

改进积分通道特征的快速多尺度行人检测

黄鹏, 于凤芹, 陈莹

HUANG Peng, YU Fengqin, CHEN Ying

江南大学 物联网工程学院, 江苏 无锡 214122

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

HUANG Peng, YU Fengqin, CHEN ying. Improved integral channel features for fast multiscale pedestrian detection. Computer Engineering and Applications

Abstract: In order to overcome the problem that integral channel features has redundant information and slow detection speed in multiscale pedestrian detection, an improved integral channel features for fast multiscale pedestrian detection is proposed. Firstly, fast feature pyramids is used to compute multiscale channel features, which avoid computing the same features at multiple locations and scales. Then the detection windows are divided into cells and blocks to make the overall description of the image and reduce the redundancy, replacing the original method of random location and size. Finally, obtaining the sums of pixels in cells and blocks as the pedestrian features and then the features are classified by soft cascade Adaboost. Simulation experiment results show that the proposed method gives improved accuracy and 15.4 times speedup compared with the original algorithm and runs at 18.5fps on 640×480 images.

Key words: pedestrian detection; integral channel features; multi-scale; fast feature pyramids

摘要: 针对积分通道特征冗余信息多, 在多尺度行人检测中检测速度较慢的问题, 提出了改进积分通道特征的快速多尺度行人检测算法。该方法首先采用快速特征金字塔计算图像不同尺度下的特征通道, 避免对图像重复缩放计算特征通道; 然后将检测窗口分为单元和块来对图像进行整体描述, 代替原始方法的随机位置和大小矩形来减少冗余特征, 最后计算单元和块内的像素和作为特征向量送入软级联 Adaboost 分类器进行分类。仿真实验结果表明, 该算法检测精度优于积分通道特征算法, 同时检测速度提高了 15.4 倍, 在 640×480 大小图像上检测速度达到 18.5fps。

关键词: 行人检测; 积分通道特征; 多尺度; 快速特征金字塔

doi:10.3778/j.issn.1002-8331.1512-0378

文献标志码: A

中图分类号: TN911.73

1 引言

行人检测在智能汽车辅助驾驶中具有极其重要的作用, 经典的 VJ^[1]和 HOG^[2]行人检测方法影响甚广, 之后很多学者对它们进行了相关改进, 近几年基于深度学习^[3]、基于部件^[4]的行人检测方法的提出, 使得检测性能不断提升, 但由于行人兼具刚性和柔性物体的特性, 外观易受穿着、遮挡、姿态和视角等影响, 行人检测离实时应用还有一段距离, 使得行人检测成为计算机视觉的研究难点与热点^[5]。

经过近十几年的发展, 学者提出了很多

有效的行人检测算法, 当前最具影响力的是 Dollar^[6]等提出的积分通道特征(ICF: Integral Channel Features)算法, 积分通道特征使用图像通道, 融合了丰富和多元化的图像信息, 具有较少的参数设置, 相比于融合多个特征的算法, 计算复杂度更低, 尤其适用于行人检测, 取得了较好的检测结果, 但是由于提取的冗余特征多, 而且要对每个尺度分别进行特征计算, 耗费了大量时间。Dollar等^[7]针对积分通道特征多尺度检测速度过慢的瓶颈, 提出了近似实时的多尺度行人检测

基金项目: 国家自然科学基金项目 (No. 61104213)。

作者简介: 黄鹏 (1992-), 男, 硕士研究生, 主要研究领域为图像与视频信号分析处理; 于凤芹 (1962-), 女, 博士, 教授, 主要研究领域为语音信号分析与处理研究, 图像信号与信息处理等; 陈莹 (1976-), 女, 博士, 教授, 主要从事计算机视觉与模式识别的研究和信号与信息处理等。

方法, 首次提出了快速计算多尺度特征的方法, 但缺乏对特征池冗余信息的改进。Benenson 等^[8]提出了基于积分通道特征的每秒 100 帧的行人检测方法, 多尺度检测时通过训练多个尺度的检测模板对输入图像进行检测, 替代了原方法对输入图像多次缩放, 加快了检测速度, 但该方法需要 GPU 进行运算加速。Benenson 等^[9]通过改进积分通道特征的特征池、训练不同尺度的检测模板等方法使得检测精度优于绝大数其他方法, 但该方法只着重于精度提升, 并没有提及检测速度。Zhang 等^[10]使用积分通道特征作为底层特征, 将站立的行人分为头部, 上半身和下半身 3 个有区别的矩形检测模板, 有效处理遮挡情况, 提升了检测率。综上对于积分通道特征的相关改进都分别着重于精度或者速度, 并没有同时对这两者综合考虑。

为了进行快速多尺度行人检测, 既要提升检测速度, 又要相应的提升检测精度, 本文提出了改进积分通道特征的快速多尺度行人检测算法, 该方法首先采用快速特征金字塔计算图像不同尺度下的特征通道, 避免了对图像重复缩放计算特征通道, 提升检测速度; 然后改进积分通道的特征池, 将随机位置和大小矩形框优化为用单元和块的方法对检测模板进行分割, 对图像进行整体描述, 减少了冗余特征, 提高检测精度; 最后使用软级联 Adaboost 分类器对特征进行分类, 通过设定固定阈值, 检测负样本时不需要遍历所有弱分类器, 更快的判断出负样本。

2 基本原理

2.1 改进的积分通道特征

积分通道特征选用 10 个通道: 6 个方向的梯度直方图、3 个 LUV 颜色通道和 1 梯度幅值。这些通道可以高效计算并且捕获输入图像不同的信息, 计算得到 10 个通道后, 分别在这 10 通道内随机选择矩形区域和大小, 求取其内部所有像素的像素值之和, 最终随机选取 30000 个矩形区域组成积分通道特征池。

灰度图像中, 像素点 (x, y) 的梯度幅值和方向分别为:

$$G(x, y) = \sqrt{[H(x+1, y) - H(x-1, y)]^2 + [H(x, y+1) - H(x, y-1)]^2} \quad (1)$$

$$\alpha(x, y) = \arctan\left(\frac{H(x, y+1) - H(x, y-1)}{H(x+1, y) - H(x-1, y)}\right) \quad (2)$$

其中, $H(x, y)$ 是像素点 (x, y) 的像素值。梯度直方图是一个加权直方图, 它的 bin 索引通过梯度的方向来计算, 而它的权值则通过梯度的幅值计算。梯度直方图通道计算公式为:

$$Q_\theta(x, y) = G(x, y) \times L[\Theta(x, y) = \theta] \quad (3)$$

式 3 中, $G(x, y)$ 和 $\Theta(x, y)$ 分别代表图像 (x, y) 像素处的梯度幅值和量化的梯度方向, L 是一个指示函数, θ 为梯度方向 $\alpha(x, y)$ 的量化范围, 本文选用 6 个方向的梯度直方图通道, 因此 θ 取值范围分别为 0-30 度、30-60 度、60-90 度、90-120 度、120-150 度、150-180 度。

训练积分通道特征检测器时, 选择特征池来训练弱分类器是关键, 构建恰当的特征池能提高检测性能^[11]。在训练和检测的时候提取太多的候选特征会导致内存负载过高, 而使用较少的候选特征会降低分类器性能, 引起检测性能的下降。在 128×64 像素的检测模板中, 多尺度检测时, 需要对图像反复缩放提取特征, 所有可能的矩形框的数量急剧增长, 降低了特征提取速度。同时, 积分通道特征检测算法随机选择候选特征包含了一定的随机因素。

为了减少候选矩形框的数量和避免随机选择的性能可变性, HOG+SVM^[2]算法中由于 HOG 特征使用单元和块的处理方法, 从而对图像局部像素点之间的关系得到很好的表征, 所以本文对特征池作了相应改进, 将 128×64 像素的检测框分别分割成 4×4 像素的单元, 每 2×2 个单元组成一个块, 并分别计算单元和块内像素和建立特征池, 由粗到细描述局部区域信息, 相比于原算法需要提取 718080 个候选块, 最终随机选取 30000 个矩形块作为特征。本文提取的特征只有 $(128 \times 64 \times 10 / 16) \times (1 + 1/4) = 6400$ 维, 而且对图像的每个位置都进行了特征描述, 避免了检测结果的可变性。

2.2 快速积分通道特征金字塔

滑动窗口检测法中, 为了检测大于检测模板的行人, 图像需要被缩小; 相反的, 为了检测小于检测模板的行人, 图像需要被放大。目前大部分检测方法所使用的梯度直方图、梯度幅值等特征都不具有尺度不变性。由于本文的积分通道特征不具有尺度不变性, 典型的方法是建立图像金字塔, 图像金字塔是一幅图像 I 的多尺度表示, 从底层开始, 每次将图像的宽/高缩小 $2^{(-1/8)}$ 得到更高的一层, 直到缩放后的图像尺寸小于检测模型 128×64 的尺寸, 需要对每个尺度 s 计算相应特征通道 $C_s = \Omega(R(I, s))$, 特征通道的重复计算, 耗费了大量时间。

本文采用文献^[12]提出的快速特征金字塔进行多尺度检测, 与典型方法在每个尺度都需要计算特征通道不同, 只计算尺度层 $s \in \{1, \frac{1}{2}, \frac{1}{4}\}$ 的特征通道 $C_s = \Omega(R(I, s))$, 其他中间层 C_s'' 用如下公式近似估计:

$$C_s'' \approx R(C_{s'}, s/s')(s/s')^{-\lambda_\Omega} \quad (4)$$

式(4)中, C_s'' 为待估计特征层, $s' \in \{1, \frac{1}{2}, \frac{1}{4}, \dots\}$ 为事先已经计算好的层中与 s 距离最近层的缩放因子, $R(C_{s'}, s/s')$ 表示将 $C_{s'}$ 的尺寸缩放为原来的 s/s' 倍, 常数 λ_Ω 取决于具体的特征通道。由 INRIA 训练集得到的颜色, 梯度幅值, 梯度方向特征的系数分别为 0, 0.1158, 0.1158。传统金字塔与快速金字塔的计算方式对比如图

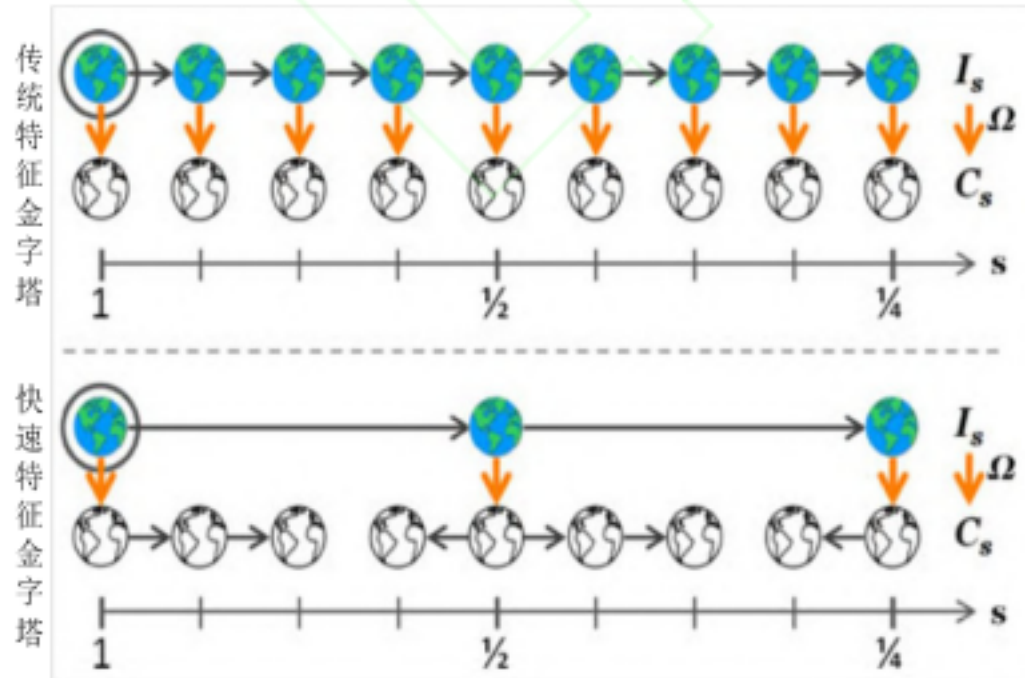


图1 特征金字塔对比图

通过近似计算特征金字塔可以显著的减少特征通道计算时间, 假设计算尺寸为 $n \times n$ 的图像的特征通道时间与其像素点个数成线性关系, 图像金字塔中每个尺度组所需计算尺度数为 m , 图像的宽/高每次缩小

$2^{1/m}$, j 为总尺度数, 则计算整个特征通道金字塔所需时间如下:

$$\sum_{j=0}^{\infty} n^2 2^{-2j/m} = n^2 \sum_{j=0}^{\infty} (4^{-1/m})^j = \frac{n^2}{1 - 4^{-1/m}} \approx \frac{mn^2}{\ln 4} \quad (5)$$

快速特征金字塔中实际每组只计算一个尺度, 所以这里 $m=1$, 时间消耗仅为 $\frac{4}{3}n^2$, 只比计算图像一个尺度下特征通道的时间多 33%。

2.3 软级联 Adaboost 分类器

一个由 K 个弱分类器级联组成的 Adaboost 分类器 H 形式如下:

$$H(x) = H_k(x) = \sum_{i=1}^k \alpha_i h_i(x) \quad (6)$$

式(6)中每个 h_i 都是一个弱分类器, 输出 -1 或者 1, α_i 是该弱分类器对应的权重, 取值范围为 $[0, 1]$ 。如果 $H(x) > 0$, 则分类为正样本, 反之则为负样本。经典的 Adaboost 分类器, 输入样本需要依次被每一个弱分类器判决, 将所有弱分类器判定的分数求和作为该样本最终检测分数, 该方法存在的问题是需要对所有弱分类器遍历, 分类速度较慢。本文采用文献^[13]用于人脸检测的软级联方法来优化分类速度。软级联 Adaboost 分类器中, 假设 s 为 k 个弱分类器分类后的累加分数和, 然后将得到的 s 与固定阈值 θ^* 比较, 如果 $s < \theta^*$, 则分类计算终止, 输出为非行人, 否则继续分类计算, 直到样本被所有弱分类器分类, 并且最终的得分 s 大于阈值 θ^* , 则输出为行人。根据文献^[14], 这里设定常量阈值 $\theta^* = -1$ 。

3 算法实现步骤

本文改进积分通道特征的快速多尺度行人检测的算法流程如图 2 所示:

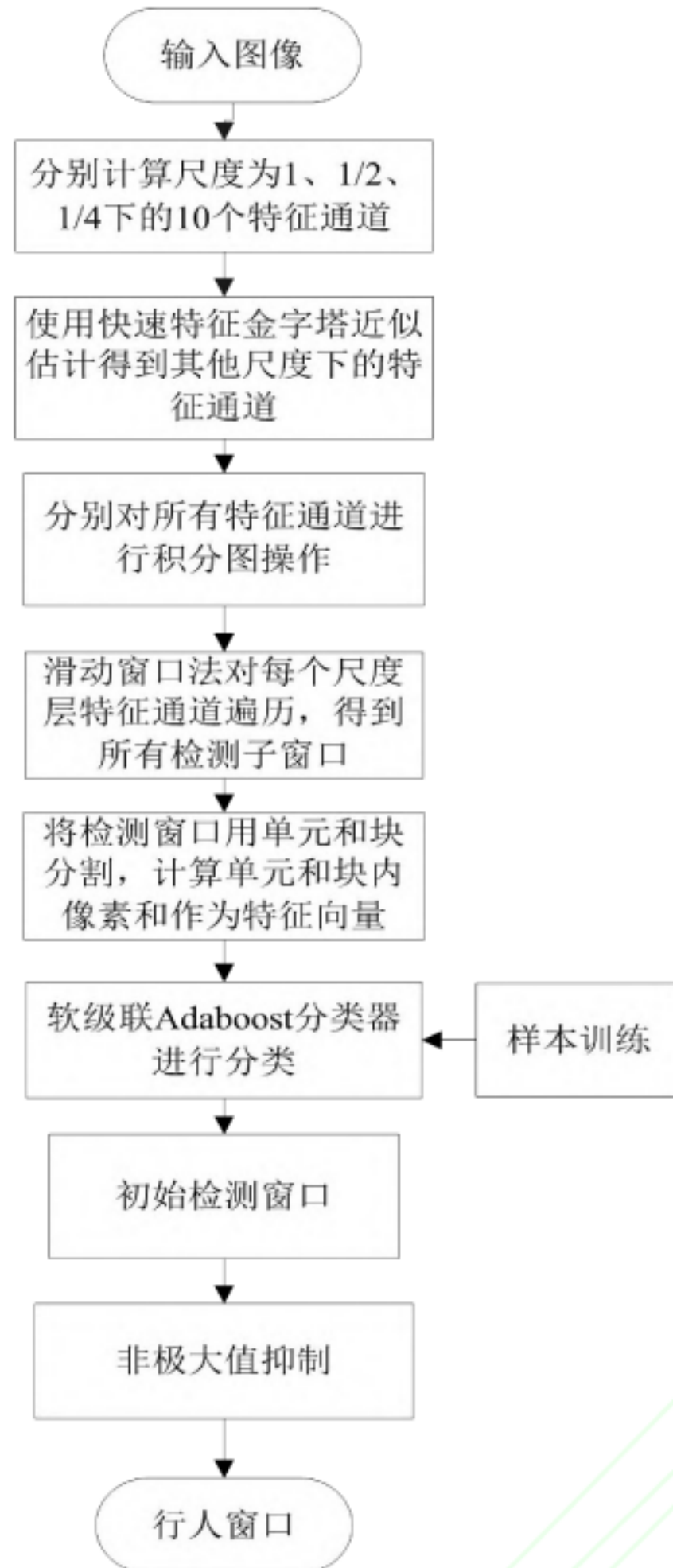


图2 本文算法流程图

具体步骤如下:

(1) 对大小为 640×480 像素的输入图像计算缩放因子为 1, 1/2, 1/4 的图像层, 使用 $[1\ 2\ 1]/4$ 模板对图像进行滤波预处理, 这样有助于去除噪声, 接着, 使用滤波后的图像分别计算这 3 个尺度下的 10 个特征通道;

(2) 分别用计算得到的 3 个尺度下的 10 个通道, 用快速特征金字塔近似估计得到其他尺度下的特征通道, 这里每 8 个尺度设定为一个尺度组, 从尺度 1 开始, 每次将图像的宽/高缩小 $2^{(-1/8)}$ 得到更高的一层, 直到缩放后的图像尺寸小于像素为 128×64 检测模型;

(3) 分别对所有尺度下的 10 个通道进行积分图操作, 利用积分图来提升求取积分通道特征的速度;

(4) 滑动窗口法用固定大小 128×64 像素的检测模板以步长为 4 像素对所有尺度下特征通道遍历, 同一尺度下同一位置 10 个

通道合并作为一个检测子窗口, 从而得到所有检测子窗口;

(5) 将检测子窗口的 10 个特征通道分别分割成 4×4 像素大小的单元, 每 2×2 个单元组成一个块, 分别求取单元和块内像素和作为特征向量, 每个检测子窗口的特征维度为

$$(128 \times 64 \times 10 / 16) \times (1 + 1/4) = 6400$$

维;

(6) 将 6400 维特征送入训练好的软级联 Adaboost 分类器进行行人与非行人的二分类, 得到检测窗口, 然后将各个尺度下的检测窗口位置映射到原图上对应位置, 得到初始检测窗口;

(7) 用非极大值抑制算法对原图上初始检测窗口进行后处理, 标记检测结果并输出。

4 仿真实验与结果分析

本文的实验分别在 INRIA 和 Caltech 行人数据库上进行。INRIA 包含 2416 个行人训练样本和 288 张含有行人的测试图片, 该行人库中行人的外观各异, 具有一定的复杂性。Caltech 数据库是目前规模较大的行人数据库, 由车载摄像机拍摄城市交通所得, 包含较小行人、光照、遮挡等复杂情况, 约 10 个小时左右, 视频分辨率为 640×480 , 标注了约 250000 帧, 350000 个矩形框, 2300 个行人, 另外对矩形框之间对应关系及遮挡情况进行了标注, 是公认规模较大, 挑战性较大的数据库。

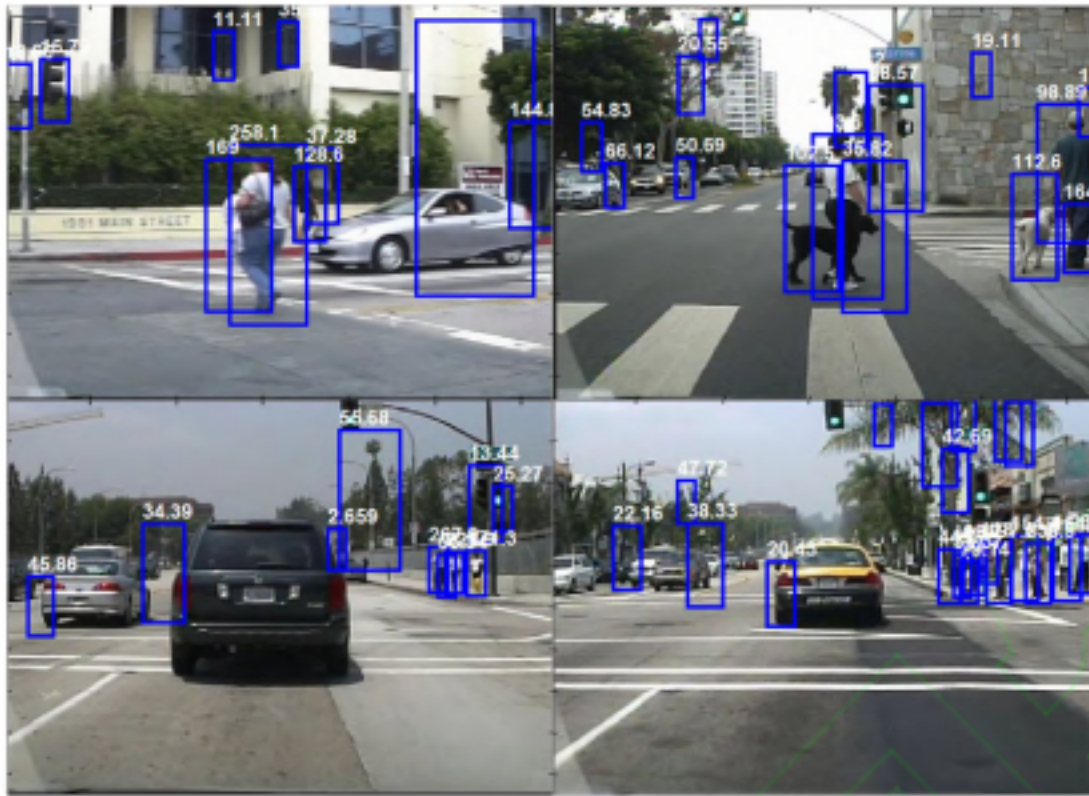
性能评价采用文献^[15]提出的漏检率 (Miss Rate) vs 每张图片平均虚警个数 (False Positives Per Image: FPPI) 曲线作为评价指标, 并以 0.1FPPI 作为参考点。实验所用软件环境为 Matlab R2012a, 系统环境为 Windows7, 内存为 8GB。

图 3 分别是 INRIA 和 Caltech 数据库上部分检测结果, 在 INRIA 数据库上, 在昏暗场景和行人重叠的情况下都取得较好的检测结果, 几乎没有误检窗口; 在 Caltech 数据库上, 对行人几乎没有漏检, 对整堆行人也取得了较好的检测效果, 但是出现了一些误检窗口, 由于是拍摄的实时路面视频, 车速较快, 视频分辨率不高, 背景相当复杂,

前方停靠车辆、树木、红绿灯柱等都会造成一些误检。



(a)INRIA 数据库部分检测结果



(b)Caltech 数据库部分检测结果

图 3 本文算法检测结果图

为客观评价本文方法, 分别就检测精度和检测速度两方面来分析。在 INRIA 数据库上将本文方法与 VJ[1]、HOG[2]、ChnFtrs[4]、FPDW[5]、VeryFast[6]、Roerei[7]和 InformedHaar[8]这七种方法作检测精度对比试验, 精度对比实验结果如图 4(a)。

图 4(a)显示, 在 0.1FPPI 参考点上本文算法检测精度达到了 80%, 跟 ChnFtrs 和 FPDW 算法相比, 分别取得了 2%和 1%的精度提升, 同时表明积分通道特征优于经典的 VJ 和 HOG 方法, 由于本文着重于多尺度检测下的速度提升, 与精心设计特征池来提升精度的 InformedHaar、Roerei 的方法以及训练多尺度检测模板提升检测速度的 VeryFast 方法相比精度有所不及。

在 Caltech 数据库上, 由于 VeryFast 方

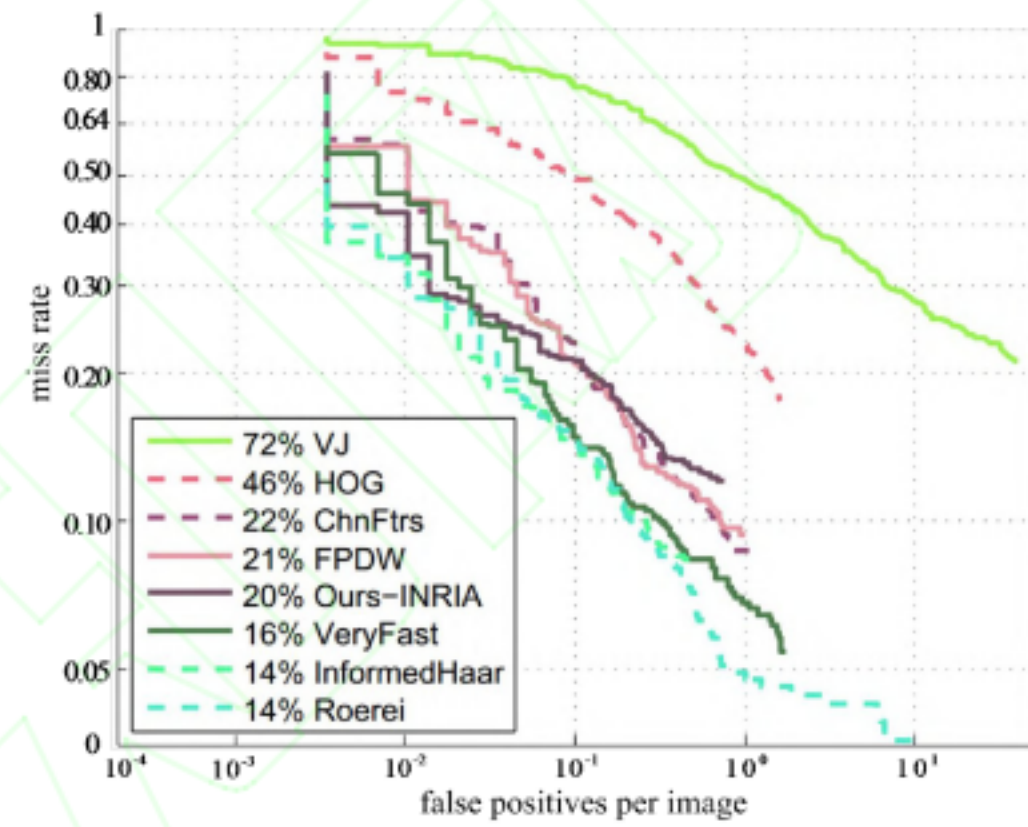
表 1: INRIA 数据库上 640×480 大小的图像上检测速度比较

方法	VJ[1]	HOG[2]	ChnFtrs[4]	FPDW[5]	VeryFast[6]	Roerei[7]	Informed Haar[8]	本文方法
帧/s(fps)	0.4	0.2	1.2	6.5	50	8.2	1.3	18.5

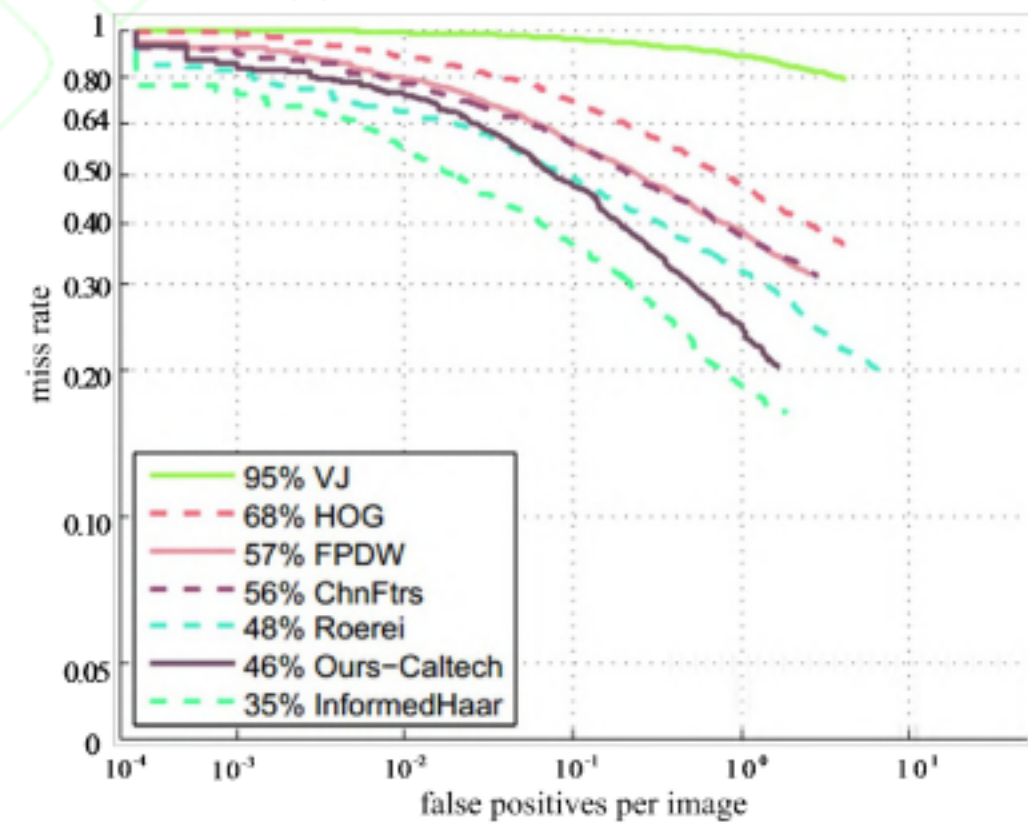
表 1 显示, 经典 VJ、HOG 算法检测速度较慢, 本文检测器速度与 ChnFtrs、FPDW、

法在 Caltech 数据库上训练所需的计算量较大, 至今没有学者在 Caltech 数据库上进行过相关实验分析, 所以这里不将此方法列入 Caltech 数据库对比方法中, 因此将本文方法与 V[1]、HOG[2]、ChnFtrs[4]、FPDW[5]、Roerei[7]和 InformedHaar[8]方法进行检测精度对比, 对比实验结果如图 4(b)。

图 4(b)显示, 在 0.1FPPI 参考点本文算法检测精度达到 54%, 跟 VJ、HOG、ChnFtrs、FPDW 和 Roerei 算法相比分别提高了 49%、22%、11%、10%和 2%, 对于注重精度提升将行人分为 3 个矩形模板的 InformedHaar 方法, 本文精度与之相比低了 9%。



(a)INRIA 数据库比较结果



(b)Caltech 数据库比较结果

图 4 不同算法检测精度对比图

为了验证本文算法的时间复杂度, 在 INRIA 数据库上用同一电脑将本文方法与其他 7 种方法进行检测速度对比试验, 实验结果如表 1:

Roerei 和 InformedHaar 算法相比, 分别取得了 15.4 倍、2.8 倍、2.2 倍和 14.2 的速度提

升, 在 640×480 大小的图像上达到 18.5fps 的检测速度, 满足了实时检测应用, 而通过 GPU 架构加速的 VeryFast 方法速度虽然较快, 但硬件配置较高, 限制了实际应用。通过对检测精度和速度分析得到, FPDW 方法相对于 ChnFtrs 方法, 速度提升了 5 倍左右, 而本文方法速度提升了 15.4 倍, 同时本文方法在 Caltech 数据库上检测精度相比 ChnFtrs 算法也取得了 10% 的提升, 着重提升检测速度的同时, 精度也得到相应提升。

5 结束语

本文针对积分通道特征算法特征池冗余信息较多, 多尺度下检测速度较慢的问题, 提出了改进积分通道特征的快速多尺度行人检测算法。该方法采用快速特征金字塔来实现快速多尺度特征计算, 同时减少了积分通道特征池的冗余性, 最后使用软级联 Adaboost 分类器进行快速分类。仿真实验结果表明, 相比于原积分通道特征算法和其他改进算法, 该算法不仅检测速度显著提高, 而且检测精度有所提升, 实现了实时行人检测。但是本文算法是基于行人整体的检测算法, 并没有考虑行人严重遮挡, 重叠等问题, 故对处在遮挡下行人的检测, 仍是需要进一步研究的内容。

参考文献

- [1] 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.
- [2] Dalal N and Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 886-893.
- [3] Ouyang W L and Wang X G. Joint deep learning for pedestrian detection[C]//Proceedings of the IEEE International Conference on Computer Vision, 2013: 2056-2063.
- [4] Ouyang W L, Zeng X Y and Wang X G. Single-pedestrian detection aided by two-pedestrian detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015: 1875-1889.
- [5] Benenson R, Omran M, Hosang J, et al. Ten years of pedestrian detection, what have we learned?[C]//European Conference on Computer Vision, 2014: 613-627.
- [6] Dollár P, Tu Z, Perona P, et al. Integral channel features[C]//Proceedings of British Machine Vision Conference, 2009, 91: 1-11.
- [7] Dollár P, Belongie S, and Perona P. The fastest pedestrian detector in the west[C]//Proceedings of British Machine Vision Conference, 2010, 68: 1-11.
- [8] Benenson R, Mathias M, Timofte R, et al. Pedestrian detection at 100 frames per second[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2903-2910.
- [9] Benenson R, Mathias M, Tuytelaars T, et al. Seeking the strongest rigid detector[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3666-3673.
- [10] Zhang S S, Bauckhage C and Cremers B. Informed haar-like features improved pedestrian detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 947-954.
- [11] Jia Y Q, Huang C, and Darrell T. Beyond spatial pyramids: receptive field learning for pooled image features[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 3370-3377.
- [12] Dollár P, Appel R, Belongie S, et al. Fast Feature Pyramids for Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1532-1545.
- [13] Bourdev L, Brandt J. Robust object detection via soft cascade[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005: 236-243.
- [14] Dollár P, Appel R, Kienzle W. Crosstalk cascades for frame-rate pedestrian detection[C]//European Conference on Computer Vision, 2012: 645-659.
- [15] Dollár P, Wojek C, Belongie S, et al. Pedestrian detection: An evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011: 743-761.