

基于矩形拟合的人体检测*

孙庆杰¹⁺, 吴恩华^{1,2}

¹(中国科学院 软件研究所 计算机科学重点实验室,北京 100080)

²(澳门大学 科学技术学院,澳门)

Human Detection Based on Rectangle Fitting

SUN Qing-Jie¹⁺, WU En-Hua^{1,2}

¹(Key Laboratory of Computer Science, Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

²(Faculty of Science and Technology, University of Macau, Macau, China)

+ Corresponding author: Phn: 86-10-62562796, Fax: 86-10-62563894, E-mail: sqj@ios.ac.cn

<http://www.iscas.ac.cn>

Received 2002-09-11; Accepted 2002-11-05

Sun QJ, Wu EH. Human detection based on rectangle fitting. *Journal of Software*, 2003,14(8):1388~1393.

<http://www.jos.org.cn/1000-9825/14/1388.htm>

Abstract: A new algorithm for human detection in a static image is proposed in this paper. Human is seen as a model constructed by a torso and four limbs. The torso is fitted by a quadrangle, and each limb is fitted by one or two quadrangles. Detecting human is to search a combination of quadrangles constrained by some geometrical and topological relations. The first step is to detect and fit rectangles in the image, and then try to search some reasonable combinations of rectangles that should satisfy the geometrical and topological relations. If the probability of the combination is more than a fixed threshold, the vertex positions of the rectangles are adjusted to get a compact model of human in that image. At last, the probability of the compact model is recalculated. Experimental results show that this algorithm can be applied in the content-based image retrieval and the analysis of human motion based on video.

Key words: human detection; rectangle fitting; probability calculation; image segmentation; human model

摘要: 提出了一种在单幅图像中检测人体的新算法.将人体看作由躯干和四肢构成.用一个四边形拟合躯干,用一个或者两个四边形拟合每个肢体.人体检测就是寻找满足一定几何和拓扑约束的四边形组合.首先检测与拟合图像中的矩形区域,然后搜索满足几何和拓扑约束的矩形组合.如果该矩形组合的概率大于阈值,则调整各个矩形顶点的坐标,以便得到一个紧凑的人体模型表示.最后,重新计算该紧凑模型的概率.实验结果表明,该方法可以用于基于内容的图像检索以及基于视频的人体运动分析.

关键词: 人体检测;矩形拟合;概率计算;图像分割;人体模型

中图法分类号: TP391 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant Nos.69873044, 60033010 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312100 (国家重点基础研究发展规划(973))

第一作者简介: 孙庆杰(1972-),男,河北河间人,博士生,主要研究领域为计算机图形图像,模式识别.

人体检测是近几年的一个研究热点.它应用在许多方面,如监控系统、驾驶员辅助系统、图像理解、图像索引、基于图像的人体建模和姿态仿真等.在实现人体检测系统时要考虑许多复杂的因素^[1].首先,人体是关节物体,可以有很多种形状和姿态,如人的胳膊和腿的相对位置可以有很大的变换范围.另外,人可以穿各种颜色、各种风格的衣服.

大多数人体检测系统都利用了运动信息或简化的背景,它们所实现的是在图像序列上进行人体的追踪^[2,3],而不是静态图像中的人体检测.本文的研究目标是单幅彩色图像中的人体检测.

人体的形态有很多变化,但人体的每个组成部分(本文只考虑躯干和肢体,每个肢体有上下两部分)变化相对比较.我们首先从图像里检测可能是人体各个部分的矩形,然后搜索图像中是否存在一个或者多个由这些矩形按照一定的几何和拓扑关系构成的组合,如果存在并且其构成人体的概率大于一个阈值,则认为检测到一个人体.

本文第1节介绍一些相关工作.第2节介绍一种图像分割方法.第3节介绍图像区域的分析和矩形拟合.第4节介绍人体检测以及概率计算.第5节给出实验结果.第6节对全文进行总结.

1 相关工作

在文献[4]中,M.Oren 利用小波模板检测静态图像中的直立人体.作者通过大量样本的学习,发现直立人体图像经过 Harr 小波变换后的系数有特定的规律,可以用一个模板来表示.他们对图像中每个特定大小的窗口以及该窗口进行一定范围的比例缩放得到的窗口进行 Harr 小波变换,然后利用支持矢量机检测变换的结果是否可以与小波模板匹配,如果匹配则检测到一个人体.该方法只适合直立人体,且四肢要尽量为向下状态.文献[1]对文献[4]做了改进.他们也利用了小波模板和支持矢量机,但并不是在每个窗口内对整个人体的小波模板进行匹配,而是在窗口的一些特定位置范围内对人体的每个组成部分进行相应的小波模板的匹配,之后对这些分量的匹配结果进行总的匹配评价.利用该方法检测直立人体时,允许四肢的相对位置可以有许多变化.

在文献[5~8]中,David A.Forsyth 等人提出一些方法用以判断静态图像中是否包含裸体.人体的皮肤在色度和饱和度方面具有一些特定的属性,将图像中具有这些属性的足够大的像素区域分割出来.然后利用局部对称性和边缘检测算子获得该图像区域中的许多矩形.最后利用采样概率判断哪些矩形可以构成一个人体配置.

文献[9]使用 LapLacian 算子检测图像中的边缘,然后利用边缘图的曲率信息检测人体.该方法可以用于灰度图像和彩色图像,但只能识别人的头部和肩膀.

对于静态图像中的人体检测,目前还没有成熟方法^[4].已有的方法需要施加一些限制(如直立、裸体、上半身等).本文假设人体躯干和四肢都没有被背景遮挡,且各个部分之间没有相互遮挡.我们采用基于分量的人体检测方法.首先,利用颜色和纹理信息将图像分割为一些区域,然后利用形态学方法对每个区域进行矩形的检测和拟合.如果两个矩形近似平行且两对对应端点临近,则将这两个矩形合并.在所有矩形中搜索满足一些几何约束和拓扑约束的矩形组合.如果某个矩形组合的概率大于预先设定的阈值,则检测到一个人体模型.

2 图像分割

为了检测人体,我们首先需要检测躯干和四肢对应的图像区域.一般而言,躯干和四肢分别对应近似一致的颜色纹理区域.为此,我们采用基于区域的图像分割技术.人体穿着的衣服可以有各种颜色,同时可能有纹理,因此图像分割时应该综合考虑颜色和纹理.但同时考虑像素间颜色和纹理的相似性会导致分割情况很复杂.我们利用 JSEG 算法^[10].将图像分割分为两步:颜色量化和空间分割.颜色量化是在颜色空间中进行的,不需要考虑颜色的空间分布.之后,每个像素的颜色值用其对应的颜色类标识所代替,这样就构成了图像的类型图.可以将该类图看成是一种特殊的纹理合成.第2步就是在该类图上进行空间分割,此时不需要考虑像素的颜色相似性.

3 图像区域的分析与矩形拟合

对图像分割得到的每个区域,我们利用形态学方法对其外边界包围的区域进行区域分析,以便得到一些拟

合矩形.如果一个区域的边界包含较长的图像最外边沿,则可以认为该区域是背景区域,不对其进行分析.

3.1 距离图

本文使用数学形态学的方法进行区域的分析.分析结果要能反映物体重要的几何特征以及相邻部分之间的拓扑关系.具体的形态学知识可见文献[11].

假设 I 是一个二值图像, R 是 I 的一个分割区域.我们为 R 定义一个距离图 D .对于 I 中的像素 p ,如果 p 不属于 R , $D(p)$ 为 0, 否则 $D(p)$ 定义为 p 到 R 的边界的最小欧式距离.如果 $D(p)$ 大于 0, 且不小于相邻像素的距离值, 则称其为局部最大值点.这些点可以反映区域的形状信息.

3.2 轴形段

对于复杂的形状,局部最大值点可能分成多个连通分量.每个连通分量的局部最大值点集可能由多个分支组成.我们从这组点集的最左下方的点开始追踪,每次遇到分支点时中止当前的追踪,得到的点集称为一个分支轴形点集.对其余的点重复该追踪过程,以得到所有的分支轴形点集.用直线段拟合每个分支轴形点集:首先确定该点集的两个端点,将它们作为线段的端点,判断其他点到该直线的距离是否大于某个阈值.如果大于该阈值,则在距离最大的像素处将该线段分割为两个线段.该过程递归进行,直到将所有点都组织到合适的线段为止,我们称这样的线段为轴形段.如果相邻的两个轴形段位于一条直线上,则将其合并,以避免在分支点处多余的分割.对轴形段上的每个像素,检测其对应的区域长度是否发生不连续变化.若是,则在该线段处将其分割为两个轴形段.这样得到的每个轴形段都对应着一个几何区域.

3.3 矩形拟合

假设正在分析的区域用前景色 f 填充,背景用背景色 b 填充.对每个轴形段,在线段的两个端点用颜色 Z 画垂直于轴形段的直线,称这些直线为潜在切割线.

对每个轴形段,我们需要判断其对应的区域是否可以用一个矩形近似.对轴形段上的每个像素,沿着垂直于线段的方向发出一条直线,计算该直线与区域边界的两个交点,称这两个交点之间的距离为该像素对应的区域长度.将第 i 个像素的区域长度记作 $lineLen_i$.

$$rectLength = \frac{1}{n} \sum_i lineLen_i,$$

$$difLen = \frac{1}{n} \sum_i |rectLength - lineLen_i|,$$

$$difRatio = \frac{difLen}{rectLength}.$$

如果 $difRatio$ 小于某个阈值 $maxLenLimit$, 则认为轴形段对应的区域可以用一个矩形近似, 否则该区域的形状不适合用矩形拟合. 如果轴形段对应的区域可以用矩形拟合, 则在轴形段两个端点处沿着另一端点的反方向扩展该线段及其对应的区域. 当扩展到某个像素时, 在该像素沿垂直方向发出一条直线, 该直线在轴形段两侧与区域边界以及所有潜在切割线的交点中寻找距离与 $rectLength$ 最接近的点对, 其距离值为该像素的区域长度. 如果该像素的区域长度发生了不连续变化, 则停止扩展. 如果 $difRatio$ 大于阈值 $maxLenLimit$, 则回退直至 $difRatio$ 小于 $maxLenLimit$ 的位置. 如果拟合后的矩形越出图像边沿的边长大于某个阈值, 则不考虑该矩形.

3.4 矩形合并

上述矩形拟合是对图像分割的每个区域进行的. 对于由不同区域拟合得到的两个相邻矩形, 如果它们的两组边分别近似平行, 有两组对应边分别近似共线, 一组对应顶点彼此相邻很近, 则将其合并为一个新的矩形.

4 人体检测

人体可以看作由一个躯干和四个肢体构成的. 当肢体伸直时, 可以用一个矩形拟合, 弯曲时上下半肢体分别由一个矩形拟合. 肢体和躯干的形状要满足一定的几何约束, 躯干和四个肢体之间也要满足一定的位置约束.

4.1 人体概率

在观察图像中的人体时,最重要的因素是人体的总体结构.人体图像受观察方向和姿态影响,形状可能发生一些改变,但这种改变并不影响人眼的观察结果.所以,我们需要设计一种强调人体总体结构的概率函数.

假设矩形 T 是一个可能的躯干矩形.其 4 个顶点附近的可能的肢体分别为 A, B, C, D .如图 1 所示,与 A 和 B 临近的顶点构成 T 的宽度边,与 B 和 C 临近的顶点构成 T 的长度边.这个组合构成人体的概率用 $P(\text{human})$ 表示.

$$p(\text{human}) = p(T = \text{torso}) * p(lb|T = \text{torso}) * p(lbFit),$$

$$p(lb|T = \text{torso}) = \text{least}\{p(x = lb|T = \text{torso})\}, (x = A, B, C, D),$$

$$p(lbFit) = (\text{fitDegree}(A, B) + \text{fitDegree}(C, D)) / 2.$$

$p(T = \text{torso})$ 表示矩形 T 构成躯干的概率. $P(A = lb|T = \text{torso})$ 表示矩形 T 和矩形 A 构成躯干-肢体关系的概率. $p(lb|T = \text{torso})$ 表示 A, B, C, D 这 4 个矩形与 T 构成躯干-肢体关系的最小概率. $\text{fitDegree}(A, B)$ 表示 A 和 B 两个矩形的相似度量对总概率的影响,可将其影响限制在一定范围内. $p(lbFit)$ 表示两对肢体的平均相似概率对总概率的调节. $p(T = \text{torso})$ 和 $p(x = lb|T = \text{torso})$ 是从大量样本中学习得到的概率函数.由于不同的人可能有不同的体形,同一个人可以有许多的姿态,而且人体可能穿着各种各样的衣服,所以四肢的形状可以有较大的变化,即表示肢体矩形的长宽比例会有较大的变化范围,因此我们不区分胳膊和腿对应的矩形,对它们使用相同的概率分布函数 $p(x = lb|T = \text{torso})$.

4.2 搜索矩形组合

人体检测就是寻找满足特定几何和拓扑约束的矩形组合.如果某个矩形的概率大于阈值,则其对应于一个人体模型.

对于复杂的图像,我们可能得到大量矩形.如果用同样的方法检测每个矩形,将会非常耗时.为简化搜索,将其分为两步:首先搜索可能的躯干矩形,然后在每个躯干矩形附近搜索可能的肢体矩形.

对于矩形 Y ,如果 $p(Y = \text{torso})$ 大于某个阈值,则称其为一个可能的躯干矩形.在 Y 的每个顶点附近搜索可能的肢体矩形.对于矩形 L ,如果其在矩形 Y 的某个端点附近,且 $p(L = lb|Y = \text{torso})$ 大于特定阈值,称其为 Y 的一个肢体.如果概率小于阈值,需要进一步检查 L 是否可以和其附近的某个矩形合并.如果可以合并,则判断其并集是否可以构成 Y 的某个肢体.如果可以为 Y 矩形的所有顶点搜索到肢体,则 Y 和其肢体矩形的组合可能对应于一个人体模型.如果 Y 的某个顶点对应于多个肢体矩形,选择使 $p(\text{human})$ 概率最大的相应肢体.

4.3 调整矩形顶点

为躯干矩形的每个顶点找到肢体矩形后,计算其对应于人体模型的概率 $p(\text{human})$.如果概率大于特定阈值,则检测到人体.由于图像分割和矩形拟合的原因,这些矩形可能彼此并不连接.我们需要调整相邻矩形的顶点,以便得到一个连通的人体模型:如果某个肢体处于非直立状态,则调整其上下肢体矩形的相邻顶点;对于每个肢体,根据其邻接的躯干矩形的顶点对其进行调整.这样可以得到一个紧凑的人体模型,之后更新该模型的概率 $p(\text{human})$.

5 实验结果

图 2 是几个人体检测的例子.第 1 列是输入图像.第 2 列是利用 JSEG 方法对图像分割得到的区域.第 3 列是对分割区域拟合后得到的矩形,其中包括一些相邻矩形合并后得到的新矩形.合并之后将长度和宽度都小于某个阈值的小矩形删除,因为我们假定身体的各部分都不会太小.第 4 列是检测到的对应于人体模型的矩形组合,其对应于人体模型的概率分别为 0.91, 0.84, 0.93.

实验中发现检测失败的一个主要原因是图像分割时,没有将人体包含的所有区域都分割出来.JSEG 不能很好地将皮肤颜色与相邻的灰度颜色分离.另外,使用形态学方法进行矩形拟合容易受噪声的影响.在未来的工作

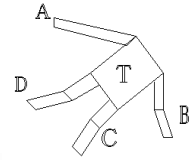


Fig.1 Human model

图 1 人体模型

中需要对这两个问题进行更深入地研究.

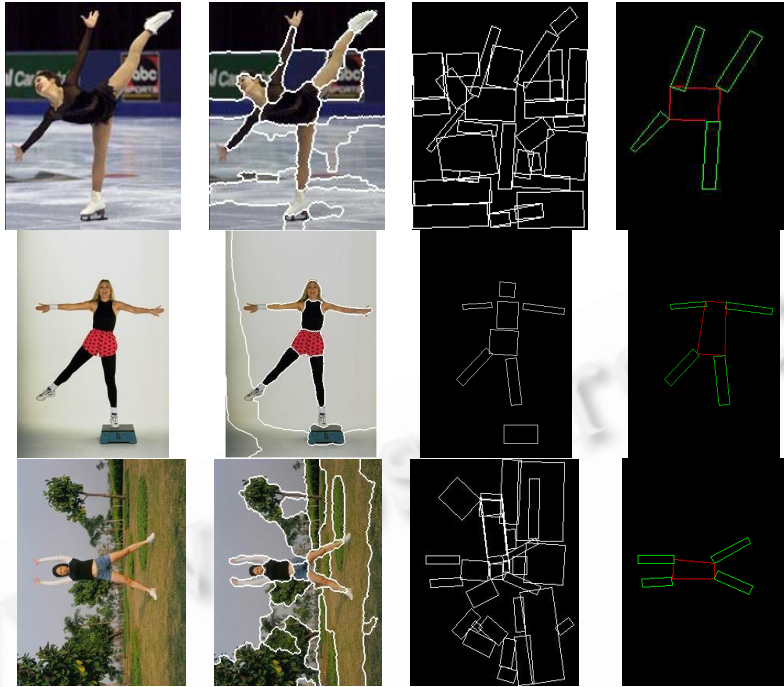


Fig.2 Examples for human detection

图2 人体检测的例子

6 结 论

本文给出了在静态图像中检测人体的一种方法.该方法首先利用 JSEG 算法将图像分割为许多区域,然后对每个区域进行形态学的分析,将其分解为一些几何区域,用矩形拟合这些几何区域.在得到的所有矩形中搜索满足特定拓扑和几何约束的矩形组合,如果某个矩形组合的概率大于预先设定的阈值,该组合就对应于一个人体模型.

本文将人体看成躯干和四肢的适当组合,要求人体的躯干和四肢都完整地出现.但我们搜索到的大多数人体图像都存在遮挡或者身体各部分之间自遮挡的现象.因此需要在未来的工作中着重研究一些新的匹配原则,以便处理人体某些部分被遮挡以及身体各部分之间自遮挡的情况.本文的人体模型只包含躯干和四肢,未来的工作可以使用更加完整的人体模型,如包含人的头部、手和脚.另外,图像分割是一个经典的难题,短时间内还不会出现完美的分割算法,所以在未来的工作中需要研究不依赖于图像分割的人体检测算法.

静态图像中的人体检测是一个非常困难的问题,因为人体有各种各样的姿态,并且人体所穿着的衣服可以有各种颜色和纹理.本文只是对该问题做了一个简单的尝试,今后还需要对其进行更加深入的研究.

References:

- [1] Mohan A, Papageorgiou C, Poggio T. Example-Based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001,23(4):349~361.
- [2] Haritaoglu I, Harwood D, Davis LS. W⁴: who? when? where? what? A real time system for detecting and tracking people. In: Mase K, ed. *Proceedings of the International Conference on Automatic Face and Gesture Recognition*. Nara: IEEE Press, 1998. 222~227.
- [3] Wren CR, Azarbayejani A, Darrell T, Pentland AP. Pfinder: real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997,19(7):780~785.
- [4] Oren M, Papageorgiou C, Sinha P, Osuna E, Poggio T. Pedestrian detection using wavelet templates. In: Nevatia R, ed. *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*. Puerto Rico: IEEE Press, 1997. 193~199.

- [5] Forsyth DA, Fleck MM. Body plans. In: Nevatia R, ed. Proceedings of the IEEE Conference Computer Vision and Pattern Recognition. Puterto Rico: IEEE Press, 1997. 678~683.
- [6] Fleck MM, Forsyth DA, Bregler C. Finding naked people. In: Buxton BF, Cipolla R, eds. Proceedings of the European Conference of Computer Vision. Berlin: Springer-Verlag, 1996. 593~602.
- [7] Forsyth DA, Fleck MM. Automatic detection of human nudes. International Journal of Computer Vision, 1999,32(1):63~77.
- [8] Ioffe S, Forsyth DA. Probabilistic methods for finding people. International Journal of Computer Vision, 2001,43(1):45~68.
- [9] Hyeon DH, Kim TE, Yoon KH, Choi JS. People detection in a static image using curvature. In: Tuceryan M, ed. Proceedings of the International Conference on Augmented, Virtual Environments and Three Dimensional Imaging. Ornos: IEEE Press, 2001. 363~366.
- [10] Deng YN, Manjunath BS. Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001,23(8):800~810.
- [11] Simmons M. 2D object abstraction: the automatic generation of hierarchical geometric representations [MS Thesis]. Berkeley: University of California, 1997.

第 8 届中国密码学学术会议

征文通知

第 8 届中国密码学学术会议拟定于 2004 年 5 月中下旬在上海（具体时间待定）举行。热忱欢迎所有涉及密码学（数学的和非数学的）、信息安全理论和关键技术方面的研究论文提交本次会议交流。

一、征文要求

提交论文必须是未公开发表并且未向学术刊物和其他学术会议投稿的最新研究成果,文稿可用中文书写,同时鼓励用英文书写,字数一般不超过 6000。会议论文集将以“密码学进展——ChinaCrypt'2004”由著名学术出版社出版发行(第 2 届~第 7 届密码学学术会议论文集由《科学出版社》、《电子工业出版社》出版发行),会议还将推荐优秀英文论文在 EI 检索源学术刊物上发表。作者应将论文全文(务必注明作者的详细通讯地址、联系电话和 E-Mail 地址)的 Word/PDF 文档,或论文全文的打印稿一式三份寄至以下地址:

二、重要日期

征文截止日期: 2003 年 07 月 31 日

文章录用通知: 2003 年 10 月 31 日

录用论文定稿: 2003 年 11 月 20 日

三、联系信息

贵州大学计算机系 李祥 教授

贵州省贵阳市

邮政编码: 550025

Tel: 0851-3621767

E-mail: lixiang@gzu.edu.cn

上海交通大学计算机系 陈克非 教授

上海市华山路 1954 号

邮政编码: 200030

Tel: 021-62932135

E-mail: kfchen@mail.sjtu.edu.cn

欲进一步了解会议的有关信息,欢迎访问有关站点 <http://www.chinacrypt.net>, <http://www.cs.sjtu.edu.cn>