

# 行人检测技术综述

苏松志<sup>1,2</sup>, 李绍滋<sup>1,2</sup>, 陈淑媛<sup>3</sup>, 蔡国榕<sup>1,2,4</sup>, 吴云东<sup>4</sup>

(1. 厦门大学信息科学与技术学院, 福建厦门 361005; 2. 厦门大学福建省仿脑智能系统重点实验室, 福建厦门 361005;  
3. 元智大学资讯工程系, 台湾; 4. 集美大学理学院, 福建厦门 361021)

**摘要:** 行人检测是计算机视觉中的研究热点和难点, 本文对 2005-2011 这段时间内的行人检测技术中最核心的两个问题——特征提取、分类器与定位——的研究现状进行综述. 文章中首先将这些问题的处理方法分为不同的类别, 将行人特征分为底层特征、基于学习的特征和混合特征, 分类与定位方法分为滑动窗口法和超越滑动窗口法, 并从纵横两个方向对这些方法的优缺点进行分析和比较, 然后总结了构建行人检测器在实现细节上的一些经验, 最后对行人检测技术的未来进行展望.

**关键词:** 行人检测; 目标检测; 智能监控; 车辆辅助驾驶

**中图分类号:** TP391      **文献标识码:** A      **文章编号:** 0372-2112 (2012) 04-0814-07

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2012.04.031

## A Survey on Pedestrian Detection

SU Song-zhi<sup>1,2</sup>, LI Shao-zi<sup>1,2</sup>, CHEN Shu-yuan<sup>3</sup>, CAI Guo-rong<sup>1,2,4</sup>, WU Yun-dong<sup>4</sup>

(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 Computer Engineering and Science, Yuan-Ze University, Taiwan, China;

4. School of Science, Jimei University, Xiamen, Fujian 361021, China)

**Abstract:** Pedestrian detection is an active area of research with challenge in computer vision. This study conducts a detailed survey on state-of-the-art pedestrian detection methods from 2005 to 2011, focusing on the two most important problems: feature extraction, the classification and localization. We divided these methods into different categories; pedestrian features are divided into three subcategories: low-level feature, learning-based feature and hybrid feature. On the other hand, classification and localization is also divided into two sub-categories: sliding window and beyond sliding window. According to the taxonomy, the pros and cons of different approaches are discussed. Finally, some experiences of how to construct a robust pedestrian detector are presented and future research trends are proposed.

**Key words:** pedestrian detection; object detection; intelligent surveillance; driver assistance systems

## 1 引言

行人检测可定义为: 判断输入图片(或视频帧)是否包含行人, 如果有, 给出位置信息. 它是车辆辅助驾驶<sup>[1-4]</sup>、智能视频监控和人体行为分析<sup>[5,6]</sup>等应用中的第一步, 近年来也应用在航拍图像<sup>[7]</sup>、受害者营救<sup>[8]</sup>等新兴领域中. 行人兼具刚性和柔性物体的特性, 外观易受穿着、尺度、遮挡、姿态和视角等影响, 使得行人检测成为计算机视觉的研究难点与热点.

自 2005 年以来, 行人检测技术的训练库趋于大规

模化、检测精度趋于实用化、检测速度趋于实时化(2.67 帧每秒<sup>[9]</sup>). 在这样的形势下, 有必要及时地对行人检测技术的现状进行梳理. 许言午等<sup>[1]</sup>从技术和原型系统这两方面总结了 2001 年至 2006 年这段时间内行人检测系统的进展. Erzweiler 等<sup>[10]</sup>和 Dollar 等<sup>[4]</sup>在公共数据集上比较了行人检测技术经典算法之间的性能差异; 文献 [2, 3] 总结了车辆辅助驾驶系统中行人感兴趣区域分割和识别技术这两方面的研究现状. 这些综述文献从不同的侧面对行人检测的技术进行梳理, 但缺少对不同领域下行人检测技术的共性问题进行深入剖析.

收稿日期: 2010-04-07; 修回日期: 2011-06-30

基金项目: 国家自然科学基金(No. 60873179); 高等学校博士学科点专项科研基金(No. 20090121110032); 深圳市科技计划基础研究(No. JC200903180630A); 深圳市科技研发基金-深港创新圈计划(No. ZYB200907110169A); 福建省教育厅基金(No. JA10196)

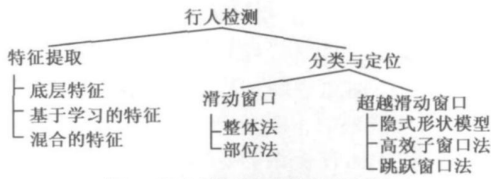


图1 行人检测技术的分类方法

本文对行人检测技术中最核心的两个问题——特征提取、分类与定位——进行综述。文章中首先将这两个问题的处理方法分为不同的类别(见图1),然后从纵横两个方向对这些方法进行分析和比较,文章最后给出了我们在构建行人检测系统时积累的一些经验,并对行人检测技术的未来进行展望。篇幅所限,本综述主要总结2005-2011这段时间内的行人检测技术进展。

## 2 特征提取

行人特征描述子可分为三类:底层特征、基于学习的特征和混合特征。底层特征指的是颜色、纹理和梯度等基本的图像特征;基于学习的特征指的是通过机器学习的方法,从大量的行人样本中学习到的行人特征表示;混合特征指的是多种底层特征的融合,或者是底层特征的高阶统计特征。这三类特征的优缺点如表1所示。本节最后依时间顺序列出了2005-2011较具代表性的行人特征,见表2。

表1 各类行人特征描述子的优缺点

行人特征类型	优点	缺点
底层特征	单一特征,计算速度快,可利用积分图技术快速计算	只从某一方面,如梯度或者纹理来描述行人特征,判别力较差
基于学习的特征	能从大量的样本中选出判别能力较强的特征	特征的选择与训练样本密切相关,若样本集不具有代表性,很难选择出好的特征
混合特征	能从不同的侧面来刻画图像特征,提高检测的准确率	特征维度增加,特征的计算和分类器的预测时间也增加,影响实时性。

### 2.1 底层特征

Dalal等提出HOG(Histogram of Oriented Gradient, HOG)<sup>[11]</sup>是目前广泛使用的行人特征描述子。HOG刻画图像的局部梯度幅值和方向特征,基于梯度特征、对块的特征向量进行归一化处理、允许块之间相互重叠,因此对光照变化和小量的偏移并不敏感,能有效地刻画出人体的边缘特征。

HOG也有其缺点:维度高、计算慢。针对这些缺点,Zhu等<sup>[12]</sup>允许HOG中块大小可变,利用积分直方图技术来快速计算HOG特征,通过Adaboost算法选择判别能力较强的块,然后构建级联的分类器,该方法的检测速度比Dalal等<sup>[11]</sup>的快将近70倍。Wojek等<sup>[13]</sup>则采用并

行技术,在GPU上实现HOG,构建了一个实时的行人检测系统。

局部二值模式(Local Binary Pattern, LBP)最早是由Ojala等<sup>[14]</sup>提出的一种用于纹理分类的特征提取方法,广泛应用在人脸识别中。Mu等<sup>[15]</sup>根据行人的特点,提出LBP的两个变种:Semantie-LBP(S-LBP)和Fourier LBP(F-LBP)。Wang等<sup>[16]</sup>则简单地将局部图像块的LBP直方图特征串联起来作为行人的特征描述子,其检测性能并不比S-LBP差,但Walk等<sup>[17]</sup>在其他行人数据上的实验表明HOG与LBP特征的结合并没有提高检测性能。究其原因LBP在图像比较模糊或者光照变化强烈等成像条件较差时,不能有效地刻画出纹理特征。与LBP特征类似的有,Wu等提出的CENTRIST特征<sup>[18]</sup>,即CENSus TRansform hISTogram。该特征能刻画场景的全局信息,最早用在场景分类中。2011年Wu等将CENTRIST应用在行人检测中<sup>[19]</sup>,利用积分图技术快速计算该特征,并与级联分类器相结合构建了一个实时的行人检测系统。

表2 2005-2011行人特征概览表

行人特征	年限	类别*	概述
HOG <sup>[11]</sup>	2005	0	刻画图像的梯度特征,最为广泛使用和成功的行人特征,可采用积分图技术快速计算。
Edgelet <sup>[23]</sup>	2005	1	描述行人的局部轮廓方向特征。
边缘模板 <sup>[24]</sup>	2007	1	从样本中学习行人的形状分布模型,模板需手工标注。
LBP <sup>[15, 16]</sup>	2008	0	计算速度快,对单调变化的灰度特征具有不变性,低分辨率下判别能力较差。
COV <sup>[26]</sup>	2008	2	局部区域图像块中各象素的坐标、灰度的一阶导数、二阶导数和梯度方向的协方差特征。
C <sub>0</sub> -HOG <sup>[27]</sup>	2009	2	通过“梯度对”更好地描述了梯度的空间分布特征,但是向量的维度太高。
积分通道特征 <sup>[29]</sup>	2009	2	一种自然的多种异质特征融合方法;参数少且对具体参数值不敏感;检测器空间定位更加准确。
自适应轮廓特征 <sup>[25]</sup>	2009	1	具有较强的判别能力,能自适应地从训练样本中学习出行人的轮廓特征。
GC <sup>[28]</sup>	2009	2	刻画了图像区域的结构信息和统计特征,可看作介于HOG(统计)和Edgelet(结构)之间的特征。
CSS <sup>[20]</sup>	2010	0	利用局部部位间的颜色相似性来刻画人体的结构特征,与HOG相结合,大大提高了检测性能。
CENTRIST <sup>[19]</sup>	2011	0	刻画图像的全局轮廓特征,计算简单快速,与级联分类器相结合,可达到实时。

注:0:底层特征,1:基于学习的特征,2:混合特征

颜色是最基本的图像特征,易受光照的影响,不适合作为行人的特征描述子,但是由于人体的结构具有相对的稳定性,不同部位的颜色之间差异较小,因此

Walk 等<sup>[20]</sup>提出了颜色自相似特征来刻画局部块特征之间的相互关系,与 HOG 特征相结合,大大提高了检测性能.

## 2.2 基于学习的特征

基于学习的特征,利用 Boosting 进行特征选择,选择出来的特征可看作是行人的中间层表示.如 Viola 等<sup>[21]</sup>提出了基于“AdaBoost+ Haar”的人脸检测方法中,利用 AdaBoost 从大量的 Haar 特征中选择判别能力较强的特征(弱分类器),后来该方法成功地应用在行人检测中<sup>[22]</sup>. Wu 等<sup>[23]</sup>提出 Edgelet 来描述行人的局部轮廓方向特征.所谓的 Edgelet 是一段长为  $k$  的直线或弧线,记其上各象素点的梯度幅值和方向分别为  $\{n_i^E\}_{i=1}^k$  和  $\{u_i\}_{i=1}^k$ ,则 Edgelet 在图像  $I$  中点  $p$  处的特征值  $Edgelet(p)$  定义为:

$$Edgelet(p) = \frac{1}{k} \sum_{i=1}^k M^l(u_i + p) | < n^l(u_i + p), n_i^E > |$$

其中,  $M^l(\cdot)$  和  $n^l(\cdot)$  表示由 Sobel 算子计算出的边缘强度和方向.

Gavrila 等<sup>[24]</sup>提出 Chamfer System 也是基于边缘特征的行人检测方法,该方法从样本中学习待检测目标的形状分布模型,采用层次聚类的方法对模板进行索引以提高检测速度,缺点是需手工标注模板;边缘模板依赖于二值化边缘的提取,在边缘不显著的情况下,可能出现漏检;层次索引中,不同层次的边缘模板仍然有很大的相似性,存在大量重复计算.与 Gavrila 的边缘模板特征相比,Edgelet 特征具有如下优点:由机器学习算法自动筛选,不需要手动标注;描述行人局部特征,能有效处理遮挡;综合考虑边缘的强度和方向特征,误检率较低;

Gao 等<sup>[25]</sup>提出了自适应的轮廓特征,利用 Adaboost 算法在方向粒度空间中进行特征选择,定义了“生长”、“合并”和“切割”三种操作,能有效刻画形状的共现特征.

## 2.3 混合特征

Tuzel 等<sup>[26]</sup>利用各种不同特征(象素的坐标、灰度的一阶导数、二阶导数和梯度方向等)的协方差矩阵来描述行人的局部区域特征,将协方差矩阵视为联通的黎曼流形,在黎曼几何空间中对行人进行分类. Watanabe 等<sup>[27]</sup>则采用类似灰度共生矩阵的特征,提出了共生梯度方向直方图特征(Co-occurrence Histograms of Oriented Gradients, 简称 CoHOG). CoHOG 通过“梯度对”更好地描述了梯度的空间分布特征,缺点是向量维度太高. Liu 等<sup>[28]</sup>提出 GGP (Granularity-tunable Gradients Partition) 特征,粒度用来定义霍夫空间中线段的空间位置和角度的不确定性,该特征刻画了图像区域的结构信息和统

计特征,并将二者统一在可调粒度的梯度空间中.

Dollar 等<sup>[29]</sup>提出了积分通道特征,利用积分图技术对图像的各特征通道(线性或非线形变换后的图像),如局部和、直方图特征和哈尔特征等进行快速计算.该方法不仅将多特征有机结合起来,且解决了多特征融合计算速度慢的缺点.有部分研究者通过行人的颜色、形状、纹理和场景等多种信息相互融合,构建鲁棒的行人检测系统.如 Schwartz 等<sup>[30]</sup>将 HOG 特征与颜色、纹理和边缘等特征相结合,构建一高维的描述子,并通过偏最小二乘降维.其他的特征混合方式有:形状与纹理信息相结合<sup>[24]</sup>,表现信息与运动信息相结合<sup>[22,31]</sup>,表现信息与深度信息相结合<sup>[20]</sup>.

## 3 分类与定位

行人检测中的分类指的是判断当前检测窗口是否包含行人,定位指的是行人在图片中的具体位置.行人的分类与定位方法可分为两大类:滑动窗口法(Sliding Window)和超越滑动窗口法(Beyond Sliding Window).滑动窗口法是目标检测中的主流方法,主要思路如下:选一固定大小的窗口在图像的尺度空间内(即缩放后得到的图片)依次滑动窗口,利用预先训练好的分类器判断窗口内是否包含行人,通过窗口的滑动对图片进行遍历,从而给出行人在图片中出现的位置.超越滑动窗口法是近年来随着基于“词袋”(bag-of-words, BOW)的场景分类技术<sup>[32]</sup>的发展而出现的,其中较具代表性的方法有隐式形状模型<sup>[33]</sup>、高效子窗口法<sup>[34]</sup>和跳跃窗口法<sup>[35]</sup>.文献<sup>[36]</sup>的 5.3.5 节通过实验比较了二者在不同性能评价准则下的检测率,有关滑动窗口法和超越滑动窗口法的优缺点分析见表 3.

表 3 两种分类与定位方法的优缺点分析

	优点	缺点
滑动窗口法	方便与级联和积分图等技术相结合,提高检测速度.	扫描窗口以固定的长度步进,搜索结果不是全局最优的,可能造成漏检.
超越滑动窗口法	全局最优搜索;通过局部特征块进行投票,能在一定程度上处理部分遮挡和多姿态问题	需要在投票空间中寻找最大值,其检测速度无法实时;局部特征的判别能力较弱,误检率较高.

### 3.1 滑动窗口法

滑动窗口法依据所利用的人体结构信息可进一步分为整体法和部位法.整体法是对扫描窗口内的图像进行特征提取,获取行人的全局信息;部位法对行人的部位(如头肩部、躯干、手和脚等)进行特征提取,构建部位间的几何关系.

#### 3.1.1 整体法

目前大部分的行人检测方法<sup>[11,13,16,21,37]</sup>采用整体法,这些方法的不同在于分类器和特征,第 2 节已对行

人的特征进行了综述, 这里主要总结目前行人检测中常用的分类器: SVM、Boosting 和多示例学习等。

Oren 等<sup>[38]</sup>最早提出基于 SVM 的行人检测方法, 随着核函数在计算机视觉的广泛应用, Maji 等<sup>[39]</sup>提出基于直方图交叉核支持向量机(Histogram Intersection Kernel SVM, HIKSVM)的行人检测方法, 并设计相应的快速近似算法, 性能优于线性核 SVM。

最早将 AdaBoost 应用到目标检测的是 Viola 等<sup>[21]</sup>, Viola 等利用哈尔特征、AdaBoost 算法和级联分类器成功地实现第一个实时人脸检测系统, 并将该方法应用到智能监控的行人检测中<sup>[22]</sup>。Chen 等<sup>[40]</sup>改进级联结构, 在每级输出中引入 Meta Stage, 利用每级的输出构建分类器, 大大提高了行人检测的速度和准确性。Tuzel 等<sup>[26]</sup>针对协方差矩阵, 在黎曼流形空间中利用 Logit-Boost 对行人进行分类。Kim 等<sup>[41]</sup>提出多分类器增强算法 MCBoost(Multiple Classifier Boosting)对图像和视觉特征进行协同聚类(Co-Clustering), MCBoost 能有效解决目标检测中的多类别和多视角问题, 在 INRIA 上的实验表明 MCBoost 的性能优于 AdaBoost。

多示例学习已成为与监督学习、非监督学习和强化学习并列的一种机器学习框架。Lin 等<sup>[42]</sup>和 Babenko 等<sup>[43]</sup>针对行人检测中因姿态问题所导致的局部特征不匹配的现象, 提出基于多示例学习框架下的基于部位行人检测方法。该方法能自动对齐行人的局部特征, 对部位的形变进行建模, 与级联结构相比, 所需特征数量更少, 提高了检测的准确率。

各个分类器都有其优缺点: 线性核 SVM 的训练和预测时间比较快, Boosting 技术和多示例学习的训练时间则很长, 在行人检测中目前尚无统一的实验平台来验证到底哪个分类器性能最优。

### 3.1.2 部位法

部位法将人体看成是部位的组合, 重点要解决以下两个问题: 构造有效的部位检测器、对部位间的几何关系进行建模。近年来这方面最具代表性的工作是 Felzenszwalb 等<sup>[44]</sup>提出的形变部位模型(Deformable Part Model, DPM)。DPM 包括低分辨率的整体模型、高分辨率下的部位模型及部位相对于整体的形变模型。为了得到形变部位模型的参数, Felzenszwalb 把各个部位相对于整体的位置偏移作为隐变量, 提出了基于隐式支持向量机的参数求解方法。DPM 虽然能在一定程度上解决人体姿态的变化, 但是检测的速度较慢。由于 DPM 包含低分辨率模型和高分辨率模型, 低分辨率模型可以快速去除大部分非行人区域, 因此 Felzenszwalb 等<sup>[45]</sup>借鉴 Viola 在人脸检测中采用的级联思想, 提出级联形变部位模型, 大大提高了检测速度。

Pedersoli 等<sup>[46]</sup>通过实验观察到: 基于部位法的目标

检测中, 部位与图像进行匹配占用了大部分的检测时间。为此他们提出了由粗到细的快速形变部位模型法: 低分辨率下将部位和图像进行匹配, 将局部窗口内检测器的响应值最大的区域保留下来, 高分辨率检测器只在低分辨率检测器保留下的区域内进行进一步的检测。该方法大大缩短了检测和训练时间。

部位法的难点在于如何对部位进行划分和构建有效的部位检测, Parikh 等<sup>[47]</sup>的研究表明: 部位法中部位检测器对检测性能的影响大于部位之间的几何关系对检测性能的影响。

整体法的优点是训练数据库的标注只需用矩形框标出行人的区域, 而部位法为了建立部位模型则需要标注出行人各部位的区域。整体法的缺点是无法克服部分遮挡的影响, 部位法采用“分而治之”的策略, 能在一定程度上克服部分遮挡的影响。

### 3.2 超越滑动窗口

“词袋”模型是场景分类中的代表性方法, 通常包含如下三个步骤<sup>[48]</sup>: 视觉词典的构建、基于词典的图片表示和分类器学习。超越滑动窗口是基于 BOW 的方法, 目前常见的有隐式形状模型<sup>[33]</sup>、高效子窗口法<sup>[34]</sup>和跳跃窗口法<sup>[35]</sup>。其中, 高效子窗口的视觉词典并没有考虑局部特征块的空间信息, 利用直方图特征来表示图片, 分类器采用线性核 SVM; 隐式形状模型和跳跃窗口均考虑了局部特征块相对于目标中心的偏移, 其几何模型为星型模型。以下分别介绍这三种方法。

高效子窗口搜索法(Efficient Subwindow Search, ESS)是 Lampert 等<sup>[33]</sup>提出的一种快速目标定位技术, 该方法利用线性支持向量机的可加性, 将分类器的输出转化为各个维度的权重累加和, 采用分支限定技术寻找参数空间的极大值。ESS 虽然能在参数空间中寻找一个最优值, 但是检测速度较慢。不少研究者对 ESS 进行改进和扩展。An 等<sup>[49, 50]</sup>改进 ESS 的搜索策略, 构建更为紧凑的上界函数, 提高了 ESS 的检测速度。Vijayanarasimhan 等<sup>[51]</sup>提出了基于区域搜索的目标检测方法, 该方法能定位不规则形状的目标。Yuan 等<sup>[52]</sup>将 ESS 推广到三维空间中, 提出了基于时空分支界限的人体行为检测方法。

隐式形状模型将行人检测视为广义的霍夫变换问题: 首先通过局部特征检测算子寻找关键点; 然后在关键点的周围选取一固定大小的图像块, 通过聚类、随机森林<sup>[53]</sup>或者最大间隔<sup>[54]</sup>等方法建立局部块的空间分布模式; 最后通过霍夫投票方式确定行人位置。隐式形状模型能有效解决遮挡, 但是并不是所有的物体都能利用局部特征检测器检测到关键点, 该方法只有在侧面行人检测时有较好的效果。

跳跃窗口法与隐式形状模型类似, 不同在于每个

局部特征不是对行人的中心点进行投票,而是对应一个行人可能出现的矩形框位置,最后对所有的矩形框位置进行融合。

## 4 实现细节

构建一个鲁棒的行人检测系统,除了与行人特征、分类与定位技术有关外,系统的性能还与训练样本的选择、预处理和后处理等密切相关。本节主要总结相关文献中提及的,以及我们在构建行人检测系统中积累的一些经验。

训练样本的收集 “巧妇难为无米之炊”,基于统计学习的方法来构建行人检测器,首先需要建立一个行人数据库。(1)对于通用场景的行人检测器来说,为了提高检测器的泛化能力,需要包含各种不同条件下的行人样本。(2)对于某一固定场景的行人检测器,若条件允许,最好收集针对某一固定场景下的行人正样本和负样本,避免样本的非独立同分布对检测器性能的影响。另外,(3)行人训练样本的标注中往往给出的是包含行人的最小矩形窗口,但训练时最好对该窗口进行扩充,尤其是向下延伸,如图2所示。因为行人一般站在地面上,而地面的表现特征通常相对固定。

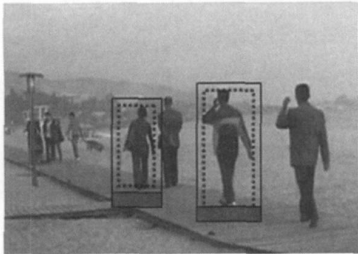


图2 训练样本的标注;图中虚线矩形框为包含行人的最小矩形窗口,训练时向下延伸窗口(图中的蓝色部分)有助于提高检测性能。注:照片来源于厦门大学数据库

**预处理** 在实时性要求较高的应用场景中,可利用场景的先验信息预先确定行人可能出现的区域,缩小检测器的搜索范围。在视频监控系统中,由于摄像头静止,背景信息相对固定,可采用减背景法获取运动目标区域,将其作为候选区域。在车辆辅助驾驶中,基于双目视觉或者其它非视觉的传感器可获取路面信息,行人检测器只需对路面区域进行扫描。

**后处理** 对检测结果进行筛选,滤掉误检的行人窗口可进一步提高检测性能。如Gavrilu等<sup>[24]</sup>通过比较立体图像对中左图像和右图像的轮廓相关性对检测结果进行验证。Leibe等<sup>[31]</sup>采用倒角匹配(Chamfer Matching)对检测结果的形状进行验证和细化。Shashua等<sup>[51]</sup>在对检测结果进行跟踪后,提取多帧的信息,如步态模式、运动特征和单帧的检测结果置信度(即分类器对单个检测窗口的输出值)等,然后根据这些特征进行验

证。

总之,通过上述对各种方法的分析和比较,围绕着行人检测器的速度和准确率,可得出如下的两条经验:

(1)提高检测速度的方法:根据应用场合,充分利用场景的先验信息,确定行人的候选区域;关于特征提取,可以利用GPU或者积分直方图技术进行快速计算;在分类器的构造上,可以结合级联、多分辨等由粗到细的方法,尽量在分类的早期去掉大量非行人的区域。(2)提高检测准确率的方法:确定行人的候选区域,降低误检率;通过多特征融合,从多个不同的方面刻画行人特征,提高准确率;采用部位法,解决因遮挡而引起的漏检。

## 5 总结及其展望

本文对2005-2011行人检测技术中最核心的两个问题—特征提取、分类与定位—进行综述,从纵横两个方向对这些方法进行分析和比较,并给出构建行人检测系统时的一些实现细节和经验。虽然行人检测技术的准确率和速度在不断的提高,但是距离实用性还依然遥远。未来的行人检测技术的研究需要从以下几个方面着手:

### (1) 多视角和遮挡问题

目前大部分行人检测的研究为单目视觉,在单目视觉下多视角和遮挡问题很难解决。为了解决该问题,可采用多摄像机或利用深度信息来检测行人。同时,在多摄像机下,人体的姿态较易获取,可探讨多目视觉中基于姿态的行人检测技术。

### (2) 特殊场景下的行人检测问题

一个鲁棒的行人检测必需在恶劣的天气环境下(如雨雪天气等)工作,系统必需能对部分遮挡、分辨率低、远距离的、携带大面积物件的行人进行准确检测,并保持低误报率。同时,建立专门针对遮挡、低分辨率和远距离的行人测试数据库。

### (3) 设计自适应的检测器

目前大部分研究者的目标是构建通用的行人检测器,而在特殊场景下尤其是摄像头静止不动的监控场合中,如何利用增量式学习、在线学习等算法,将通用的行人检测器迁移到特殊场景中,使得行人检测器在检测的过程中通过自学习提高性能将是未来的研究重点。

总之,行人检测是当今计算机视觉领域的核心难点问题,其解决具有重要的理论意义和良好的应用前景,也吸引了大量的研究人员投入该领域的研究。虽然取得了一定的成效,但有效的解决真实复杂场景下的行人检测问题,还有待进一步的研究。

## 参考文献

- [1] 许言午, 曹先彬, 乔红. 行人检测系统研究新进展及关键技术展望[J]. 电子学报, 2008, 36(5): 368–376.  
Xu Yanwu, Cao Xianbin, and Qiao Hong. Survey on the latest development of pedestrian detection system and its key technologies expectation[J]. Acta Electronica Sinica, 2008, 36(5): 368–376. (in Chinese)
- [2] 贾慧星, 章毓晋. 车辆辅助驾驶系统中基于计算机视觉的行人检测研究综述[J]. 自动化学报, 2007, 33(1): 84–90.  
Jia Huixing, Zhang Yujin. A survey of computer vision based pedestrian detection for driver assistance systems[J]. Acta Automatica Sinica, 2007, 33(1): 84–90. (in Chinese)
- [3] Geronimo D, Lopez A, Sappa A, et al. Survey of pedestrian detection for advanced driver assistance systems[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1239–1258.
- [4] Dollár P, Wojek C, Schiele B, et al. Pedestrian detection: an evaluation of the state of the art. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 99: 1–20.
- [5] Aggarwal J, Ryoo M. Human activity analysis: a review[J]. ACM Computing Surveys, 2011, 43(3), 16: 1–47.
- [6] 杜友田, 陈峰, 徐文立, 李永彬. 基于视觉的人的运动识别综述. 电子学报, 2007. 35(1): 84–90.  
Du Youtian, Chen Feng, Xu Wenli, et al. A survey on the vision-based human motion recognition[J]. Acta Electronica Sinica, 2007. 35(1): 84–90. (in Chinese)
- [7] Reilly V, Solmaz B, and Shah M. Geometric constraints for human detection in aerial imagery[A]. In Proc. ECCV[C], 2010.
- [8] Andriluka M, Schnitzspan P, Meyer J, et al. Vision based victim detection from unmanned aerial vehicles[A]. In Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)[C]. Taipei, Taiwan, 2010.
- [9] Dollar P, Belongie S, Perona P. The fastest pedestrian detector in the west[A]. In Proc. BMVC[C], 2010.
- [10] Enzweiler M, Gavrila D. Monocular pedestrian detection: survey and experiments[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2179–2195.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[A]. In Proc. IEEE CVPR[C], 2005, 886–893.
- [12] Zhu Q, Yeh C, Cheng T. Fast human detection using a cascade of histograms of oriented gradients[A]. In Proc. IEEE CVPR[C], 2006, 2: 1491–1498.
- [13] Wojek C, B. Schiele. A performance evaluation of single and multi-feature people detection[A]. In Proc. DAGM[C], 2008.
- [14] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on feature distributions[J]. Pattern Recognition, 1996, 19(3): 51–59.
- [15] Mu Y, Yan S, Liu Y, et al. Discriminative local binary patterns for pedestrian detection in personal album[A]. In Proc. IEEE CVPR[C], 2008.
- [16] Wang X, Han X, Yan S. A HOG-LBP human detector with partial occlusion handling[A]. In Proc. IEEE ICCV[C], 2009.
- [17] Walk S, Majer N, Schindler K, et al. New features and insights for pedestrian detection[A]. In Proc. IEEE CVPR[C], 2010.
- [18] Wu J X, Rehg J M. CENTRIST: A visual descriptor for scene categorization[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1489–1501.
- [19] Wu J X, Geyer C, Rehg J M. Real-time human detection using contour cues[A]. In Proc. IEEE ICRA, 2011. 860–867.
- [20] Walk S, Majer N, Schindler K, et al. New features and insights for pedestrian detection[A]. In Proc. of CVPR[C], 2010.
- [21] Viola P, Jones M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2): 137–154.
- [22] Viola P, Jones M, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision, 2005, 63(2): 153–161.
- [23] Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors[A]. In Proc. IEEE ICCV[C], 2005.
- [24] Gavrila D. A bayesian, exemplar-based approach to hierarchical shape matching[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(8): 1–14.
- [25] Gao W, Ai H, Lao S. Adaptive contour features in oriented granular space for human detection and segmentation[A]. In Proc. IEEE CVPR[C], 2009.
- [26] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on Riemannian manifolds[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(10): 1713–1727.
- [27] Wantanbe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for pedestrian detection[A]. In Proc. of The 3<sup>rd</sup> Pacific Rim Symposium on Image and Video Technology [C], Tokyo, Japan, 2009. 37–47.
- [28] Liu Y, Shan S, Zhang W, et al. Granularity-tunable gradients partition (GGP) descriptors for Human Detection[A]. In Proc. IEEE CVPR[C], 2009.
- [29] Dollar P, Tu Z W, Perona P, et al. Integral channel features[A]. In Proc. BMVC[C], 2009, 1–11.
- [30] Schwartz W, Kembhavi A, Harwood D, et al. Human detection using partial least squares analysis[A]. In Proc. IEEE ICCV[C], 2009.
- [31] Wojek C, Walk S, Schiele B. Multi-cue onboard pedestrian detection[A]. In Proc. IEEE CVPR[C], 2009.
- [32] Li Fei-fei, Perona P. A bayesian hierarchical model for learning natural scene categories[A]. In Proc. IEEE CVPR[C], 2005.

- [33] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation [J]. International Journal of Computer Vision, 2008, 77(3): 259–289.
- [34] Lampert C, Blaschko M, Hofmann T. Efficient subwindow search: a branch and bound framework for object localization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2129–2142.
- [35] Chum O, Zisseman A. An exemplar model for learning object classes [A]. In Proc. IEEE CVPR[C], 2007.
- [36] 苏松志. 行人检测若干关键技术研究[D]. 福建厦门, 厦门大学博士学位论文, 2011.  
Su Song-zhi. Research on key technologies of pedestrian detection[D]. Fujian, Xiamen, Xiamen University, Ph. D Thesis, 2011. (in Chinese)
- [37] 田广, 戚飞虎. 移动摄像机环境下基于特征变换和 SVM 的分级行人检测算法[J]. 电子学报, 2008, 36(5): 1024–1028.  
Tian Guang, Qi Fei-hu. Feature transformation and SVM based hierarchical pedestrian detection with a monocular moving camera[J]. Acta Electronica Sinica, 2008, 36(5): 1024–1028. (in Chinese)
- [38] Oren M, Papageorion C, Sinha P, et al. Pedestrian detection using wavelet templates[A]. In Proc. IEEE CVPR[C], 1997. 193–199.
- [39] Maji S, Berg A, Malik J. Classification using intersection kernel SVMs is efficient[A]. In Proc. IEEE CVPR[C], 2008.
- [40] Chen Y T, Chen C S. Fast human detection using a novel boosted cascading structure with meta stages[J]. IEEE Trans. on Image Processing, 2008, 17(8): 1452–1464.
- [41] Kim T K, Cipolla R. MCBoost: Multiple classifiers boosting for perceptual co-clustering of images and visual features[A]. In Proc. IEEE NIPS[C], 2008.
- [42] Lin Z, Hua G, Davis S. Multiple instance feature for robust part-based object detection[A]. In Proc. IEEE CVPR[C], 2009.
- [43] Babenko B, Dollár P, Tu Z, et al. Simultaneous learning and alignment: multi-instance and multi-pose learning [A]. In Proc. IEEE ECCV[C], 2008.
- [44] Felzenszwalb P, Girshick R, McAllester D, et al. Object detection with discriminatively trained part based models[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627–1645.
- [45] Felzenszwalb P, Girshick R, McAllester D. Cascade object detection with deformable part models[A]. In Proc. IEEE CVPR[C], 2010.
- [46] Pedersoli M, Vedaldi A, Gonzalez J. A coarse-to-fine approach for fast deformable object detection [A]. In Proc. IEEE CVPR[C], 2011.
- [47] Parikh D, Zitnick C. Finding the weakest link in person detectors [A]. In Proc. IEEE CVPR[C], 2011.
- [48] Agarwal A, Awan A, Roth D. Learning to detect objects in images via a sparse, part-based representation [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1475–1490.
- [49] An Senjian, Peursum P, Liu Wanquan, et al. Efficient algorithms for subwindow search in object detection and localization [A]. In Proc. IEEE CVPR[C], 2009.
- [50] An Senjian, Peursum P, Liu Wanquan, et al. Exploiting monge structures in optimum subwindow search [A]. In Proc. IEEE CVPR[C], 2010.
- [51] Vijayanarasimhan S, Grauman K. Efficient region search for object detection [A]. In Proc. IEEE CVPR[C], 2011.
- [52] Yuan J S, Liu Z C, Wu Y. Discriminative video pattern search for efficient action detection [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 99: 1–17.
- [53] Gall J, Lempitsky V. Class-specific hough forests for object detection [A]. In Proc. IEEE CVPR[C], 2009.
- [54] Maji S, Malik J. Object detection using a max-margin hough transform [A]. In Proc. IEEE CVPR[C], 2009.
- [55] Shashua A, Gdalyahu Y, Hayun G. Pedestrian detection for driving assistance systems: single-frame classification and system level performance [A]. In Proc. IEEE Intelligent Vehicles Symposium [C], 2004. 1–6.

#### 作者简介



苏松志 男, 厦门大学信息科学与技术学院助理教授, 1982年11月出生于福建泉州。2005年毕业于山东大学, 2008年、2011年于厦门大学获得工学硕士和理学博士学位。主要从事行人检测和人体行为分析的研究。  
E-mail: ssz1021@gmail.com



李绍滋 男, 厦门大学信息科学与技术学院教授, 博导, 1963年3月出生于湖南常德。1983年毕业于湖南大学, 1988年毕业于西安交通大学, 获工学硕士学位, 2009年毕业于国防科技大学, 获工学博士学位。主要研究方向为: 人工智能及其应用、计算机视觉、机器学习等。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、深圳科技计划项目等研究。 E-mail: Hszlig@xmu.edu.cn



陈淑媛 女, 台湾元智大学教授, 博导。1958年11月出生于台湾花莲。1980年、1982年、1990年于台湾国立交通大学分别获得工学学士、硕士和博士学位, 主要从事计算机视觉、模式识别和图像检索等研究。  
E-mail: cschen@saturn.yzu.edu.tw