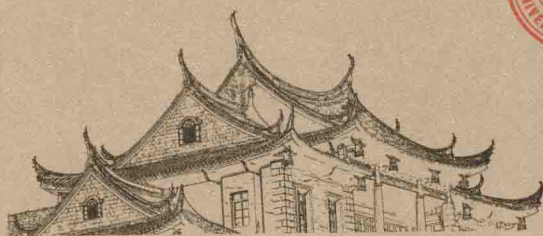


厦门大学南强丛书【第六辑】



Xiamendaxue
Nanqiang Congshu

行人检测：理论与实践

苏松志 李绍滋 蔡国榕◎著



厦门大学出版社
XIAMEN UNIVERSITY PRESS

国家一级出版社
全国百佳图书出版单位

厦门大学南强丛书

【第六辑】

国家自然科学基金项目(60873179 ,61202143,61572409)

高等学校博士学科点专项科研项目(20090121110032)

深圳市科技研发基金项目(ZYB200907110169A,JC200903180630A)

行人检测：理论与实践

苏松志 李绍滋 蔡国榕◎著



厦门大学出版社
XIAMEN UNIVERSITY PRESS

国家一级出版社
全国百佳图书出版单位

图书在版编目(CIP)数据

行人检测:理论与实践/苏松志,李绍滋,蔡国榕著. —厦门:厦门大学出版社,2016.3
(厦门大学南强丛书. 第6辑)

ISBN 978-7-5615-5951-2

I. ①行… II. ①苏… ②李… ③蔡… III. ①计算机视觉-研究 IV. ①TP302.7

中国版本图书馆 CIP 数据核字(2016)第 052074 号

出 版 人 蒋东明

责任编辑 陈进才

装帧设计 李夏凌

责任印制 许克华

出版发行 厦门大学出版社

社 址 厦门市软件园二期望海路 39 号

邮政编码 361008

总 编 办 0592-2182177 0592-2181253(传真)

营销中心 0592-2184458 0592-2181365

网 址 <http://www.xmupress.com>

邮 箱 xmupress@126.com

印 刷 厦门集大印刷厂印刷

开本 720mm×1000mm 1/16

印张 15.75

插页 4

字数 260 千字

版次 2016 年 3 月第 1 版

印次 2016 年 3 月第 1 次印刷

定价 55.00 元

本书如有印装质量问题请直接寄承印厂调换



厦门大学出版社
微信二维码



厦门大学出版社
微博二维码

“厦门大学南强丛书”（第六辑）编委会

主任委员：朱崇实

副主任委员：李建发 韩家淮

委 员：（以姓氏笔画为序）

万惠霖 朱崇实 孙世刚 李建发 宋文艳

陈支平 陈武元 陈振明 周 宁 周涵韬

洪永淼 蒋东明 郭永华 周永华 周永华

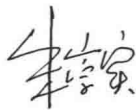
作者简介 **苏松志**，男，厦门大学博士，副教授，硕士生导师。2011年留校任教至今。主要研究方向为计算机视觉、机器学习及其应用，具体包括人脸识别与行人检测、无人驾驶中的智能视觉技术、三维场景的感知与理解和航拍图片分析。主持过国家自然科学基金和福建省自然科学基金项目，并作为核心技术骨干开发多个行人检测的相关项目，具有丰富的实际项目经验。已发表SCI论文12篇。

李绍滋，男，博士，教授，博士生导师。现任厦门大学智能科学与技术系主任，福建省人工智能学会副理事长，中国计算机学会协同计算专委会副主任。主要研究方向为人工智能及其应用、计算机视觉、机器学习、运动目标检测与识别、智能电网及电力综合自动化系统、自然语言处理与多媒体信息检索、网络多媒体及CSCW技术、智能中医信息处理等。先后主持或参加过多项国家863项目、国家自然科学基金项目、教育部博士点基金项目、省科技重点项目、省基金项目、深圳科技计划项目等各级项目的研究，获福建省科学技术三等奖两项。已发表论文200多篇，其中30多篇SCI检索，175篇EI检索，Google Scholar他引达到430多次，单篇他引达37次。

蔡国榕，男，厦门大学博士，副教授，硕士生导师。主要研究方向为计算机视觉、三维重建，具体包括图像三维重建、目标检测与识别、图像配准、深度学习理论及其应用等。近年来主持省部级自然科学基金，参与多项国家自然科学基金、国家支撑计划项目的研究。已发表论文30余篇，其中SCI收录8篇，EI收录17篇。

总 序

厦 门 大 学 校 长
“厦门大学南强丛书”编委会主任



厦门大学是由著名爱国华侨领袖陈嘉庚先生于1921年创办的,有着厚重的文化底蕴和光荣的传统,是中国近代教育史上第一所由华侨出资创办的高等学府。陈嘉庚先生所处的年代,是中国社会最贫穷、最落后、饱受外侮和欺凌的年代。陈嘉庚先生非常想改变这种状况,他明确提出:中国要变化,关键要提高国人素质,要提高国人素质,关键是要办好教育。基于教育救国的理念,陈嘉庚先生毅然个人倾资创办厦门大学,并明确提出要把厦大建成“南方之强”。陈嘉庚先生以此作为厦大的奋斗目标,蕴涵着他对厦门大学的殷切期望,代表着一代又一代厦门大学师生的志向。

1991年,在厦门大学建校70周年之际,厦门大学出版社出版了首辑“厦门大学南强丛书”,共15部优秀的学术专著,影响极佳,广受赞誉,为70周年校庆献上了一份厚礼。此后,逢五逢十校庆,“厦门大学南强丛书”又相继出版数辑,使得“厦门大学南强丛书”成为厦大的一个学术品牌。值此建校95周年之际,我们再次遴选一批优秀著作出版,这正是全校师生的愿望。入选这批“厦门大学南强丛书”的著作多为本校优势学科、特色学科的前沿研究成果。作者中有院士、资深教授,有全国重点学科的学术带头人,有新近在学界崭露头角的新秀,他们都在各自的学术领域中受到瞩目。这批学术著作的出版,为厦门大学95周年校庆增添了浓郁的学术风采。

至此,“厦门大学南强丛书”已出版了六辑。可以说,每一辑都从一个侧面反映了厦大学人奋斗的足迹和努力的成果,丛书的每一部著作都是厦大发展与进步的一个见证,都是厦大人探索未知、追

求真理、为民谋利、为国争光精神的一种体现。我想这样的一种精神一定会一辑又一辑地传承下去。

大学出版社对大学的教学科研可以起到很重要的推动作用,可以促进它所在大学的整体学术水平的提升。在95年前,厦门大学就把“研究高深学术,养成专门人才,阐扬世界文化”作为自己的三大任务。厦门大学出版社作为厦门大学的有机组成部分,它的目标与大学的发展目标是相一致的。学校一直把出版社作为教学科研的一个重要的支撑条件,在努力提高它的学术出版水平和影响力的过程中,真正使出版社成为厦门大学的一个窗口。“厦门大学南强丛书”的出版汇聚了著作者及厦门大学出版社全体同仁的心血与汗水,为实现厦门大学“两个百年”的奋斗目标做出了一份特有的贡献,我要借此机会表示我由衷的感谢。我不仅期望“厦门大学南强丛书”在国内学术界产生反响,而且更希望其影响被及海外,在世界各地都能看到它的身影。这是我,也是全校师生的共同心愿。

2016年3月

前 言

无人驾驶汽车是今年市场上的一大热点。在无人驾驶汽车平台中,智能视觉模块占据一定的比重,而行人检测则是其中基础而又关键的一个模块。这是一本关于行人检测的理论和实践的专著,主要总结近年来行人检测技术的发展状况。

本书第一章介绍了行人检测问题的来源和难点、公共的数据集和性能评价指标。第二章将行人检测分为四个模块:候选区域的确定、特征提取、分类与定位、验证,并分别综述了每个模块的研究现状。第三章详细介绍了基于滑动窗口的行人检测方法的基本思想,并对各种窗口融合方法进行详细介绍;在本章中,还对目前常用的行人特征描述子的计算方法、积分图和积分直方图技术等开发行人检测系统过程中常用的技术进行详细的介绍。第四章分别梳理了四种常见的基于部位的检测方法(基于组件的行人检测方法、基于图案结构模型的行人检测方法、基于形变部位模型的行人检测方法和基于姿态描述子的行人检测方法)。

局部特征描述子在图片检索等其他领域中也有广泛的应用,本书第五章在介绍完描述子之后重点介绍了基于局部块的行人检测方法。视频监控场景中,背景相对固定,笔者在第六章介绍了固定背景下的行人检测方法。运动信息和场景特定的几何信息能在一定程度上提高行人检测的性能,因此本书第七章和第八章对这方面的知识进行详细的介绍。

行人检测是一类特殊的目标检测问题,是计算机视觉中的经典难题,很难找到一个检测器能在任何场景下鲁棒地检测出行人来。实际上,本人在一开始从事行人检测的研究中,一直试图寻找这样一个“万能”的检测。随着研究的深入,才明白行人检测技术只能“具体问题具体分析”,因此领域的相关先验信息就特别重要(比如监控中的固定背景,车辆辅助驾驶中的路面信息等),这也是行人检测方法如此之多的原因。

本书在写作的过程中,实验室的学生提供了很多帮助。他们是黄敏、

杨柳、刘志辉、和文丽、雷明仪、袁德东、敖翔和董洁等，在此表示感谢。感谢家人对我的理解和支持，尤其是患病中的妻子给予的支持。此次出版，恰逢厦门大学 95 周年校庆，感谢母校对我的培养，谨以此书献给她。

行人检测技术随着市场需求的推动和机器学习理论的不断进步，发展相当快，很难对所有的方法进行深入的剖析。笔者只是将自己从事行人检测具体项目开发的过程中所掌握的方法和一些经验整理出来，和大家一起分享。才疏学浅，略懂而已，兼具时间和精力有限，书中错谬之处在所难免，恳请指出，不胜感激。

苏松志

2016 年 3 月

目 录

第一章 引 言	001
1.1 行人检测问题的来源与难点	001
1.2 公共数据集	006
1.3 性能评价标准	012
1.4 本书的组织结构	016
第二章 行人检测技术的研究现状	021
2.1 确定行人候选区域	022
2.2 基于 Stixel 的候选区域确定	024
2.3 行人特征提取	030
2.4 分类与定位	035
2.5 验 证	042
2.6 特定场景下的行人检测	043
第三章 基于滑动窗口法的行人检测方法	052
3.1 滑动窗口法	052
3.2 行人特征描述子	059
3.3 分类器	078
3.4 快速检测技术	094
3.5 基于 GPU 的快速行人检测方法	100
第四章 基于部位法的行人检测方法	113
4.1 基于组件的行人检测方法	113
4.2 基于图案结构模型的行人检测方法	116
4.3 基于形变部位模型的行人检测方法	129

4.4	基于姿态子的行人检测方法	140
第五章	基于局部块的行人检测方法	147
5.1	检测算子	147
5.2	描述算子	150
5.3	词袋模型	161
5.4	基于隐式形状模型的行人检测方法	165
5.5	高效子窗口搜索法	167
第六章	固定背景下的行人检测	178
6.1	减背景法	178
6.2	后处理	194
6.3	运动区域分类方法	200
第七章	基于运动信息的行人检测	209
7.1	光 流	209
7.2	基于 Haar 运动信息的行人检测	214
7.3	基于光流直方图信息的行人检测方法	218
第八章	特定场景下的行人检测方法	225
8.1	基于自学习的半监督训练方法	225
8.2	基于自学习的无监督训练方法	227
8.3	基于迁移学习的训练方法	229
8.4	基于非真实样本的训练方法	235
8.5	算法优劣分析	237
第九章	总结和展望	240

第一章 引言

行人检测是当前图像处理和计算机视觉中的研究热点,可看作一类特殊的目标检测问题。目标检测包括广义的目标检测和狭义的目标检测。广义的目标检测指的是类别层次的目标检测,比如花卉、车辆、人脸和行人等的检测。狭义的目标检测指的是类别中某个具体个体的检测,比如检测某个具体品牌的汽车或者寻找某个特定人的人脸,这本质上是目标识别。目标检测与识别是计算机视觉中的经典难题,大部分研究者研究的目标包括人脸、车辆、行人等。其中,人脸和车辆属于刚性物体,目标的结构稳定,研究起步较早,相应的检测技术也较为成熟,且有相关的产品问世,比如人脸考勤和车牌识别等。行人兼具刚性物体和柔性物体的特性,与人脸和车辆相比,表观和姿态的变化更加多样化,检测难度更大。此外,行人检测是视频监控、车辆辅助驾驶和人体行为分析等应用领域中基础而又关键的第一步。因此,行人检测成为继人脸检测之后,计算机视觉目标检测领域中的又一研究热点。本书主要总结近 15 年来行人检测技术的发展。本章中,我们首先描述行人检测问题的来源和难点,然后总结一下目前的公共数据集和性能评价标准,最后介绍本书的组织结构。

1.1 行人检测问题的来源与难点

人体运动分析的相关研究最早可追溯到 1982 年 Hoffman 等的工作^[1]。这一时期该领域中的大部分研究工作都假设运动的目标已经被检测出来了,研究的重点放在运动目标的识别、分类和长序列运动模式的分析上。在 2000 年前后的视频监控系统如 Pfinder^[2]、W⁴^[3]和 VSAM^[4]等项目中,行人检测也只是被视为系统中的一个预处理步骤,并未引起研究者的足够重视。

近年来随着智能视频监控、车辆辅助驾驶、基于内容的图像或者视频检索和人体行为分析等领域的广泛应用,以及一些新的应用领域的出现,如家庭服务机器人、基于航拍图像的行人及受害者检测^[5,6]、Google 街景图中的行人消除^[7]等,场景中的目标(包括车辆、行人、道路、交通标志和树木等)多样化。在这复杂的场景中分析人体的行为,不能像早期的研究那样经过简单的预处理或者手动标注出人体的位置,而是迫切需要程序自动地给出人体在图像或者视频中的位置。因此行人检测技术的研究逐渐受到研究人员的重视。

与人脸检测的定义^[8]类似,行人检测可定义为^[9]:

行人检测 判断输入图片(或者视频中的某一帧)中是否包含行人,如果有,给出位置信息。

如图 1-1 所示,对于给定的输入图片,这些图片可能是单张的静态图像,也可能是来自视频中的某一帧,行人检测要求最终给出行人在图片中

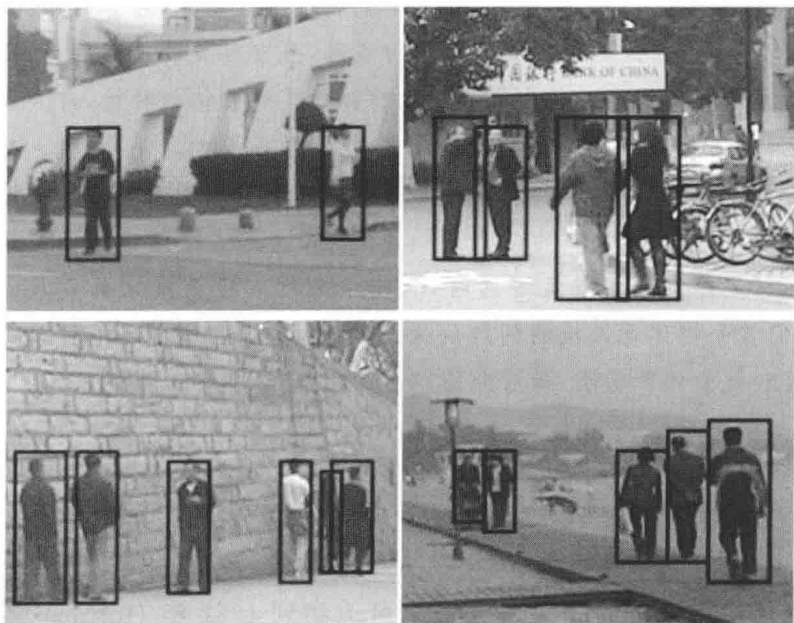


图 1-1 行人检测示意图

对输入图片(或者视频中的某一帧)进行检测,给出图片中行人的位置信息,通常用矩形框 (x, y, w, h) 表示。 (x, y, w, h) 表示矩形框左上角的 x 坐标、 y 坐标、宽和高

的具体位置,即用一个矩形框框出人体。矩形框的信息包括四个参数:左

上角的坐标(x 和 y)、矩形框的宽(w)和高(h)。注意,定义中的“图片”指的是广义的图片,可以是可见光的彩色或者灰度图片,也可以是深度图像、红外图像或者热成像图像等其他非可见光图像。

1997 年,Oren 等^[10]提出的“Haar 小波特征+支持向量机”可能是最早的行人检测方法。虽然目前许多行人的特征表示检测性能优于小波特征的检测性能,但 Oren 等人^[10]的主要贡献在于将机器学习的思想引入行人检测中。这篇文章的贡献点可以归纳为三个:

(1)提出了 Haar 小波特征,该特征后来与 AdaBoost 技术相结合,使人脸检测问题得到成功的解决;

(2)引入支持向量机(Support Vector Machine,SVM)作为分类器;

(3)使用自举(Bootstrap)的方法收集行人负样本。

其中(2)(3)这两点仍广泛应用于当前的行人检测技术中。2001 年,Mohan 等^[11]提出基于样例学习的行人检测方法,将人体看成是部件的组合,能在一定程度上解决行人检测中的多姿态和遮挡问题,进一步提高了行人检测的性能,但离实用化仍有一定距离。2005 年,Dalal 等^[12]提出了基于梯度方向直方图特征(Histogram of Oriented Gradient,HOG)的行人检测方法,将行人检测技术的研究推向新的高潮。

车辆辅助驾驶是行人检测的另一重要应用领域。2000—2005 年,Gavrila^①领导的研究团队成功开发了实时的行人检测与保护系统,并应用在智能车辆辅助驾驶系统中。欧盟从 2000 年到 2010 年连续资助 PROTECTOR、SAVE-U^[13]、APALACT^② 和 FNIR^③ 等项目,开发了多个行人检测系统;意大利 Parma 大学的 ARGO^[14]智能车辆也包括一个行人检测模块;以色列的 MobilEye 公司^[15]开发了芯片级行人检测系统;日本本田汽车公司^[16]开发了基于红外摄像机的行人检测系统;国外研究机构如 Caltech^[17-19]、TUD^[20]、巴塞罗那自治大学^{④[21-23]}和国内的西安交通大学、清华大学、北京大学、复旦大学、厦门大学等^[24-26]也在该领域做了许多研究工作。

尽管拥有巨大的市场前景,研究者发现行人检测的挑战性不亚于跟踪、识别等高层的视觉分析。由于行人检测涉及模式识别、图像处理、计算

① <http://gavrila.net/>.

② <http://www.prevent-ip.org/>.

③ <http://www.fnir.eu>.

④ <http://www.cvc.uab.es/adas/index.php?section=publishations>.

机视觉和机器学习等多学科的知识,同时受到衣着、光照变化、身体姿势、尺度变化、视角变化和复杂背景等的影响(如图 1-2 所示),因而行人检测至今仍然没有达到令人满意的效果。以下详细分析这些因素对行人检测的影响,并对其常见的解决方案进行扼要介绍。



图 1-2 行人检测难点

行人的表观受衣着颜色(a)、场景光照(b)、行走姿态(c)和视角、遮挡、携带附属物(d)等的影响,使得其表观模式千变万化

光照使得同一人体在不同场景中呈现出不同的表观。光照问题可以通过设计鲁棒的特征描述子,即不受光照影响(或者影响较小)的图像特征来解决。常见的特征有:Haar^[10]、梯度方向直方图(Histogram of Oriented Gradient)^[12]、局部二值模式(Local Binary Pattern)^[27]、积分通道特征(Integral Channel Feature, ICF)^[17]等。

尺度因素的影响,即图片或者视频中的人体大小不一。产生尺度变化的原因有二:一是人体有高矮胖瘦之分;二是摄像机与人体之间的距离不一样,使得在成像时“近大远小”。针对尺度因素,目前常用的解决方案是滑动窗口法。滑动窗口法是目标检测中的常用定位技术,通常是训练一个

固定尺度的模型,然后利用该模型在图片中进行多尺度扫描,即缩放原始图片后,在尺度空间中进行扫描。在滑动窗口中,通常的做法是将训练图片归一化到某一固定的宽高,然后训练一个固定尺度的模型。由于多尺度扫描过程中需要重新对图片进行特征提取,而特征提取是滑动窗口法实时化的瓶颈,因此近年来不少研究者提出了多尺度模型的滑动窗口法。多尺度模型的滑动窗口法在检测的过程中,在同一尺度的图片中应用多个不同尺度的模型进行检测,因此减少了缩放图片的次数,即减少了图片特征计算的次数,检测速度较快。

姿态变化使得:(1)同一个人体呈现出不同的轮廓;(2)目标内相互遮挡,即在行走过程中,肢体之间会相互遮挡;(3)训练时样本较难对齐。针对姿态因素的影响,研究者们提出了基于部位的检测方法和基于局部特征块的检测方法。基于部位的检测方法对人体的各个部位建立表观模型,同时建立部位之间的空间位置关系模型,代表性工作有 Felzenszwalb 等提出的形变部位模型^[28]、图案结构模型^[29]、Poselet^[30]等;基于局部特征块的方法首先通过局部点检测算子获取人体身上的关键点,然后利用已经训练好的局部块与目标中心的相对位置关系进行投票,最后通过均值漂移(Mean Shift)等方法获取投票中心。基于局部块的方法本质上是广义的霍夫变换(Hough Transformation),代表性工作有隐式形状模型、高效子窗口搜索模型。

由于不同场景下的拍摄角度不同,因而人体的表观呈现多样化。在车辆辅助驾驶中,摄像头的加载高度一般在 0.8~1.5 m 之间,拍摄出来的人体轮廓较为完整,透视变换的影响较小;在智能视频监控或者智能家居中,摄像头的安装高度变化范围较大,最高可能达数十米,因此拍摄出来的人体表观受透视变换的影响较大。目前基于单摄像头很难提取出视角不变性特征,因此常用的解决方案是:(1)根据摄像头的几何参数对场景进行几何变换;(2)在不同的视角下重新训练分类器。

遮挡包括肢体内的相互遮挡、人体和场景中的其他物体之间的遮挡等情况。遮挡使得提取的行人特征中部分数据损坏,因此在遮挡情况下的行人检测问题可视为是部分特征丢失或者损坏下的模式识别问题。目前常用的解决方案是采用基于部位的行人检测方法,其基本思路是:分而治之。通过将人体划分为不同的部位,即使某些部位被遮挡了,其他部位的特征仍然是完整的,通过这些完整的部位检测器来判断当前检测窗口是否包含

行人。

复杂场景下人体的表现与背景相似,使得在分类的过程中类间间距较小,容易造成误检率高。常用的解决方案是对负样本进行多轮的重采样:利用当前样本集中的正样本和负样本训练一个分类器模型,然后将该分类器模型在负样本的图片上进行检测,将误判的人体样本重新加入负样本集中,利用扩充后的负样本集和原来的正样本集,训练新的分类器,重复上述过程。为了防止过拟合,一般训练 3~5 遍。该方法在训练的过程中,一方面要求负样本的训练要足够大,另一方面由于内存的限制,不可能把所有的负样本全部加载进去,Felzenszwalb 等针对这种情况提出了一种负样本数据挖掘方法。

综上所述,行人检测受各种因素的影响,挑战性极大,但已取得了很好的进展,归纳起来如下:训练数据集的规模更大,检测的准确率更高,检测速度更快(见第二章行人检测技术的综述)。尽管如此,行人检测离实用化还有一定距离,需要研究者们继续努力。以下总结一下目前行人检测常用的一些公共数据集。

1.2 公共数据集

在计算机视觉中,数据集对一个领域的发展起到了巨大的推动作用。比如,立体视觉中的 Middlebury^① 和 KITTI 数据集^②、人脸检测中的 FDDB^③ 数据集、人脸识别的 MegaFace 数据集^④、图片分类中的 ImageNet 数据集^⑤、人体行为识别中的 UCF 数据集^⑥,这些数据集都极大地促进了计算机视觉各自领域的发展。行人检测也不例外。

当前行人检测的数据集呈现出如下的发展趋势:规模由小变大,由静态图像数据集发展为视频库,由单目视觉采集的图像库发展为基于双目视

① <http://vision.middlebury.edu/stereo/>.

② http://www.cvlibs.net/datasets/kitti/eval_stereo.php.

③ <http://vis-www.cs.umass.edu/fddb/>.

④ <http://megaface.cs.washington.edu/>.

⑤ <http://www.image-net.org/>.

⑥ <http://crcv.ucf.edu/data/UCF101.php>.