

基于约束树形图结构外观模型的人体姿态估计

王浩 刘则芬 方宝富 陈金金

(合肥工业大学计算机与信息学院 合肥 230009)

摘要 人体姿态估计是计算机视觉领域中的一个关键问题,可用于行人检测、人体活动分析、人机交互以及视频监控等方面。针对基于树形图结构模型的人体姿态估计算法中人体部件外观模型容易受到背景干扰的问题,提出了基于先验分割和外观转换的外观模型的人体姿态估计算法以改进人体的外观模型。根据PS模型,使用人体检测器和前景加亮进行预处理,确定人体的大致位置和大小,同时移除背景上的杂斑,基于先验分割和外观转换机制来估计人体部件的外观模型。实验表明在不同的图像数据库中,该算法在使用人体检测器和前景加亮算法减少部件搜索空间的同时,提高了人体姿态估计的准确性。

关键词 图像处理,人体姿态估计,外观模型,图形结构,先验分割

中图分类号 TP181 文献标识码 A

Human Pose Estimation Based on Appearance Model for Constraint Tree Pictorial Structure

WANG Hao LIU Ze-fen FANG Bao-fu CHEN Jin-jin

(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract Human pose estimation is one of the hot topic in the field of computer vision, and can be used for pedestrian detection, human activity analysis, human-computer interaction and video surveillance. Aiming at the problem that the human body appearance model of human parts is vulnerable to background interference in the body pose estimation algorithm of the tree pictorial structure model, we put forward body pose estimation algorithm of appearance model which is based on priori segmentation and the appearance transfer mechanism in order to improve the human body appearance model. According to PS model, using the human body detector and foreground highlighting by preprocessing, it can find an approximate position and size of the human body, meanwhile remove the background clutter. We can estimate the appearance model of the human body parts based on a priori segmentation and appearance transformation mechanism. Experiments show that using the algorithm of the human body detector and foreground highlighting can not only reduce the search space of the components, but also improve the accuracy of the body pose estimation.

Keywords Image processing, Human pose estimation, Appearance model, Pictorial Structure, Piori segmentation

1 引言

人体姿态估计是计算机视觉领域的一个重要而富有挑战性的研究课题,可应用于行人检测、视频监控、图像和视频检索、行为分析等领域。准确地估计人体在复杂场景中的姿态,在图像和视频检索等领域具有很重要的意义。近年来,图形结构(Pictorial Structure, PS)^[1]成为越来越多学者研究的热点,这种方法借助图模型,将人体的各个部件看作图中的节点,各个部件间的连接关系表示图形的边,然后利用图模型中的推理方法估计人体的姿态。但是在实际生活中,大多数的图像受到背景、光照、外观、人的穿着、遮挡等因素的影响,导致单独依靠图形结构模型的姿态估计方法的检测率低,检测速度慢。

为了解决这个问题, Felzenszwalb^[1]针对已知背景的图

片,提出了基于前、背景差分的外观模型,但是,这种模型在复杂、动态的背景中的检测准确性受到很大的限制,因此不具有推广性。Andriluka^[2]等人使用形状上下文和 adaboost 算法构造了判别式的外观分类器,它能够有效地检测人体各个身体部件,但是该分类器由于只使用了形状信息作为判断的依据,仍然会产生误检,这种模型不适用于人体部件有遮挡的场景中。Ramanan^[3]等人提出了基于边缘的描述算子,它根据边缘对背景和着装的敏感性,在边缘模板的基础上,进一步学习人体区域模板,对于每一幅测试图像,先用边缘变形模板初始化人体的姿态,然后使用得到的前景区域和背景区域的颜色直方图,构造区域变形模板,最后通过使用姿态估计和前/背景直方图之间的反复迭代得到最终的姿态。这种方法易受到背景的影响,在背景比较复杂的情况下,计算量大。以上3种方法都只是直接对图像做处理,没有对图像进行预处理,背

到稿日期:2013-05-15 返修日期:2013-07-22 本文受国家自然科学基金(61070131, 61175051, 61175033, 61075076), 国家高技术研究发展计划(863计划)(2012AA011005), 安徽省自然科学基金(1308085QF108)资助。

王浩(1962—),男,博士,教授,CCF高级会员,主要研究领域为人工智能与机器人、数据挖掘、面向对象技术等, E-mail: jsjxwangh@hfut.edu.cn; 刘则芬(1989—),女,硕士生,主要研究领域为人工智能机器应用; 方宝富(1978—),男,博士,副教授,主要研究领域为机器视觉、机器学习; 陈金金 男,硕士生,主要研究领域为人工智能机器应用。

景会对最终的结果造成一定的干扰。

Ferrari^[4]等人对检测的图像进行了预处理,提出了一个通用的人体姿态的检测器。使用 HOG 方法进行人体检测,需要手动指定一个矩形区域(包含人体头部和肩部)作为初始化信息,进而将图形划分为几个区域,如前景(包含背景)区域和背景区域,然后使用 Grabcut 算法对前景区域进行进一步划分,得到一个只包含前景区域和其他区域以及手臂可能出现的区域,从而减少姿态搜索的空间。但是该算法开始时对头部区域的输入参数进行了限制,减少了算法的自动化程度。Ferrari^[5]等人在最新的改进版本中,结合了人脸检测和梯度向量直方图等,减少了初始时对头部区域输入参数的限制,在一定程度上提高了算法的自动化程度,但是对于人脸不可见或半可见的情况,还是没有合适的办法来解决这个问题。

本文针对基于树形图模型的人体姿态估计算法中人体部件外观模型容易受到背景干扰的问题,提出了基于先验分割和外观转换的外观模型的人体姿态估计算法以改进人体的外观模型。本文是在 Ferrari^[4]的方法的基础上,借助 PS 模型,使用人体检测器和前景加亮算法进行预处理,前者获得的检测窗口能够确定人在图像中的大致位置,后者用 Grabcut 算法根据检测窗口提供的信息进行初始化,移除背景上人体周围的杂斑,获得前景区域内人体各个部件的位置信息。然后通过学习得到各个部件关于检测窗口的先验分割,再使用外观转换机制来估计各部件的外观模型,最后使用 PS 模型来估计人体的姿态。图 1 所示是本文姿态估计的全过程。

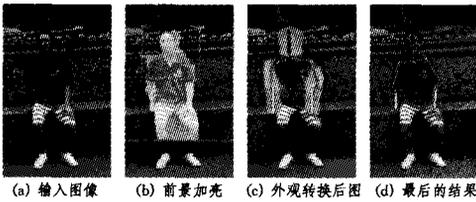


图 1 人体姿态估计的全过程

2 姿态估计的预处理

2.1 基于约束的 PS 模型

PS 模型将人体分成若干个部件,然后对各个部件以及它们之间的关系进行建模。这种方法将人体描述成一个无向图,每个节点表示一个身体部件,无向图的边表示运动学上相连的两个部件。节点通常用矩形来表示: $l=(x,y,\theta,s)$,其中, (x,y) 表示部件的位置, θ 表示部件的方向, s 表示部件的尺度。则人体的姿态可定义成 $L=(l_1,l_2,\dots,l_n)$,人体的上半身包括:头部、躯干、左上臂、右上臂、左下臂、右下臂。本文使用的是基于树形结构的 PS 模型,如图 2(a)所示。

设 D 表示人体结构模型参数集, I 表示图像信息, L 表示人体各部件位置信息。假设人体的各个部件之间是相互独立的,根据 Bayes 理论,估计某一幅图像 I 中人体的姿态 L ,其后验概率可以表示为:

$$P(L/I,D) \propto \exp\left(\sum_{(i,j) \in E} \psi(l_i,l_j) + \sum_i \phi(l_i,D)\right) \quad (1)$$

式中, $P(L/I,D)$ 表示当模型为 D 、图像 I 的情况下,人体的姿态是 L 的后验概率,二元约束项 $\psi(l_i,l_j)$ 表示运动学上相连的两个部件 i 和 j 的位置的先验概率, $\phi(I/l_i,D)$ 表示在特定的位置 l_i 的图像特征和外观模型上的部件 i 的似然程度。

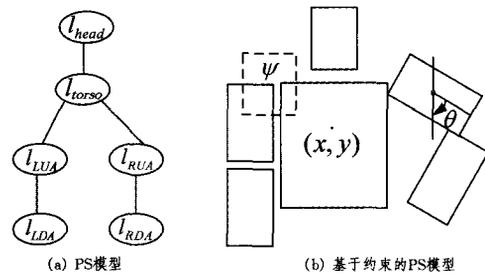


图 2 上半身 PS 模型和基于约束的 PS 模型

在图像处理的过程中,通常遇到的图像中的人只有上半身可见,在这种情况下,假设人体的头部位于躯干的上方。本文使用了一种约束的 PS 模型,如图 2(b)所示,它通过给式(1)增加约束条件来实现,并增加 $\gamma(l_{head})$ 和 $\gamma(l_{torso})$ 来限制躯干和头部的方向是竖直的或接近竖直的。

$$p(L/I) \propto \exp\left(\sum_{(i,j) \in E} \psi(l_i,l_j) + \sum_i \phi(l_i) + \gamma(l_{head}) + \gamma(l_{torso})\right) \quad (2)$$

式中, $\gamma(\cdot)$ 表示在竖直方向附近的 θ 值概率是均匀的,在其他方向上的概率为零,这样可以减少头部和躯干的搜索空间,提高正确估计人体姿态的可能性。确定人体的躯干位置后,根据运动学上的先验概率 ψ 来限制手臂的动作,有利于上、下手臂的姿态估计。

2.2 人体检测器

通用检测器^[7]可以用来检测各种各样的物体,在以前的工作中^[4,5],使用通用检测器来获得人体在检测窗口中的大致位置。本文使用通用检测器训练许多标记人体姿态的图像来获得人体检测器,从而确定图像中人体上半身所在的检测窗口,根据先验知识:人的头部一般在检测窗口的中上方,而人体躯干位于头部的下方,头部在身体中的比例是比较稳定的,可以确定最终的检测窗口的大小。本文根据检测出的窗口,选取头部、肩部所在的位置,找出肩部所在的检测窗口和棍状图(stickmen)的矩形间的合理的比例,使扩大后的检测窗口能够覆盖所有的人体区域,包含手臂可能出现的区域。根据检测窗口来确定图像中的人体所在的大致位置,以达到在复杂场景下正确估计人体姿态的目的。

2.3 前景加亮

根据人体检测器获得人体在检测窗口中大致的位置,在很大程度上减少了部件的搜索空间。在使用的过程中,需要选择一个合适的检测窗口,使图像中的人在任何姿态的情况下,都能够限制在这个检测窗口中。通过调整能够得到一个合适的检测窗口,把搜索区间定位在检测窗口区域内,此时检测窗口所覆盖的区域面积相对于人体部件还是比较大,检测的结果仍然会受到背景的干扰。本文将得到的检测窗口作为 Grabcut 算法的初始输入,根据 Grabcut 分割出前/背景区域,进一步确定矩形框中人体各部件所在的区域。

3 基于改进的外观模型的姿态估计

3.1 基本思想

通过人体检测器和前景加亮算法的预处理,在很大程度上减少了部件的搜索空间。经过前面两步的预处理,使用 Grabcut 分割算法之后,人体能够被比较完整地分割出来,根据分割后的目标区域能够得到人体的外观模型。因此,本文提出了一种基于先验分割和外观转换的姿态估计算法。通过训练大量的样本图像,可以得到这样的先验知识:一些身体部件相对于检测窗口 $W=(x,y,s)$ 的位置是稳定的;同时,不同

的身体部件间的外观是相互关联的。先验知识是建立在统计学基础上的,通过训练大量的标记真实信息的棍棒图样本,可以学习先验分割(segmentation prior)获得每个部件相对于 W 的位置分布,根据给出的 W 和先验分割信息来估计初始的外观模型,这种模型得到的是图像的粗略信息,通过外观转换机制进行优化,最后得到人体部件的外观模型。如图 3(a)所示,通过学习得到各个部件的先验分割:躯干、左右上臂、左右下臂、头部,把先验分割值作为外观模型的初始值,如图 3(b)所示,根据运动学上的联系进行外观转换,得到转换后的颜色模型,如图 4(c)所示,最后得到各个部件分割后的颜色模型,如图 4(d)所示,根据分割后的颜色模型,使用 PS 模型进行推断,最后能够估计出人体的姿态。

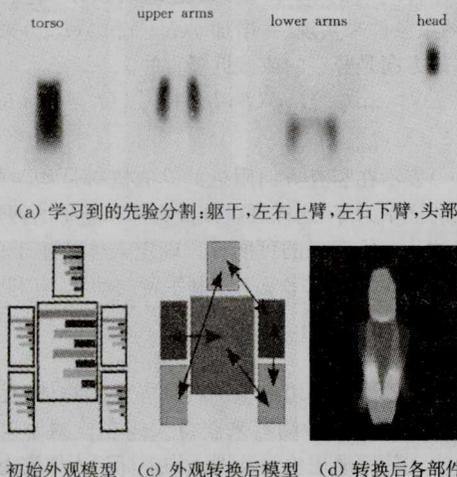


图 3 根据先验分割和外观转换机制得到人体外观模型的过程

3.2 先验分割

先验分割是在不考虑实际数据的情况下,部件 i 包含像素 (x, y) 的先验概率,因此对于每个身体部件 i ,其先验分割 $SP_i(x, y) \in [0, 1]$ 。通过训练标注真实信息的棍棒图,可以学习每个部件相对于检测窗口 W 的先验分割。学习的过程:首先,根据运行在图像上的检测器来获得检测窗口,然后,把棍棒图结合到这个检测窗口上,根据检测窗口可以得到所有训练棍棒图的一个通用的坐标系,在通用坐标系中能够确定人体各个部件大致的位置。在学习先验分割时,检测器得到的结果可能会有误差,此时,通过极大似然估计来学习先验分割。每个部件 i 的先验分割 $SP_i(x, y)$ 是训练图像中人体部件 i 包含像素 (x, y) 的小数部分。

3.3 不同部件之间的外观转换

在下手臂有遮挡的情况下,可以根据人体衣着或人体皮肤来估计人体的外观模型,由于不同身体部件的外观间有着固然的联系,可以根据季节或场景等进行判断。因此,需要一个结合不同身体部件之间外观模型的转换机制来估计人体的外观模型。外观转换机制可以根据一个部件的外观模型来估计另一个部件的外观模型,本文是根据文献[8],使用混合权重的方式来实现外观转换的,在已知部件 i 的外观模型 AM_i^{TM} 的情况下,求部件 t 的外观模型 AM_t^{TM} 的公式如下:

$$AM_t^{TM} = \sum_i w_{it} AM_i^{SP} \quad (3)$$

其中, w_{it} 表示部件 i 结合部件 t 的混合权重, AM_i^{SP} 是根据先验分割得到的初始的外观模型。混合权重是转换机制的参数,混合权重是通过转换机制产生的外观模型 AM_i^{TM} 和从训练棍棒图得到的真实值 AM_i^{GT} 差的平方的最小值获得的。公式如下:

$$\min_{w_{it}} \sum_s \sum_k (\sum_i w_{it} AM_i^{SP} - AM_i^{GT})^2 \quad (4)$$

$$\text{s. t. } 0 \leq w_{it} \leq 1, \sum_i w_{it} = 1$$

其中, i 取值为身体部件的个数, s 表示所有训练实例的个数, k 表示所有的外观模型的组件。通过求解这些混合权重可以得到各个部件之间的关系,进一步减少姿态搜索的空间,从而提高姿态估计的准确性。

根据学习到的先验概率和外观转换机制,估计人体的外观模型时,使用先验分割来估计初始的外观模型,根据文献[9]使用 RGB 颜色空间的颜色直方图模型,将每个颜色带分成 16 份(bins),所以总共为 $16 * 16 * 16$ 份。定义每份的颜色为 c ,则每个部件 i 的前景概率为 $P_i(c/fg)$,背景概率为 $P_i(c/bg)$,根据前景和背景概率能够确定像素 (x, y) 属于某个身体部件 i 的先验概率,从而计算每个部件位置的后验概率 $P(i/D)$ 。把检测窗口 W 转换到学习先验分割得到的标准坐标系中,通过修剪检查窗口 W 外的图像,使其有一个固定的大小。把先验分割的值作为颜色模型的初始值,其中先验分割的值表示颜色直方图的高度。根据外观转换式(3)来重新调整颜色模型,能够得到每个部件最终的颜色模型 $P_i(c/fg)$ 。根据文献[3]中的方法、每个部件的先验分割的补集,即每个身体部件的背景模型 $P_i(c/bg)$,假设 $P_i(bg) = P_i(fg)$,可以计算出像素 (x, y) 属于某个身体部件 i 的先验概率 $P_i(fg/c)$ 即每个部件分割后的颜色模型,从而计算式(1)的 $\phi(I/l_i, D)$ 。根据 Bayes 公式可得:

$$P_i(fg/c) = \frac{P_i(c/fg)}{P_i(c/fg) + P_i(c/bg)} \quad (5)$$

3.4 基于先验分割和外观转换机制的人体姿态估计算法步骤

(1)输入:当前固定大小的测试图像(设测试图像的大小为 $m \times n$, m 表示宽度, n 表示高度)。

(2)减少搜索空间

①根据人体检测器获得人体在图像中的大致位置;

②在①的基础上扩大检测窗口使其包含整个人体,根据假设(人的头部一般在检测窗口的中上方,而人体躯干位于头部的下方),可以有效地确定哪些是背景部分,哪些是前景部分,从而有效地使用 Grabcut 算法进行初始化。

(3)学习先验分割

把棍棒图结合到(2)②的检测器上,可以得到一个通用坐标系,得到各个部件的先验分割值。

(4)使用外观转换机制估计部件的外观模型

①把先验分割值作为外观模型的初始值;

②估计每个部件的颜色模型,具体过程可分为 3 步:

a)把检测窗口 W 转换到通用坐标系中,修剪检测窗口 W 外的图像,使其有一个固定的大小;

b)把各个部件先验分割值作为颜色模型的初始值;

c)根据外观转换式(3)来重新调整颜色模型,得到每个部件最终的颜色模型 $P_i(c/fg)$;

③估计颜色分割,根据式(5)来获得每个部件分割后的颜色模型 $P_i(c/fg)$,即确定像素 (x, y) 所在的部件 i ,从而计算式(1)的 $\phi(I/l_i, D)$ 。

(5)输出:根据 PS 模型推断出人体的姿态。

4 实验

4.1 实验数据

实验中采用了 3 个不同的数据库,第一个数据库是 Buffy

数据库,取自美剧《Buffy the Vampire Slayer》第5集的第2集到第6集,共748帧图像,其中人体上半身姿势经过手工标定。第2个数据库取自 ETHZ PASCAL 中的549幅图片,同样上半身的姿势也经过手工标定。第3个数据库是 PARSE 数据库,包含305张图像,前100张图像用来训练,其余的图像用来测试。

4.2 实验结果和分析

给定一个检测图片,能够得到姿态估计的分布 $P(L|I)$, 这些分布中用式(6)把分数最高的作为最后的结果。

$$-\frac{1}{T} \sum_i \log P(L_i^A | I_i) \quad (6)$$

式中, L_i^A 是标记的真实姿态, T 表示需要测试的检测图片。

本文在实验的过程中,当测试图片选自《Buffy the Vampire Slayer》的第2集、第5集、第6集时,用其第3集、第4集和 ETHZ PASCAL 数据库中的图片训练。当测试 ETHZ PASCAL 数据库中的图片时,利用 Buffy 数据库中的图片做训练。当测试 PARSE 数据库时,PARSE 数据库前100张训练图片学习观测模型,然后将其余图片作为估计图片。

如图4所示,(a)是本文的实验结果,(b)是 Ferrari 的方法^[5]、Wang 的方法^[10]、费的方法^[10]的实验结果。图5是通过训练 Buffy 数据库和 ETHZ PASCAL 数据库,然后对 PARSE 数据库上的图像进行姿态估计的结果。通过观察可以发现,由于人的手臂经常是运动的,具有很大的灵活性,可能和背景融成一片,因此在姿态估计的过程中容易造成误检。而本文的方法增加了基于部件的外转换机制,减少了部件的搜索空间,提高了姿态检测的准确性。



(a) 本文的实验结果



(b) Ferrari 的实验结果



(c) 费的实验结果



(d) Wang 的实验结果

图4 在 Buffy 和 ETHZ PASCAL 数据库上本文方法和 Ferrari 的方法、Wang 的方法、费的方法比较的结果



图5 PARSE 数据库使用本文方法的测试结果

表1列出本文算法和 Ferrari 的方法、Wang 的方法、费的方法在3个数据库中的实验结果。由表可知本文算法在3个

数据库中进行测试时,姿态估计的正确率有一定的提高,由于 PARSE 中的图像估计的都是人体全身的姿态,因此在 PARSE 数据库中的正确率略低于其他两个数据库,但本文的方法比 Ferrari 的方法、Wang 的方法、费的方法具有优势。

表1 Buffy、PARSE、ETHZ PASCAL 3个数据库中估计正确率

估计正确率	Buffy (748张图片)	ETHZ PASCAL (549张图片)	PARSE (305张图片)
本文的方法	78.5%	65.7%	22.3%
Ferrari 的方法	70.8%	57.1%	17.2%
Wang 的方法	68.3%	53.4%	15.6%
费的方法	48.5%	47.9%	13.4%

结束语 本文使用了基于外观模型的人体姿态估计算法,其在上半身是直立的图像中取得很好的效果,对复杂背景下人体姿态估计有了很大的提高。本文通过使用人体检测器和前景加亮进行预处理,获得人体在检测窗口中的大致位置,根据学习得到的先验分割作为外观模型的初始值来进行外观转换,得到人体的外观模型。最后根据 PS 模型进行推理,估计出人体的姿态。本文基于约束的树形图结构的外观模型人体估计算法,在估计全身的人体姿态时,全身各个部件之间的外观转换需要花费大量的时间,下一步工作,将对人体全身人体部件外观之间的转换过程进行细化,完善自己的算法。

参考文献

- [1] Felzenszwalb P F, Huttenlocher D P. Pictorial structures for object recognition[J]. International Journal of Computer Vision, 2005,66(1):55-79
- [2] Andriluka M, Schiele S R B. Pictorial structures revisited: people detection and articulated pose estimation[C]// Irfan Essa. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009, Florida: IEEE Computer Society, 2009: 1014-1021
- [3] Ramanan D. Learning to parse images of articulated bodies [C]// Neural Information Processing Systems, 2007. Cambridge: MIT Press, 2007: 1129-1136
- [4] Ferrari V, Marin-Jimenez M, Zisserman A. Progressive search space reduction for human pose estimation[C]// Computer Vision and Pattern Recognition, 2008. Anchorage: IEEE Computer Society, 2008: 1-8
- [5] Ferrari V, Marin-Jimenez M, Zisserman A. Pose search: Retrieving People Using Their Pose[C]// Computer Vision and Pattern Recognition, 2009, Florida: IEEE Conference Publication, 2009: 1-8
- [6] 贾慧星,章毓晋. 基于梯度方向直方图特征的多核跟踪[J]. 自动化学报, 2009, 35(10): 1283-1289
- [7] Eichner M, Marin-Jimenez M. 2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images [J]. Int J Comput Vis, 2012, 99: 190-214
- [8] Felzenszwalb P, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models[J]. IEEE Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645
- [9] Wang Hao, Meng Fan-hui, Fang Bao-fu. Iterative Human Pose Estimation Based on A New Part Appearance Model[J]. Applied Mathematics & Information Sciences, 2014, 8(1L): 311-317
- [10] 费英兰,范宁宁. 基于人体图模型和背景减除的姿势估计[J]. 苏州大学学报, 2012, 32(3): 8-14