View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Xiamen University Institutional Repository

文章编号:1008-7826(2011)04-0023-04

基于单张静态图像的人体行为识别方法综述

姜夕凯[1,2] , 苏松志[1,2] , 李绍滋[1,2], 成 运[3]

(1. 厦门大学 信息科学与技术学院, 福建 厦门 361005; 2. 福建省仿脑智能重点实验室(厦门大学), 福建 厦门 361005; 3. 湖南人文科技学院 通信与控制工程系, 湖南 娄底 417000)

摘 要: 人体行为识别是计算机视觉的研究难点与热点,目前大部分研究者主要针对视频中的行为展开研究. 然而,人类的视觉往往根据单张图片就可判断图片中发生的行为. 基于单张静态图像的人体行为识别,挑战性更大,是近年来人体行为识别研究的一个趋势,更是探索人类视觉奥秘的一个很好切入点. 本文对单张静态图像的人体行为识别方法进行梳理,将其分为三类,最后对其未来研究方向进行展望.

关键词: 计算机视觉 ; 行为识别 ; 静态图像

中图分类号: TP182 **文献标识码:** A

A Survey of Recognizing Action from Single Still Images

JIANG Xi-kai[1,2] , SU Song-zhi[1,2] , LI Shao-zi[1,2] , CHENG Yun[3]

(1. School of Information Science and Technology, Xiamen University, Xiamen, Fujian 361005 , China; 2. Fujian Key Laboratory of the Brain-like Intelligent Systems (Xiamen University), Xiamen, Fujian 361005 , China; 3. Department of Communication and Control Engineering Hunan Institute of Humanities, Science and Technology, Loudi, Hunan 417000, China)

Abstract: Human action recognition is a difficult and active research area in computer vision. At present, most of researchers in this field focus on recognizing action from video. However, human can understand human action based on a single picture. Recognize action from single still images has more challenge and is a trend in action recognition in recent years, but also a good entry point to explore the mysteries of human vision. In this paper, we sort out the methods of recognizing action from single still images and classify these methods into three categories. At last, the future research directions are discussed.

Key words: computer vision ; action recognition ; still images

1 引言

行为识别在计算机视觉中是一个非常活跃的研究课题,它有着许多重要的应用,如人机界面、基于内容的图像/视频检索、智能视频监控、家庭服务机器人等等. 人的行为识别可以理解为是对个体行为、人与人之间以及人与对象之间的交互行为的识别和表示. 但是由于客观环境的多样性以及人体行为的复杂性,想要在计算机视觉中实现行为识别变得非常困难.

过去大部分关于行为识别的研究集中在如何从一段未知的视频中识别正在进行的行为[1,2,3],然而静态图像的人体行为识别的相关研究却非常少. 在视频这项技术没有发明的很长一段时间里,都是静态图像

收稿日期: 2011-11-11

基金项目: 国家自然科学基金项目(60873179); 高等学校博士学科点专项科研基金项目(20090121110032); 深圳市科技计划项目-基础研究(JC200903180630A); 深圳市科技研发基金项目-深港创新圈计划(ZYB200907110169A); 湖南省科技厅科研项目(2010TC2006)和教育厅科研项目资助(09A046)

作者简介: 姜夕凯(1986-), 男, 在读研究生.

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

在向我们传递行为信息，我们可以很容易的识别出静态图片中人的行为. 所以静态图像的人体行为识别是可行的，而且也有很好的应用前景，比如说新闻和体育图片的检索，静态图像行为识别的研究成果也可以直接应用到视频情况下，推动视频中行为识别方面的发展. 然而，相对于视频中行为识别，静态图像中行为识别更加困难和具有挑战性. 在视频中，“动作”线索可以提供充足的信息来进行行为识别，但是静态图像中可以依靠的信息却是非常少的. 目前对于静态图像行为识别的研究也慢慢的得到关注，许多关于这方面问题的解决方法也已经提出，本文主要内容就是对这些方法的一个概述和分类.

但是，静态图像的人体行为识别不能简单地看作是一个图片分类问题. 因为: (1)同一类的图片中可能存在不同的行为，比如说踢足球图片类中可能有跑、跳等多种行为；(2)即使是同一行为类的图片，也可根据图片之间的细小差别将它们看做不同行为类的图片. 比如说演奏某一乐器的图片集，我们可以将正在演奏乐器和没有演奏只是拿着乐器的图片看做是不同的行为类.

基于上述考虑，本文将静态图像中人体行为识别的方法分成了三类:基于人体形状或姿势的识别方法、基于相互作用的识别方法和其他识别方法.

2 基于人体形状或姿势的识别方法

基于人体形状或姿势的识别方法是利用静态图像中人体形状或姿势的表示信息进行行为识别的方法. Wang et al.[4]使用无监督的方法来实现静态图像中的人体行为识别. 作者使用图像中人体的大体形状来匹配图像对，并使用线性规划松弛技术来计算图像之间的距离，利用从可变形形状匹配中得到的距离信息来对不同的人体姿势进行聚类. Thurau et al.[5]使用姿势基元直方图来表示行为，姿势基元通过非负矩阵分解得到的. 文中提到的方法，在学习模式下，对于姿势和行为的参数化表示是通过视频进行估计的. 但在运行模式中，此方法在视频和静态图像的情况下都是适用的. Ikizler et al.[6]使用解析概率图中矩形区域的空间和方向直方图来表示姿势，使用了线性判别分析（LDA）来获得一个更简洁和具有识别力的特征. 文中的方法可用于无监督环境下. Ikizler-Cinbis et al.[7]想要实现在无监督的情况下从网络上搜集的图像中获得行为表示，并且将其应用到视频中行为的自动注释. 文中使用了概率边界运算符（pb）和 HOG[8]描述符来进行行为的估计.

以上方法的优点是，可以很好的利用姿势和形状估计领域的理论和方法来实现静态图像中的行为识别. 缺点是它们都假设图像的表示基于整体的模板，即使用从整个图像中提取出来的一个特征描述符来表示这个图像. 这种表示方法由于其在行人检测方面的成功而变得非常受欢迎，尤其是在 Dalal 和 Triggs[8]提出梯度方向直方图（HOG）后. 该表示方法适用于行人检测，因为大多数行人是直立行走的，所以可以使用一个整体的模板来表示所有行人. 但是对于行为识别来说，用一个整体的模板来表示变化度非常大的行为时就显得不是那么的灵活了.

3 基于相互作用的识别方法

基于相互作用的识别方法是指利用图像中人和对象或者人体姿势和对象的相互作用来构造行为分类器.

Gupa et al.[9]利用贝叶斯方法集成各种感知元素，例如对象识别、场景理解，并以此来获取图像中人和对象的相互作用. 而且对每一个感知元素使用了空间和函数限制来达到连贯语意解释. Yao et al.[10]认为图像中的人体姿势和对象可以看做是互为环境，对其中一个元素的识别会有助于另一个元素的识别. 在这篇文章中，作者提出了一个新的随机场模型来编码人和对象相互环境. 文中将模型学习任务看做是一个结构

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

学习问题，对象、人体姿势和身体各部分的结构连接通过一个结构搜寻方法来进行估计，模型参数的估计是通过一个新的最大利润算法实现的. 文章[13]是作者对[10]的延续，作者认为图像中的人体姿势和对象可以互为环境，使用一个相互环境模型来结合的对图像中的对象和人体姿势进行建模. 该文的目标不仅是实现静态图像的行为识别，还希望实现计算不同行为图像之间的相似性. 与传统行为识别不同，相似性计算是为了计算不同行为图像之间因为图像中人体姿势和对象等因素的变化而产生的差异. 该文利用静态图像中的对象和人体姿态信息进行行为图像之间相似性的测量，而且得到了很好的效果. Desai et al.[11]利用人和对象的相互关系来表示行为，使用一个统一的判别模型来进行行为识别. Prest et al.[12]介绍了一个弱监督方法来构造图像中人和对象之间的相互关系来进行行为识别.

上面提到的方法会受到错误的对象检测和行为估计的影响. 但是，与基于人体形状或姿势的方法相比，基于相互作用的方法更好的利用了图像中提供的信息进行行为识别. 通常我们日常生活中的行为几乎都伴随着与对象的相互作用，例如打电话、打篮球等，所以行为并不是孤立进行的，而是有目的的与附近的对象的相互作用. 因此，对于图像中对象的很好的识别和位置确定对于行为识别有很大的帮助.

4　其他识别方法

这一节我们将举例介绍一些实现静态图像人体行为识别的其他方法. Li et al.[14]通过对静态图像中出现的场景和对象进行识别和分类来进行行为识别. 然而该文并没有像基于相互作用识别方法所提到的那样，使用图像中元素之间的相互作用和相互关系. 相反，作者只是孤立的使用图像中的场景和对象信息，忽略了图像中对象之间和人与对象之间的位置和相互作用信息. 这样导致的结果是，如果一副图像中被识别出有人和山，那么不管这个人是否在山上或者是否在登山，该文的算法都会将这幅图像标记为登山行为. Yao et al.[15]想要实现分辨出同一类的人与对象相互作用行为的静态图片之间的差别. 比如，给定两幅图像，一幅是人在演奏小提琴，另一幅是人拿着小提琴并不演奏，该文的目的就是区别出这样的图像. 大多数解决此类问题的基本思路是识别出图像中人的姿势和对象，以及姿势和对象之间的相互位置关系. 但是，目前大多数姿势估计算法对于身体各部分的检测精度并不能达到解决此类问题的要求，尤其是在图像中存在部分遮挡和背景混乱的情况下. 对象检测也面临同样问题. 所以该文提出了一种新的图像特征表示“grouplet”. 通过编码图像中的可分辨图像特征和它们的空间配置，使得“grouplet”具有图像的结构化信息. Yang et al.[16]提出了一种新颖的方法来识别静态图像中的行为. 以前的方法将姿势识别和行为识别看做是两个分离的学习问题，姿势识别算法的输出是行为识别算法的输入. 但是该文却将姿势识别和行为识别相结合，来达到最终行为识别的目的. 该文将图像中人体的姿势看做“潜变量”，而且选择基于样例的姿势表示方法“poselet”来表示姿势. “poselet”首先在[19]中提出，是用来表示一组有相同三维姿势结构的图像块. Yao et al.[17]使用密集特征表示来实现图像的分类，文中作者使用随机森林和决策树算法来确定可分辨图像区域. 与传统的决策树算法不同，该文算法在每一个树的节点处使用强分类器，而且在树的不同深度处结合信息，以达到对密集抽样空间的有效挖掘. Subhransu et al.[18]使用“姿态子激活向量”(poselet activation vector)对静态图像中的对象和人的姿势实现分布式表示. “姿态子激活向量”是建立在前面提到的 poselet 的基础上的. 文中证明了这种方法可以用来进行姿势估计，而且对于遮挡、视角变换有很好的鲁棒性. 文中还提到了将这种表示与图像中其他信息结合，这些信息包括图像中人与对象的相互作用和图像中其他人的行为，以此来进行行为识别.

5　总结与展望

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

本文章对静态图像的人体行为识别方法进行了一个综述. 我们将这些识别方法分成了三类，而且对每一类方法的研究现状都进行了介绍和分析. 从本文可以看出，静态图像的人体行为识别相对于视频中的人体行为识别来说还是有一定难度的，目前对于这方面的研究还存在许多难点和问题，比如如何利用静态图像中的有限信息来很好的表示和识别行为,如何从静态图像中恢复出人体的姿态，如何构建图像中人和物体之间的交互模型. 但是，随着这方面问题越来越多的得到人们的关注，不久的将来相信这些问题和难点都是可以得到解决的.

6 致谢

本文受国家自然科学基金项目(60873179)、高等学校博士学科点专项科研基金项目(20090121110032)、深圳市科技计划项目-基础研究(JC200903180630A)、深圳市科技研发基金项目-深港创新圈计划(ZYB200907110169A)，湖南省科技厅科研项目(2010TC2006)和教育厅科研项目(09A046)资助.

参考文献：

[1] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies [C]. In CVPR, 2008.

[2] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words [C]. In BMVC,volume 3, pages 1249–1258, 2006.

[3] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: a local SVM approach [C]. In ICPR, volume 3, pages 32–36, 2004.

[4] Y. Wang, H. Jiang, M. S. Drew, Z.-N. Li, and G. Mori. Unsupervised discovery of action classes[C]. In CVPR, 2006.

[5] C. Thurau and V. Hlaváč. Pose primitive based human action recognition in videos or still images[C]. In CVPR, 2008.

[6] N. Ikizler, R. G. Cinbis, S. Pehlivan,and P. Duygulu.Recognizing actions from still images[C]. In ICPR, 2008.

[7] N. Ikizler-Cinbis, R. G. Cinbis, and S. Sclaroff. Learning actions from the web [C]. In ICCV, 2009.

[8] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection[C]. In CVPR, 2005.

[9] A. Gupta, A. Kembhavi, and L. S. Davis. Observing human-object interactions: Using spatial and functional compatibility for recognition [J]. IEEE T. Pattern Analysis and Machine Intelligent, 31(10): 1775–1789, 2009.

[10] B. Yao and L. Fei-Fei. Modeling mutual context of object and human pose in human-object interaction activities[C]. In CVPR, 2010.

[11] C. Desai, D. Ramanan, and C. Fowlkes. Discriminative models for static human-object interactions[C]. In SMiCV, 2010.

[12] A. Prest, C. Schmid, and V. Ferrari. Weakly supervised learning of interactions between humans and objects [R]. Technical report, INRIA, 2010.

[13] B. Yao, A. Khosla and L. Fei-Fei. Classifying Action and Measuring Action Similarity by Modeling the Mutual Context of Object and Human Poses[C]. In ICML,2011.

[14] L. J. Li and L. Fei-Fei. What, where and who? Classifying events by scene and object recognition[C]. In ICCV, 2007.

[15] B. Yao and L. Fei-Fei. Grouplet: A structured image representation for recognizing human and object interactions[C]. In CVPR, 2010.

[16] W. Yang, Y. Wang, and G. Mori. Recognizing human actions from still images in latent poses [C]. In CVPR, 2010.

[17] B. Yao, A. Khosla, and L. Fei-Fei. Combining randomization and discrimination for fine-grained image categorization[C]. In CVPR, 2011.

[18] S.Maji, L.Bourdev, and J.Malik. Action Recognition from a Distributed Representation of Pose and Appearance[C]. In CVPR,2011.

[19] L. Bourdev and J. Malik. Poselets: Body part detectors training using 3d human pose annotations[C]. In ICCV, 2009.

[责任编辑：林宝德]

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net