如人的某些标准姿势和一些简单的动作行为,如走、跑、蹲、站等简单规范的动作,而且,识别的对象通常是针对单个人体的运动。所以,如何针对在复杂场景下的多个人之间的交互动作行为进行识别,是未来的一个发展趋势。

特征选择的困难。特征向量的选择非常关键,直接影响到所采用的识别方法以及识别性能。如果选取过多的特征,特征向量维数就会过大,相应地就会增加计算的复杂度,相反,如果选取的特征过少,又可能无法获取较高的识别准确率。因此必须在选取合适的特征和获取较高的识别准确率之间进行折中权衡。

机器学习的局限性。利用机器学习工具进行人体动作行为的识别取得了一定的进展,但仍然还处于初级阶段,在识别过程中对人体运动的模型需要增加约束条件来减少歧义性,而这些约束条件与一般的现实情形通常是不吻合的。而且,机器学习的方法要求广泛的训练,即每个动作姿态都需要经过大量的训练才能识别,因此,机器学习方法只能识别一些预先定义的动作集,没有学习新动作的能力。因此机器学习仍是人体动作行为识别的一个难点问题。

识别算法的性能评价。鲁棒性、实时性以及准确率是衡量一个算法性能的三个重要指标。在当前研究工作中,准确率是最受关注的,而另外两个指标关注较少,特别是算法的鲁棒性。但算法的鲁棒性和实时性是实际应用中必须考虑的问题,因此,在衡量一个算法的性能时,必须对这个三个方面进行综合考虑。因此,在考虑某个具体应用时,需要开发新技术既能提高识别的性能,同时又可以降低算法的计算复杂度。

4 结束语

人体动作姿态识别本质上是一个人工智能问题,在智能监控、虚拟现实以及感知用户接口等领域具有广泛的应用前景。该文简单介绍了人体动作姿态的分类和人体动作姿态识别的方法,并对各种识别方法的研究进展情况以及优缺点进行了详细介绍,最后对当前的研究难点和未来的发展趋势做了详细阐述,希望能对相关领域的研究人员有所裨益。

参考文献:

- [1] Yamato J, Ohya J, Ishii K. Recognizing human action in time sequential images using Hidden Markov model [A]. In Proc CVPR[C], IEEE, 1992:379-385.
- [2] Brand M, Oliver N, Pentland A. Coupled hidden markov models for complex action recognition [A]. In Proc CVPR [C] .USA :IEEE Computer Society Press, 1997:994-999.
- [3] Nguyen N T, Phung D Q, Venkatesh S, et al. Learning and detecting activities from movement trajectories using the hierarchical hidden Markov model [A]. In Proc CVPR[C]. San Diego: IEEE Computer Society Press, 2005:955-960.
- [4] Duong T V, Bui H H, Phung D Q, et al. Activity recognition and abnormality detection with the switching hidden semi-Markov model [A]. In Proc CVPR[C]. San Diego: IEEE Computer Society Press, 2005:838-845.
- [5] Luo Y, Wu T D, Hwang J N. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks[J]. CVIU, 2003, 92(2):196-216.
- [6] Gong S, Xiang T. Recognition of group activities using dynamic probabilistic networks [A]. In Proc ICCV [C]. Washington: IEEE Computer Society Press, 2003:742-749.
- [7] Oliver N, Horvitz E. A comparison of HMMs and dynamic Bayesian networks for recognizing office activities [J] . Lecture Notes in Artificial Intelligence, 2005: 199-209
- [8] N. Robertson, I. Reid. Behaviour understanding in video: a combined method. In: International Conference on Computer Vision, 2005:15-21.
- [9] 谢林海, 刘相滨. 基于不变矩特征和神经网络的步态识别[J]. 微计算机信息, 2007, 23(7-1):279-281.
- [10] Bobick A F, Wilson A D. A state-based approach to the representation and recognition of gesture [J] .IEEE Trans PAMI, 1997, 19(12): 1325-1337.
- [11] Bobick A F, Andy Wilson. Using configuration states for the representation and recognition of gestures. MIT Media Lab Perceptual Computing Section Technical Report, No. 308, 1995.
- [12] Pengyu Hong, Turk M, Huang T.S. Constructing finite state machines for fast gesture recognition. Proceedings of 15th International Conference on Pattern Recognition, 2000, 3: 691-694.
- [13] Yamamoto M, Mitomi H, Fujiwara F, et al. Bayesian classification of task-oriented actions based on stochastic context-free grammar. In: International Conference on Automatic Face and Gesture Recognition, Southampton, UK, April 10-12, 2006.