CN43-1258/TP
ISSN 1007-130X

计算机工程与科学
COMPUTER ENGINEERING & SCIENCE

2011 年第 33 卷第 11 期
Vol. 33, No. 11, 2011

文章编号:1007-130X(2011)11-0089-09

单帧图像人体姿态估计综述*

A Survey of Human Pose Estimation from Single Images

汤泽胜,王兆仲
TANG Ze-sheng, WANG Zhao-zhong
(北京航空航天大学图像处理中心,北京 100191)
(Image Processing Center, Beihang University, Beijing 100191, China)

摘　要:人体姿态估计是指从图像中检测人体各部分的位置并计算其方向和尺度信息,姿态估计的结果分二维和三维两种情况,而估计的方法分基于模型和无模型两种途径。本文首先介绍了人体姿态估计的研究背景和应用方向,然后对姿态估计的相关概念作了阐述,分析了姿态估计的输出表示,接着从人体目标检测和姿态估计两大类进行了详细分析和讨论,从实际应用的角度对各种方法做了理论上的比较和分析。最后,对相关研究还存在的问题和进一步研究的趋势作了归纳和总结。

Abstract: Pose estimation from images is to detect the position where the human body is located in the image and computes the orientation and scale for each body part, the results of pose estimation is always classified as 2-D and 3-D, and the approaches to these results can be divided as model-based approaches and model-free approaches. In this paper we first give a simple introduction to the background and applications of pose estimation, then conceptions relevant to pose estimations are discussed, and the different representations of the results produced are compared, the remaining part of this paper focuses on the human body detection and human pose estimation, which are discussed in detail, all the different methods are discussed theoretically by applicability, at the end of this paper. The problems remaining to be located are analyzed.

关键词:人体姿态估计;目标检测;人体模型;单帧图像

Key words: human pose estimation; object detection; human body model; single image

doi: 10.3969/j.issn.1007-130X.2011.11.017

中图分类号:TP391.41　　**文献标识码**:A

1　引言

随着数码相机以及拍照手机等消费电子产品的普及,包含人体目标的数字图像广泛存在于家庭计算机和互联网中,如何对这些图像进行分类、索引和检索成为一个难题。针对单帧图像的人体姿态估计是基于内容的图像索引的重要方面,成为近年来的一个研究热点。与此相近的一个研究热点是针对序列图像的人体姿态估计,它是行为理解与行为监控[1~3]以及基于视觉的人机交互技术等的基础,同样有着广泛的应用前景。从单个摄像机获

* 收稿日期:2011-05-10;修订日期:2011-08-20
基金项目:国家自然科学基金资助项目(60803071);教育部博士点基金项目(200800061067)
通讯地址:100191 北京市海淀区学院路 37 号北京航空航天大学图像处理中心
Address: Image Processing Center, BeiHang University, 37 Xueyuan Rd, Haidian District, Beijing 100191, P. R. China

取的序列图像用于姿态估计[4]的优势在于可以充分利用人体运动的时空一致性特征，针对同一目标在光照近似不变条件下采用多种方法如背景差、光流等提取人体特征，然后估计人体姿态。更进一步，有研究者采用多个摄像机采集的图像来获取人体姿态[5~7]，解决了单目视觉丢失深度信息的问题，从而获取更加可靠的三维人体姿态。Shan[8]、Yap[9]、LI[10]和 WANG[11]等人的综述文章针对2007年以前从单个摄像机获取的序列图像用于姿态估计的研究现状作了归纳和总结。本文将主要针对近几年从单帧图像出发的人体姿态估计的相关研究成果进行综述。需要说明的是，有相当一部分的研究是针对含标记图像的，由于真实场景下的图像通常不包含标记，手动标记对于海量图像的索引也不切实际，所以这一部分文献不在本文涵盖范围之内，本文主要针对全身或者上半身姿态估计的文献进行综述，仅针对手臂或者头部姿态的估计也不在涵盖范围内。

2 问题描述与分析

人体的姿态分二维和三维两种情况：二维人体姿态是指人体各关节在图像二维平面分布的一种描述，通常用线段或者矩形来描述人体各关节在图像二维平面的投影，线段的长度和角度分布或者矩形的大小和方向就代表了人体二维姿态，二维姿态不存在二义性问题；三维人体姿态是指人体目标在真实三维空间中的位置和角度信息，通常用关节树模型来表述估计的姿态，也有一些研究者采用更加复杂的模型，三维姿态的获取通常是通过模型反投影的方法。

无论是二维姿态估计还是三维姿态估计，都可以分为基于模型的方法和无模型的方法。基于模型的方法通常是对人体目标建立关节树模型[4,12~16]，无模型的方法则通常是基于统计学习理论或者流形的方法。当前姿态估计研究的主流是基于模型的方法。姿态估计结果的可视化输出一般分为关节树模型[15~17]、2-D 模型[18]和立体模型[19,20]三种不同方式。由于立体模型的构造比较复杂，而关节树模型和 2-D 模型相对简单，易于操控，所以大多数研究者选择关节树模型和 2-D 模型作为姿态估计的输出结果表示，但关节树模型更为常见。

对于基于模型的一类方法，需要将人体模型与图像中人体各部分的特征如形状、色彩、轮廓等信息对应起来，从而计算模型的参数，达到估计人体姿态的目的；对于基于无模型的方法，也需要从图像中人体特征如边缘、色彩等学习与人体姿态之间的映射关系，从而达到估计人体姿态的目的。因此，无论是哪一种方法，对人体目标的检测都是必不可少的步骤。故在本文接下来的部分对人体目标检测和姿态估计的相关研究作一个回顾。

3 人体目标的检测

当前研究姿态估计的主流方法是基于模型的方法，这意味着人体目标检测是必不可少的预处理步骤。与序列图像不同，在没有标记的单帧图像中找到人体目标或者人体特征是一项极具挑战性的工作，因为单帧图像中不存在可以利用的光流信息，也没有可以利用的背景知识。而真实场景中的背景通常都是比较复杂的。有效地提取人体目标或者特征是进行姿态估计时必不可少的步骤。Seemann[21]对 2005 年以前的行人检测方法做了比较全面的分析和对比。下面对近几年的研究中一些比较典型的人体检测方法做一个概述。

3.1 基于分割与匹配的检测方法

图像分割是一种通用的目标检测方法，通过分割能够有效地定位目标在图像中的位置。对于已知摄像机固定的图像序列，通常的目标检测方法是背景差分割方法。对于其他图像，一般通过建立统计模型，从而将分割问题转化为能量最小化问题。

Rother[22]等提出的基于改进的 Graph-cuts 的 Grabcut 算法用于静止图像分割方法，将分割问题变为一个能量最小化问题。Grabcut 算法与 Graph-cuts 算法相比在两个方面做了改进：一是相比于 Graph-cuts 是执行一次运算估计混合高斯模型参数，Grabcut 方法采用迭代估计的方法，多次估计以提高精度；二是相比于 Graph-cuts 算法是完全标记(即把所有像素点进行一次分类，分为前景与背景)方法，Grabcut 算法采用不完全标记，即先采用硬分割的方法，将图像划分为前景、背景以及边界区域，边界区域采用 Matting 的方法，用 Grabcut 进行优化分割，最终到令人满意的结果。Sun[23]等人同样对图像建立马尔可夫随机场模型并将分割问题视为能量最小化问题，通过引入阴影分量达到了准确分割同时消除阴影的能力，他们同样采取了 Graph-cuts 方法来求解能量最小值。Han[24]等人在多尺度框架下，结合 Grabcut 和闭环带状区域局部线性估计的方法对图像进行分割，提

高了分割的速度和精度。

使用基于分割的方法进行目标检测最终得到的将是包含整个目标的区域。与分割方法不同，基于匹配的方法既可以实现局部匹配、定位，也可以实现全目标匹配定位。匹配通常有基于特征点的方法和基于轮廓的方法[25~27]。Wang[25]等人提出综合考虑全局的形状变形和局部几何特征，将形状匹配问题转化为匹配代价最小化问题，其代价函数由形状差异分量和形状计算误差两部分构成。Shotton[27]等则提出了基于部分轮廓的目标检测方法，首先从分割好的图像中学习相应的目标轮廓模型，然后使用 Boost 分类器查找图像中目标的轮廓，从而实现了目标的检测。

基于分割的方法能为基于外观模型(Appearance Model)的方法与基于轮廓或形状的方法提供可靠的信息，但由于在处理海量数据时无法使用确定的先验信息(如衣服的色彩、肤色等)，图像分割结果的不确定性在低对比度图像上尤为明显，这种不确定性带来的直接后果就是给基于轮廓或形状的方法带来误差。而基于轮廓匹配的方法大致分为两类：一是从分割的结果中提取目标的轮廓，分割的不稳定特性对轮廓方法造成不少影响；二是直接从边缘提取的方法中寻找轮廓，但在复杂场景下从目标与背景中提取的边缘难以区分。

3.2 基于梯度信息的检测方法

Dalal[28]等人提出的方向梯度直方图(Histograms of Oriented Gradients)是一类基于图像梯度信息的目标检测方法。其检测过程可以归纳为以下几个步骤：

(1)使用伽玛变换增强图像对比度，使用高斯平滑以减小噪声影响。

(2)计算图像梯度，分两种情况考虑：对于灰度图，直接计算图像梯度，对于 RGB 三通道图像，对每一个通道单独计算图像梯度，取最大值。

(3)划分方向网格，对网格中的每一个像素计算一个权值，用于计算当前网格的梯度方向，并对网格进行规范化，减小局部光照不均匀的影响。

(4)训练线性支持向量机，决定每一个网格的权重，用于最终决策哪些方向向量表征着人体目标的位置。

Freeman[29,30]使用梯度方向直方图检测方法做手势识别和游戏交互，曾春[31]等在文献[28]的基础上采用基于感兴趣区域的梯度方向直方图来检测人体目标，减少了计算量和时间。

基于梯度信息的方法是目前人体检测研究中使用较多的一类方法[15,16,32~34]，其优点是结果比较稳定。但是，通常使用此方法的研究都是先给定一个包含目标的区域或者称之为感兴趣区域(ROI)，然后再整体计算梯度方向直方图或者依据先验信息分片计算，以提高算法的速度。Ferrari[15]在其最新的算法中，加入了人脸识别的过程，从而免去了手动指定 ROI 的过程，使得算法更具有通用性。

3.3 基于统计学习的检测方法

Ronfard[35]通过标记图像中的人体各个部分，采用支持向量机和相关向量机学习人体各个部分的分类器，从而较为准确地定位人体各个部分在图像中的位置，达到人体检测目的。Mohan[36]基于支持向量机训练为人体的每一个部分训练一个分类器，通过检测部分躯体达到最终检测整体的目的，也取得了良好的效果。基于学习的方法通常有较好的效果，但为了获得好的结果，通常意味着花费大量时间和精力去构造一个训练样本库和大量的分类器训练时间。

4 人体姿态的估计

姿态估计的方法主要有基于模型(Model-Based)的和无模型(Model-Free)的两类。而无模型的又可以分为监督学习的方法和半监督学习的方法。基于模型的方法是目前研究的主流方法。下面以基于模型的方法为重点进行概述。

4.1 基于模型的姿态估计

Ferrari[15]等人主要考虑到现实生活中的许多场景中人都只有上半身可见，如被沙发遮挡、坐在游艇里、站在花丛中等等，提出了一个主要针对上半身正面二维姿态估计的解决方法，但也可以推广到全身姿态的估计。他们的方法可以归结为以下四步：

(1)使用 HOG 方法进行人体检测。要求手动指定一个矩形区域包含人的头部和肩部作为初始化信息，进而将图像划分为几个区域，如前景(包含部分背景)和背景区域，主要依据人体目标在图像中的空间分布先验信息，并对前景区域进一步划分，得到一个只包含前景的区域和其他区域，及手臂可能出现的区域。在最新的改进版本中，他们结合人脸检测和梯度向量直方图等，减少了初始制定头部区域的输入参数要求，提高了算法的自动化程

度与可用性。

(2)前景高亮,在前景区域搜索人体各关节部分,使用人体结构的先验信息,由于前一步处理已经明确了绝对前景区域和绝对背景区域,所以可以采用基于 Grabcut 的分割方法完成前一步中指定的前景区域中前景和背景的分割。

(3)单帧图像人体姿态解析,经过前一步的处理之后,在不考虑人体被截断的情况下,目标会被比较完整地分割出来。这一步使用分割后的图像中的边缘信息,其中隐含着内外轮廓信息,这样可以很好地获取人体姿态,无论人体姿态中是否存在自遮挡。同时,从分割后的目标区域中学习目标外观模型(Appearance Model)。

(4)考虑到同一视频序列中人的外观、衣着等不会发生明显变化,通过上一步学习到的外观模型,以及人体在图像序列中的空间连续性和人体运动的几何连续性,还可以使用最大熵的方法对视频中后续帧的人体姿态进行估计。

Marcin[16]等在文献[15]的基础上,通过增加更多的约束,如两腿穿的裤子是相似的,而手臂露出的部分皮肤和脸部的皮肤具有相似性,手臂穿着衣服的部分和上半身的衣服也应相似,两条手臂之间相似等相似性约束,从而简化了对初始化的要求和对人体各部分的检测过程,进一步提高了估计的精度。

Singh[17]等人对图像中的手臂和腿部进行姿态估计。他们的算法可以简单地用以下几个步骤表述:

(1)对给定的图像进行分割,提取手臂或者腿部;

(2)对分割后的图像提取骨架;

(3)将提取的骨架使用 Radon 变换映射到 Radon 空间;

(4)使用 SMM(Spatial Maxima Mapping,简称 SMM)算法将未知的姿态与已知的姿态进行匹配;

(5)将匹配的最佳结果作为姿态估计的最终结果。

Jiang[18]对分割得到的人体目标图像使用多个矩形块做一个一致最大的覆盖,每个人体部分对应一个矩形区域,从而将姿态估计问题转化为矩形区域覆盖与原分割图像的一致性问题,而从各矩形区域的分布恢复人体姿态是一个线性可解的问题。

Andriluka[37]在已有单帧图像二维姿态估计的研究基础上,以二维姿态估计为过渡估计序列图像中的三维人体姿态。他们的工作首先是在单帧图像的图像结构模型(Picture Structure Model)中估计人体关节树模型的位置、姿态和视角,然后在贝叶斯最大后验概率下利用模型反投影的方法从这些二维姿态信息中估计三维姿态。Daubney[38]采用了类似于文献[37]的方法基于图像结构模型恢复人体三维姿态,所不同的是,他们没有估计二维姿态作为过渡,直接从图像中检测人体特征并映射到三维空间,然后从映射到三维空间中的特征估计三维人体姿态。

通过对基于模型的方法分析发现,这类方法一方面有比较稳定的结果,但对图像分割结果和先验信息有较大的依赖,并且由于人体模型的限制,通常一个算法只针对一类行为导致的姿态所作的姿态估计能够令人满意。因此,基于模型的方法针对单帧图像或者序列图像能取得较好的效果,适合用于运动分析和行为理解的应用场合,但对于海量图像的分析索引,有着先天不足。

4.2 基于无模型的姿态估计

Wang 和 Qian[39]给出了一个由粗到精的姿态识别算法,用于在图像中查找与给定的参考姿态相匹配的姿态。他们的算法可以归为以下三步:

(1)给定一个参考姿态(带标记),依据图像中标记的分布,去除不匹配的姿态;

(2)使用基于快速傅立叶变换实现的循环卷积的快速直方图匹配算法计算图像中姿态与给定参考姿态之间的旋转角度;

(3)使用非线性最小二乘法修正前一步计算得到的角度。

Guo 和 Qian[19]提出的基于流形学习的方法用于三维人体跟踪,无需建立复杂的人体模型,使用支持向量机的方法构造人体关节角度的映射,然后使用动态高斯过程模型跟踪人体的三维运动。

Rogez[32]使用 HOG 方法提取图像中人体的位置信息,然后使用随机树(针对单个人体)和随机森林(针对多人体图像)进行识别和分类,从数据库中训练了 192 个不同的直立人体三维姿态类别,使用二维流形将这些类别及摄像机视角信息映射为一个类别空间,从而将姿态估计问题转化为类别空间的分类问题得以解决。

Chen[33]等人通过设计特征选取机制,从众多特征如傅立叶描述子、Shape Context、边缘、梯度等特征中找到全局最优特征,快速准确地完成特征

到三维姿态的反投影过程。

Kohli[40]从基于 Dynamic Graph Cuts 方法结合形状先验信息求解的条件随机场分割得到的目标图像出发，然后再次使用 Dynamic Graph Cuts 求解由条件随机场描述的三维人体姿态，从而在同一的框架下将分割和姿态估计结合起来。

Okada[34]等人选择梯度方向直方图作为恢复人体姿态的特征，通过训练多个局部线性回归量来恢复单帧图像中的三维人体姿态。通过选择最相关的特征结合支持向量机方法，把相近的姿态区分开来，经过训练后，每一个线性回归量都可找到一组独立的且和三维姿态相关的特征来恢复人体姿态。

基于无模型的方法意味着对庞大的训练样本集、更多的训练时间、更稳定的训练算法以及优化参数的要求，但无模型的方法从另一个角度来说，不会被模型所限制，更具备通用性；而且基于无模型的算法可以从已经索引的图像中继续学习，从而改善姿态估计的结果。因此，无模型的方法更适合海量图像数据的索引应用场合。

5 实验与分析

我们采用的实验数据来自公开的测评图像库 buffy stickmen[41]。该图像库中的图像均截自美剧 Buffy the Vampire Slayer，背景复杂、光照变化明显、不同图像中的人体目标尺度相差明显、衣着各不相同，存在各种遮挡情况。我们在此图像库中选出 100 幅图像作为测评样本。我们测评的对象来自公开源代码的算法实现，考虑算法设计的适用性，测试的算法包括多目标姿态估计的 Ferrari[15,42]算法，以及单目标姿态估计的 Yang Yi[43,44]算法、Ramanan[45,46]算法和 Andriluka[47,48]算法。除了 Ramanan 算法过于消耗内存而在一台 Windows 7 操作系统搭配四核 Intel i3 处理器和 8G 内存的工作站上完成测试外，其他三个算法均在微机上完成，平台参数为 Windows XP sp3 操作系统搭配 Intel 酷睿 2 双核处理器和 2G 内存。

对于实验，我们分析统计的对象是人体目标数以及相应的关节数。可见目标数是在考虑遮挡与分辨率的情况下，人眼可见的目标数目统计；可见关节数是对于可见目标的可见关节进行统计，同样考虑图像的对比度和目标之间的遮挡和自遮挡；当目标图像中的人没有出现在检测和姿态估计的结果中的时候，判定为漏检；当一个检测结果明显偏离图像中目标人的分布或者不符合人体结构的分布时，判定为误检；而准确率的计算则以图像中可见的关节数为分母，以姿态估计正确的关节数为分子。

为了估计图像中多个目标的姿态，通常的办法是记录检测到的目标的个数和位置，然后分别估计该目标的姿态。

图 1 为 Ferrari 算法测试的一个结果，结果表示方式为关节树模型。

a 原图

b 目标1估计结果

c 目标2估计结果

d 目标3估计结果

图 1 Ferrari 算法多目标姿态估计示例

单目标姿态估计算法不能处理多目标的原因在于其检测算法未考虑多目标情况，针对一幅图像，只能检测到一个目标。图 2 是 Andriluka 算法的 2 个测试结果，其输出结果的表示方式为 2D 模型。

5.1 多目标姿态估计算法

Ferrari 算法基于直立人体假设训练人体模型，然后基于人体模型匹配估计人体姿态，其人体模型考虑了人体关节之间的连接关系。本文采用半身模型进行实验，即对每个目标人体上半身六个关节进行检测。对于该算法的评测结果如表 1 所示。

a 原图

b 估计结果1

c 估计结果2

d 估计结果3

图 2　Andriluka 算法单目标姿态估计示例

表 1　Ferrari 算法多目标姿态估计统计结果

可见目标数	217
可见关节数	1 183
漏检目标数	47
误检目标数	54
检测关节数	704

由表中数据可知，该算法的准确率、漏检率和误检率分别为 59.5%、21.7%和 24.9%。

5.2　单目标姿态估计算法

由于单目标姿态估计算法一幅图像中只估计一个目标的姿态，因此其漏检率与错检率相同，而多目标姿态估计的漏检率和误检率则存在差异，这是由检测算法决定的。评测结果如表 2 所示。

表 2　单目标算法姿态估计统计结果

	Yang Yi	Ramanan	Andriluka
可见目标数	100	100	100
可见关节数	558	558	558
漏检/误检数	44	100	54
检测关节数	348	0	178

Yang Yi 算法设计用于检测目标图像中单个人体全身姿态，其算法准确率、漏检率/误检率分别为 62.4%、44%。Andriluka 算法的准确率为 31.9%，漏检率/误检率为 54%。Ramanan 算法使用条件随机场推断人体姿态，虽然在他的论文中给出了较好结果，但在此数据库上的准确率为 0。

5.3　实验结果分析

图 3 给出了上述四个算法中三个算法的执行时间趋势图，Ramanan 算法由于复杂度过高，未画出相应曲线。图 3 中，横坐标表示图像的大小比例，纵坐标代表相应比例下的图像执行姿态估计算法所需要的时间，单位为秒。需要说明的是：就源代码实现来说，Ferrari 算法、Yang Yi 算法采用 Matlab 实现，而 Andriluka 使用 C++实现。由图 3 可以看出，三个算法都是近似线性复杂度的，Ferrari 算法的耗时曲线出现了上下波动的情况是因为图像大小不同，其检测到的特征数出现了波动，它的姿态估计依赖于特征，故而他的耗时曲线与图像大小不直接相关。从图 3 还可以看出不同算法的时间复杂度对比关系。

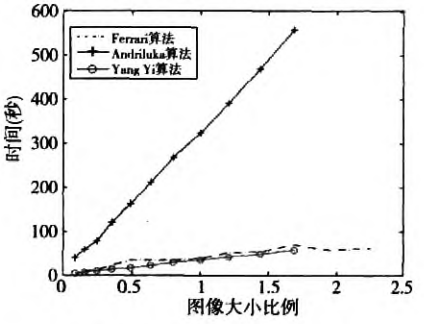

图 3　不同算法耗时曲线

通过对比实验结果我们发现，影响算法性能的因素主要有：(1)检测算法的性能，漏检降低算法的准确率，而误检则增加了不必要的计算时间。(2)模型的设计，Ferrari 算法中设计人体模型时考虑了关节之间的连接关系，增加了一个较弱的约束，使得其算法性能远远高于另外三个算法，而 Andriluka 和 Yang Yi 算法中人体模型的设计将各个关节视为彼此独立，这使得算法在图像中包含多个人体目标或者背景较为复杂的情况下，准确率受到极大影响；相反，Ramanan 算法中人体模型约束过强，一个关节的估计结果对另一个影响过大，导致完全没有正确的结果出现。

6　已有研究的不足和进一步研究的趋势

6.1　算法自动化程度

基于模型的姿态估计方法要求先对图像进行

分割[15,16,18]，而针对单帧图像的分割方法如 Grabcut 等需要初始给定参数，这个要求使得算法不适合用于批量处理图像，交互式分割方法则更是如此。Ferrari[15]利用人脸识别算法，在一定程度上提高了算法的自动化程度，但正确率还有待提高。对于人脸不可见或者半可见的情况，还没有合适的办法解决问题，这个问题的解决对于处理单幅图像中多个人体目标的姿态估计也是必不可少的。

6.2 效率

目前的姿态估计方法完成一幅原始图像到姿态参数输出的过程通常需要数秒到数百秒，为了最终达到应用于索引和检索互联网上海量图像数据的目的，提高姿态估计算法的效率则是一个迫切需要解决的问题。

6.3 输出结果的衡量标准

现有研究在结果的可视化输出表示上没有形成统一的标准，常见的就有关节树模型、2-D 模型和立体模型[49]，也造成了不同的研究者做的研究无法直接比较，因而姿态估计的相关研究论文很少有与其他人的研究成果作对比。对于真实图像而言，在采集的时候基本都是不知道姿态的参数的，因此算法性能的好坏多是主观判断的，没有一个客观的评价标准。在这一个方向要有所突破有赖于建立一个庞大的包含姿态参数的完备的数据库，这需要投入大量的人力物力。投机的办法是使用由姿态参数控制生成的近似真实人体模型投影生成的图像用于检验算法的准确性。

6.4 鲁棒性与通用性

人体姿态固有的复杂性导致的遮挡是姿态估计无法忽视的一个重要问题，在仅有一帧图像可以利用的情况下，为了获得完整的人体姿态，目标人体通常是正面或者侧面但旋转角度在一定范围内的图像，以保证人体的各个肢体关节可见，这使得有相当一部分包含人体目标的图像无法被处理。二义性问题的解决有赖于基于模型的方法，通过对模型的各个关节施加关节约束，从而保证解的唯一性。无论是二维姿态还是三维姿态，目前的研究都针对人做一些特定动作（如步行、跑步、打棒球）时的姿态，还没有算法能针对所有姿态作估计，如何提高算法的通用性是算法面向实际应用必须要解决的问题。

6.5 算法稳定性与准确率

从我们的对比实验中可以得出这样的结论，只有良好的检测算法与设计合理的人体模型彼此配合才是得到令人满意的估计结果的关键所在。从已有研究来看，单一检测方法的准确率比较有限，提高检测准确率的途径在于结合多种检测方法，而设计人体模型的关键在于如何利用先验信息，如关节约束、纹理等等。

7 结束语

针对单帧图像的人体姿态估计是基于内容的图像索引和检索的重要方向，也是人体运动分析和行为理解的相关研究的基础。本文针对从单帧图像中恢复人体姿态的相关研究作了一个基本回顾，按照人体姿态估计的一般过程，从检测和姿态估计两个方面对现有的研究作了分析和归纳，并从实际应用的角度分析了研究的难点和进一步研究的趋势，希望能对后来者有所裨益。

参考文献：

[1] Virone G, Sixsmith A. Activity Prediction for In-Home Activity Monitoring[C]// Proc of International Conference on Intelligent Environments, 2008:1-4.

[2] Vincze M, Zillich M, Ponweiser W, et al. Integrated Vision System for the Semantic Interpretation of Activities Where a Person Handles Objects[J]. Computer Vision and Image Understanding, 2009, 113(6):682-692.

[3] Jansen B, Temmermans F, Deklerck R. 3D Human Pose Recognition for Home Monitoring of Elderly[C]// Proc of the International Conference of Engineering in Medicine and Biology Society, 2007:4049-4051.

[4] Ramanan D, Forsyth D A, Zisserman A. Strike a Pose: Tracking People by Finding Stylized Pose[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2005:271-278.

[5] Huo Feifei, Hendriks E, Paclik P, et al. Markerless Human Motion Capture and Pose Recognition[C]// Proc of Workshop on Image Analysis for Multimedia Interactive Services, 2009:13-16.

[6] Peng Bo, Qian Gang, Ma Yunqian. Recognizing Body Poses Using Multilinear Analysis and Semi-Supervised Learning [J]. Pattern Recognition Letters, 2009, 30(14):1289-1294.

[7] Bray M, Kohli P, Torr P H S. Posecut: Simultaneous Segmentation and 3D Pose Estimation of Humans Using Dynamic Graph-Cuts[C]// Proc of European Conference on Computer Vision, 2006:642-655.

[8] Shan Gan-lin, Ji Bing, Zhou Yun-feng. A Review of 3d Poses Estimation from a Monocular Image Sequence[C]// Proc of International Congress on Image and Signal Processing, 2009:1-5.

[9] Hen Yap Wooi, Paramesran R. Single Camera 3d Human Pose Estimation: A Review of Current Techniques[C]// Proc

of International Conference for Technical Postgraduates, 2009:1-8.

[10] 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.

[11] 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.

[12] Felzenszwalb P F, Huttenlocher D P. Efficient Matching of Pictorial Structures[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2000:66-73.

[13] Ioffe S, Forsyth D A. Human Tracking with Mixtures of Trees[C]// Proc of International Conference on Computer Vision, 2001:690-695.

[14] Ramanan D, Forsyth D A. Finding and Tracking People from the Bottom up[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2003:467-474.

[15] Ferrari V, Marln-Jimnez M J, Zisserman A. Pose Search: Retrieving People Using Their Pose[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2009:1-8.

[16] Marcin E, Vittorio F. Better Appearance Models for Pictorial Structures[C]// Proc of British Machine Vision Conference, 2009.

[17] Singh M, Mandal M, Basu A. Pose Recognition Using the Radon Transform[C]// Proc of Midwest Symposium on Circuits and Systems, 2005:1091-1094.

[18] Jiang Hao. Human Pose Estimation Using Consistent Max-Covering[C]// Proc of International Conference on Computer Vision, 2009:1357-1364.

[19] Guo Feng, Qian Gang. 3D Human Motion Tracking Using Manifold Learning[C]// Proc of International Conference on Image Processing, 2007:357-360.

[20] Agarwal A, Triggs B. 3D Human Pose from Silhouettes by Relevance Vector Regression[C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004:882-888.

[21] Seemann E, Leibe B, Mikolajczyk K, et al. An Evaluation of Local Shape-Based Features for Pedestrian Detection[C]// Proc of British Machine Vision Conference, 2005.

[22] Rother C, Kolmogorov V, Blake A. "Grabcut": Interactive Foreground Extraction Using Iterated Graph Cuts[J]. ACM Trans on Graphics, 2004, 23(3):309-314.

[23] Sun Yunda, Yuan Baozong, Miao Zhenjiang, et al. Better Foreground Segmentation for Static Cameras via New Energy form and Dynamic Graph-Cut[C]// Proc of International Conference on Pattern Recognition, 2006:49-52.

[24] Han Shoudong, Tao Wenbing, Wu Xianglin, et al. Fast Image Segmentation Based on Multilevel Banded Closed-Form Method[J]. Pattern Recognition Letters, 2010, 31(3):216-225.

[25] Wang Song, Kubota T, Richardson T. Shape Correspondence Through Landmark Sliding[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2004:143-150.

[26] Belongie S, Malik J, Puzicha J. Shape Matching and Object Recognition Using Shape Contexts[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 24(24):509-522.

[27] Shotton J, Blake A, Cipolla R. Contour-Based Learning for Object Detection[C]// Proc of International Conference on Computer Vision, 2005:503-510.

[28] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// Proc of IEEE Conference on Computer Vision and Recognition, 2005:886-893.

[29] Freeman W T, Roth M. Orientation Histograms for Hand Gesture Recognition[C]// Proc of International Workshop on Automatic Face and Gesture Recognition, 1994:296-301.

[30] Freeman W T, Tanaka K, Ohta J, et al. Computer Vision for Computer Games[C]// Proc of IEEE International Conference on Automatic Face and Gesture Recognition, 1996:100.

[31] 曾春,李晓华,周激流.基于感兴趣区域梯度方向直方图的行人检测[J].计算机工程,2009,35(24):182-184.

[32] Rogez G, Rihan J, Ramalingam S, et al. Randomized Trees for Human Pose Detection[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2008:1-8.

[33] Chen Cheng, Yang Yi, Nie Feiping, et al. 3d Human Pose Recovery from Image by Efficient Visual Feature Selection [J]. Computer Vision and Image Understanding, 2011,115(3):290-299.

[34] Okada R, Soatto S. Relevant Feature Selection for Human Pose Estimation and Localization in Cluttered Images[C]// Proc of European Conference on Computer Vision, 2008:434-445.

[35] Ronfard R, Schmid C, Triggs B. Learning to Parse Pictures of People[C]// Proc of European Conference on Computer Vision, 2002:700-714.

[36] Mohan A, Papageorgiou C, Poggio T. Example-Based Object Detection in Images by Components[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(4):349-361.

[37] Andriluka M, Roth S, Schiele B. Monocular 3D Pose Estimation and Tracking by Detection[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2010:623-630.

[38] Daubney B, Gibson D, Campbell N. Monocular 3D Human Pose Estimation Using Sparse Motion Features[C]// Proc of International Conference on Computer Vision, 2009:1050-1057.

[39] Wang Yi, Qian Gang. Robust Human Pose Recognition Using Unlabeled Markers[C]// Proc of Workshop on Applications of Computer Vision, 2008:1-7.

[40] Kohli P, Rihan J, Bray M, et al. Simultaneous Segmentation and Pose Estimation of Humans using Dynamic Graph Cuts[J]. International Journal of Computer Vision, 2008, 79(3):285-298.

[41] http://www.robots.ox.ac.uk/~vgg/data/stickmen/.

[42] http://www.vision.ee.ethz.ch/~calvin/articulated_human_pose_estimation_code/.

[43] Yang Yi, Ramanan D. Articulated Pose Estimation with Flexible Mixtures-of-Parts[C]// Proc of the 24th IEEE Conf on Computer Vision and Pattern Recognition, 2011:1385-1392.

[44] http://phoenix.ics.uci.edu/software/pose/.

[45] Ramanan D. Learning to Parse Images of Articulated Objects[J]. Neural Info Proc Systems, 2006.

[46] http://www.ics.uci.edu/~dramanan/papers/parse/index.html.

[47] Andriluka M, Roth S, Schiele B. Pictorial Structures Revisited: People Detection and Articulated Pose Estimation[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[48] http://www.mis.tu-darmstadt.de/node/381.

[49] 杜友田,陈峰,徐文立,等. 基于视觉的人的运动识别综述[J]. 电子学报,2007, 35(1):84-90.

[50] Montabone S, Soto A. Human Detection Using a Mobile Platform and Novel Features Derived from a Visual Saliency Mechanism[J]. Image and Vision Computing, 2010, 28(3):391-402.

[51] Sigal L, Bhatia S, Roth S, et al. Tracking Loose-Limbed People[C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition, 2004:421-428.

汤泽胜(1987-),男,江西九江人,硕士生,研究方向为目标检测与姿态估计。**E-mail**:tangzesheng@sa.buaa.edu.cn

TANG Ze-sheng, born in 1987, MS candidate, his research interests include object detection, and pose estimation.

王兆仲(1975-),男,辽宁大连人,博士,讲师,研究方向为计算机视觉、数字图像、视频处理、图像科学中的动力系统理论和非线性方法。**E-mail**:zwang@buaa.edu.cn

WANG Zhao-zhong, born in 1975, PhD, lecturer, his research interests include computer vision, digital image, video processing, and dynamic system theory in image technology and nonlinear method.