

基于 PHOG 特征的行人检测算法研究

包晓安, 朱晓芳, 张娜, 高春波, 胡玲玲, 桂江生

(浙江理工大学 信息学院, 杭州 310018)

摘要: HOG 特征对行人轮廓有很好的描述能力, 但基于 HOG 特征的行人检测存在检测速度慢、漏检率较高的问题, 使得该算法的实践应用范围受限; 针对检测速度慢、漏检率较高的问题, 提出了一种基于 PHOG 特征的行人检测算法; 首先, 提出了 PHOG 特征, 该特征对 cell 内的梯度特征进行强化, 增大了目标与背景的梯度分布区别, 从而使目标更容易被分类器学习和识别; 然后提出了构建特征金字塔的方法, 并对 PHOG 特征进行有效地降维, 大幅度减少了检测时间; 实验结果表明, 提出的 PHOG-PCA 特征将漏检率从 35% 降到了 22%, 检测速度也比一些流行算法快。

关键词: 行人检测; HOG 特征; PHOG 特征; PHOG-PCA 特征; 特征金字塔

Pedestrian Detection Algorithm Based on PHOG Feature

Bao Xiaolan, Zhu Xiaofang, Zhang Na, Gao Chunbo, Hu Lingling, Gui Jiangsheng

(School of Information, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: The HOG feature has a good description of the pedestrian profile, but the pedestrian detection based on the HOG feature has the problems of slow detection speed and high missed detection rate, which limits its practical application range. Aiming at the problem of slow detection speed and high undetected rate, this paper proposes a pedestrian detection algorithm based on PHOG features. Firstly, the PHOG feature is proposed. This feature enhances the gradient features in the cell and increases the gradient distribution of the target and the background, so that the target can be easily learned and identified by the classifier. Then, a method of constructing characteristic pyramid is proposed, and the PHOG features are effectively reduced in dimension, which greatly reduces the detection time. The experimental results show that the proposed PHOG-PCA feature reduces the missing detection rate from 35% to 22% and the detection speed is faster than some popular algorithms.

Keywords: pedestrian detection; HOG feature; PHOG feature; PHOG-PCA feature; feature pyramid

0 引言

行人检测技术是智能视频监控领域一种重要的技术, 它是行人跟踪、行为分析等一系列技术实现的前提。行人检测技术可以广泛用于自动驾驶^[1]、机器人视觉^[2]、智能监控^[3]等应用中。行人检测算法自研究以来, 一直是研究热点, 吸引了无数的爱好者研究。其检测的主要内容是使用计算机识别出对视频中或者静态图像中行人, 并输出相应的位置信息。行人检测算法的主要内容是在检测过程中, 提取出行人的显著特征, 用于分类器判别, 从而输出行人的判别结果。

近年, 涌现了很多行人检测算法, 其中基于机器学习的算法表现较好。这类算法主要涉及到分类器和特征描述子的选择。2005 年, Dalal^[4] 等人提出 HOG (Histogram of

oriented gradients) 特征, 该特征因为很好的稳定性和描述能力而被广泛的研究与应用。HOG 算法的特征描述子是对行人轮廓梯度分布的一种描述, 该特征描述对局部形变和光照保持较好的鲁棒性。在 HOG 算法的基础上, 提出了很多改进算法^[5-7], 典型的有 DPM^[8] 算法。DPM 算法是 Felzenszwalb 等人提出的, 该算法将行人分解成好几部分再基于不同分辨率进行检测, 并增加了 HOG 特征缺少的颜色信息, 一定程度上提高了行人检测准确率。还有一些行人检测的改进算法, 例如基于积分通道^[9]等算法, 有效加快了行人检测的速度。其它使用较多的行人检测算法有 HOG+LBP^[10], Haar 特征^[11], 还有基于深度学习的算法^[13-14]。这些算法都存在检测速度较慢或者漏检率较高等问题。

由于行人的非刚性及环境复杂等原因, 使得行人检测存在漏检、实时性差等问题。针对 HOG 特征检测中速度过慢, 漏检率较高的问题, 提出一种基于 PHOG 特征的行人检测算法。该算法解决了以下问题: 1) 降低了漏检率。通过提出 PHOG 特征, 加强了对局部特征的描述能力, 增大了目标和背景的差异; 2) 减少了检测时间。通过构建特征图的金字塔, 从而避免了对多层提取特征, 减少了计算量。并且进行了 PCA 降维, 降低了特征复杂度。

1 HOG 特征原理

HOG 特征的提取方法是: 先灰度化、归一化图像, 计

收稿日期:2017-12-09; 修回日期:2018-01-05。

基金项目: 国家自然科学基金(61379036, 61502430); 国家自然科学基金委中丹合作项目(61361136002); 浙江省重大科技专项重点工业项目(2014C01047); 浙江理工大学 521 人才培养计划(20150428)。

作者简介: 包晓安(1973-), 男, 浙江东阳人, 硕士, 教授, 主要从事计算机视觉、智能信息处理方向的研究。

桂江生(1978-), 男, 安徽安庆人, 博士, 副教授, 主要从事视频图片处理, 模式识别方向的研究。

算图像的方向梯度特征, 再将图像划分成多个 cell 和 block。最后串联窗口里所有的 block 的特征向量, 就得到了该窗口的 HOG 特征描述子。用训练好的行人分类器对该描述子进行判别, 再输出判别结果。特征提取过程中, 对 cell 里的像素梯度进行三线性插值, 每个 cell 提取出一个 9 维的直方图。用该 9 维直方图代表这个 cell 的梯度特征, 再串联每个 block 里 cell 的直方图特征, 得到一个高维的 HOG 特征描述子, 该描述子就是窗口的 HOG 特征。基于 HOG 特征的行人检测试验结果如图 1 所示。

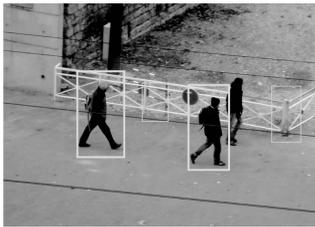


图 1 基于 HOG 特征的行人检测

由图 1 可知, 基于 HOG 特征的行人检测存在漏检和误检的情况。误检如图中细方框, 漏检如图中未标记出的行人。分析产生误检的根本原因, 发现是该 HOG 特征对行人的描述子总体上偏向于直立目标的特征, 而忽略了行人特有的一些肢体分布特征。产生漏检的原因是, 在提取 HOG 特征时, 该特征对行人目标的描述子较弱, 目标特征与背景的差异不大, 故产生漏检。针对这些问题问题, 本文提出 PHOG 特征。

2 本文算法

2.1 PHOG 特征的提取

PHOG 特征的主要思想是调整方向梯度直方图的对比度, 即增强对行人梯度分布的描述能力, 使背景和行人目标的梯度差异更大, 从而减少误检和漏检的产生。PHOG 特征主要的具体提取过程如下:

- 1) 先计算图像里每个像素的梯度信息, 包括方向和大小。
- 2) 将图像划分成多个 cell 区域。并用含有 9 个方向的梯度直方图来统计每个 cell 里像素的梯度幅值, 生成每个 cell 的特征描述子 (descriptor)。
- 3) 对每个 cell 的方向梯度直方图进行调整, 得到 PHOG 特征。根据公式 (1) 和 (2) 进行调整, 使该 cell 的梯度直方图整体对比度增大, 如图 2 所示。

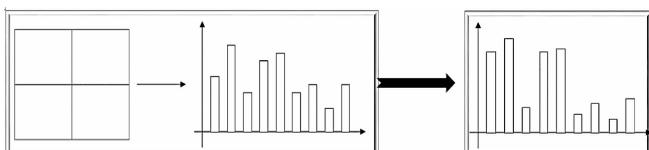


图 2 直方图对比度增强示意图

- 4) 将多个 cell 组合成一个 block 区域, 将 block 区域内 cell 的 descriptor 串联起来, 生成 block 区域的 descriptor。

5) 最后将窗口中所有 block 区域的 descriptor 串联起来, 得到整个滑动窗口的特征描述子。该特征描述子可用于训练分类器。

步骤 3) 中, 为了增强特征的局部表现力, 对每个 cell 的方向直方图进行调整, 增大每个 cell 直方图的方差 (即梯度值大的更大, 梯度值小的更小)。根据判别函数来调整 cell 的直方图特征, 具体的判别函数如式 (1)、(2) 所示:

$$g_{s, Avg} = \frac{\sum_{i=1}^b g_s(x, y)}{b} \quad (1)$$

$$g'_s = \begin{cases} a1 \cdot g_s(x, y), & g_s(x, y) \geq g_{s, Avg} \\ a2 \cdot g_s(x, y), & g_s(x, y) < g_{s, Avg} \end{cases} \quad (2)$$

$g_s(x, y)$ 是该 cell 内的 (x, y) 位置的梯度幅值, 该 cell 属于 s 这个 block。 b 是 cell 内的像素数目。 $a1$ 和 $a2$ 分别是调整系数, 要求 $a1 \geq a2$, 且 $a1 > 1.0$, $a2 > 0$ 。当 $a1 > a2$ 时, 就会将梯度值大于均值的梯度以 $a1$ 比例增大, 梯度值小于均值的像素梯度会以 $a2$ 的比例减小; 经过上述对梯度幅值的调整, 可以增大每个 cell 直方图的方差, 使 cell 内的差异更突出。本文实验中 $a1 = 1.5$, $a2 = 0.8$ 。特征增强过程如图 2 所示。

由图 2 可以看出, 该直方图纵坐标值小于均值的 bin 均被压缩, 大于均值的均被放大。所以整个直方图的对比度被显著增强, 方差也被增大, 对该 cell 内部特征的描述力更强。根据式子 (1) 和 (2) 进行了 cell 特征的调整, block 大小为 2×2 cell。选取其中一块 block 特征, 进行特征调整, 调整前后的 block 特征进行统计对比, 如图 3 所示。对调整前后的特征进行数值分析, 如图 4 所示。

0.109390035	0.128416985	0.105480708	0.145890296	0.187046096	0.240481704
0.068987638	0.090419948	0.088513456	0.139790714	0.092932925	0.10127914
0.027062574	0.056222029	0.080699913	0.074579246	0.060153868	0.162770808
0.181248769	0.161556572	0.166185707	0.241056442	0.272724807	0.272724807
0.077869363	0.155736417	0.077146627	0.272724807	0.272724807	0.250595123
0.027840663	0.142078534	0.153211534	0.16798158	0.223634362	0.272724807



0.164085053	0.192625478	0.158221062	0.218835444	0.280569144	0.192385363
0.05519011	0.135629922	0.070810765	0.111832571	0.07434634	0.081023312
0.021600059	0.044977623	0.06455993	0.059663397	0.048123094	0.244156212
0.271873154	0.242334858	0.249278561	0.192845154	0.409087211	0.409087211
0.06229549	0.124589134	0.061171302	0.409087211	0.409087211	0.200476098
0.02227253	0.113662827	0.122569227	0.134385264	0.17890749	0.498087211

图 3 调整前后的 block 特征值

观察图 4 可知, 在 $0 \sim 38$ bin 的范围内, block-PHOG 特征比 block-HOG 特征局部对比度更大, 但整体趋势保持一致。block 内特征的方差由 $0.063\ 414\ 083$ 增大到 $0.098\ 959\ 88$, 提高了大约 3.5% 。对该 block 所在的窗口提取 PHOG 特征, 将其可视化展现, 并与 HOG 特征进行对比, 如图 5 所示。

观察图 5 可知, 不仅是行人目标的梯度分布特征被增

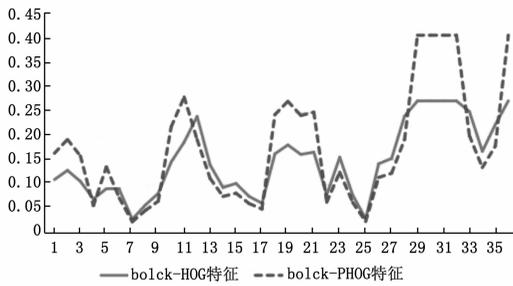


图 4 block 特征

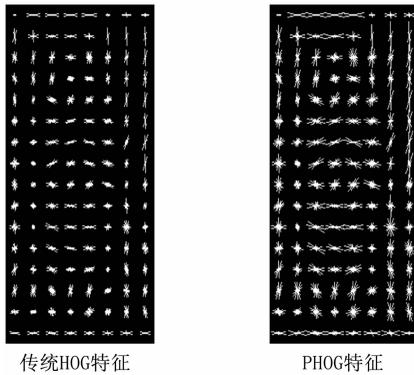


图 5 传统 HOG 特征与 PHOG 特征对比图

强，背景的特征也被增强，从而目标与背景的整体差异被增大。

上述基于 PHOG 特征检测的详细步骤如下：

1) 预处理部分。对图像进行灰度化，再利用 gamma 变换降低噪声和光照的干扰。gamma 压缩如公式 (3) 所示。实验中 $\gamma = 1/3$ 。

$$I(x, y) = I(x, y)^{\gamma} \quad (3)$$

2) 根据公式 (4) 和 (5) 计算像素的梯度。

$$g(x, y)_x = I(x + 1, y) - I(x - 1, y) \quad (4)$$

$$g(x, y)_y = I(x, y + 1) - I(x, y - 1) \quad (5)$$

其中： $I(x, y)$ 是像素灰度值， $g(x, y)_x$ 和 $g(x, y)_y$ 分别是该像素点处的 x 方向和 y 方向的梯度。该像素点处的梯度幅值和方向分别由式 (6) 和 (7) 计算。 $\partial(x, y)$ 是该点的梯度方向。本文实验水平、垂直梯度算子取 $[-1, 0, 1]$ 、 $[-1, 0, 1]^T$ 。

$$g(x, y) = \sqrt{g(x, y)_x^2 + g(x, y)_y^2} \quad (6)$$

$$\partial(x, y) = \tan^{-1}\left(\frac{g(x, y)_y}{g(x, y)_x}\right) \quad (7)$$

3) 构建 cell 梯度方向直方图。

将 cell 内像素的梯度插值，再投影到 9 维的直方图，用来描述 cell 内的梯度分布特征。其中每个 cell 内的梯度方向规定分为 9 个 (0~180 度)，每个 bin 对应着该方向的梯度幅值加权。考虑到部分像素的梯度方向处于两个 bin 的临界区域附近，则必须计算对相邻区域的影响。所以对该像素梯度方向的相邻区域进行投影，最后再叠加该像素在相邻区域的投影值，获得综合的梯度幅值。将该幅值作为直

方图的纵坐标。

4) 归一化 block 值。

由于光照不均匀或者噪声的影响，导致某些 block 区域与周围的区域差异很大，生成的 HOG 特征变化较为剧烈。这种 HOG 特征训练得到的分类器的泛化能力会大大降低。所以为了增加该特征的鲁棒性，对 block 区域内的像素梯度进行归一化处理，减弱局部剧烈特征对总体特征的不良影响。本文采用的是 L2-Norm 进行归一化，如式 (8)。 ϵ 是一个很小的常数值，本文实验 $\epsilon = 0.23$ 。

$$L2-Norm: v \rightarrow v / \sqrt{\epsilon^2 + \|v\|_2^2} \quad (8)$$

根据上述检测步骤，PHOG 特征提取过程如图 6 所示。

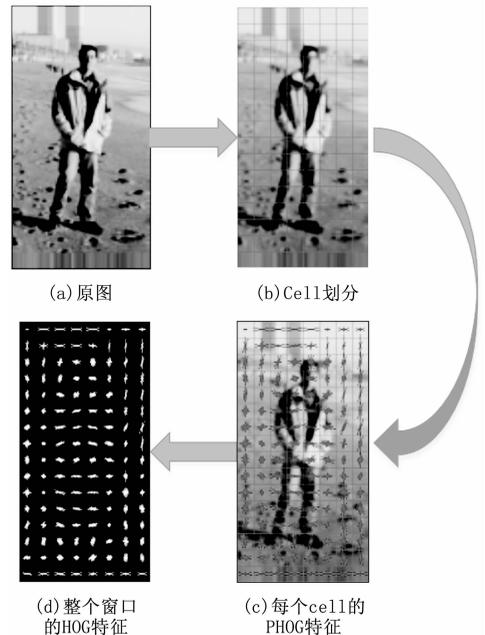


图 6 PHOG 特征提取示意图

本文实验使用的固定滑动窗口，大小为 64×128 ，cell 大小为 16×16 ，block 大小为 2×2 的 cell，扫描步长设置为 8 个像素，所以该窗口的 HOG 特征向量的大小为 $9 \times 4 \times 7 \times 15 = 3780$ 。由于该特征维数较高，本文在后面章节会进行 PCA 降维。

3 基于 PHOG 特征的行人检测

基于 PHOG 特征的行人检测流程如图 7 所示。

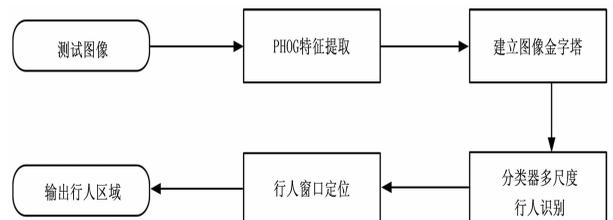


图 7 基于 PHOG 特征的行人检测

对检测图像先进行预处理后提取原图像 PHOG 特征，得到该图像的特征图如图 8 所示。观察图 8 可看出，

PHOG 特征图相比于传统的 HOG 特征, 对行人目标梯度分布的刻画更强。在该实验原图的特征图中, 传统的 HOG 特征图对行人的描述几乎融于背景, 观察不出行人的位置, 而 PHOG 特征可以大致描述出目标的位置及周边环境的特征。

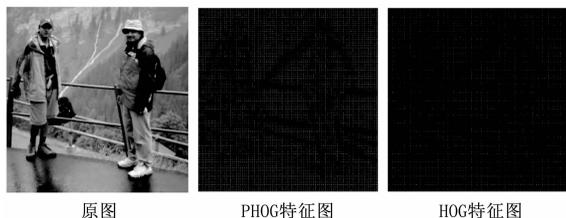


图 8 PHOG 和 HOG 特征对比

3.1 PHOG 特征金字塔

在提完 PHOG 特征后, 为了加快检测速度 (在不降低检测精度的前提下), 本文提出对 PHOG 特征图构建 8 层特征金字塔, 替代传统的构建原始图片的金字塔。特征金字塔构建的具体步骤是:

先获取第 $i=0$ 层 (最底层) 特征图, 该特征图即是原图的 PHOG 特征图。获取第 $i=i+1$ 层的特征图。对第 i 层特征图进行抽样, 抽样长度根据 2:1。即对第 i 层特征图的一个 block 内的 36 维 PHOG 特征抽样, 得到第 $i+1$ 层一个 cell 内的 9 维的 PHOG 特征。抽样计算的公式如式 (9)。

重复步骤 2, 直到 $i=8$, 获取完整的特征金字塔。

$$d_{i+1,j} = \begin{cases} dmax_{i,k}, & 0 < k \leq 5 \\ dmax_{i,k}, & 32 < k \leq 36 \end{cases} \quad (9)$$

$k = 1, 2, \dots, 36$

式 (9) 的原理是计算 36 维特征数据中, 纵坐标值最大的前 5 个 bin 值和纵坐标值最小的倒数 4 个 bin 值。分别记录这些 bin 的横、纵坐标值, 组合得到新一层的 9 维 cell 特征。 $dmax_{i,k}$ 是第 i 层的特征图内 block 的特征值排序 (降序) 函数, k 是该函数的第 k 个值。 $d_{i+1,j}$ 是第 $i+1$ 层的第 j 个 cell 的特征值。根据上述步骤构建的 PHOG 特征金字塔如图 9 所示。



图 9 PHOG 特征金字塔示意图

构建的 PHOG 特征金字塔如图 9 所示。由于是对特征图进行向下采样获得的特征金字塔, 该计算过程中只用到了简单的抽样等计算, 相比于传统的 HOG 图像金字塔每层都要重新计算 HOG 特征, 少了大量的计算量。该特征金字塔构建完成后, 每层的滑动窗口都得到一个 PHOG 特征。整体的检测时间都减少了很多。

3.2 PHOG-PCA 特征

如上节所述, 每个窗口的都得到一个 PHOG 特征。但

该特征维数高达 3 780 维, 为了加快检测速度, 本文对 PHOG 特征进行有效地 PCA^[15] (Principle Component Analysis) 降维, 得到 PHOG-PCA 特征。降维实验数据统计如表 1 所示。

由表 1 可知, 本文基于 PHOG 特征的 PCA 降维实验中, 当主成分维数 $N=180$ 时, 有保持较高的识别率。当特征维数 N 在 180 附近时, 检测的识别率都有所下降。所以本文降维后的 PHOG 特征取前 180 个主成分, 将该 180 维的 PHOG 特征记为 PHOG-PCA 特征。其中, 该降维实验中用到的行人分类器 SVM 是基于 INRIA 数据集训练得到的。在检测到行人目标后, 再进行窗口融合, 就获得了精确的行人目标位置。

表 1 PHOG 特征降维实验

PHOG-PCA 特征 主成分维数	支持向量数 数量	测试样本 检测率/%
$N_2=1\ 000$	820	93.58
$N_2=500$	657	93.20
$N_2=200$	536	93.31
$N_2=180$	494	94.35
$N_2=150$	503	93.90
$N_2=120$	488	93.96
$N_2=100$	468	93.88

4 实验设计及分析

4.1 实验设置

本文实验采用 INRIA 数据集, 该数据集背景种类较多, 行人目标行为变化多样, 对于本文算法的测试具有很好的挑战性。INRIA 数据集含有训练的正、负样本和测试样本。训练的正、负样本分别有 2416 和 1218。测试集有正样本 568, 负样本 462 个。本文算法实验的硬件的运行环境设置为 Intel (R) Core (TM) i3-2410M CPU, 4 G 内存的笔记本。实验采用 PHOG-PCA+SVM 的检测结构进行检测。

4.2 SVM 训练

本文行人检测基于的是线性 SVM (Support Vector Machine) 分类器。为提升分类器的泛化能力, 本文对负样本进行扩充。通过对 INRIA 负样本图片进行随机窗口采样, 获取到 11120 张训练负样本。将该训练样本的 PHOG-PCA 特征集, 用于进行 SVM 交叉训练。本文将训练负样本分成 4 组, 不断获取新的难例, 来提升 SVM 分类器的泛化性能。本文实验基于该 SVM 分类器进行行人检测实验。

4.3 实验结果及分析

本实验参数: 滑动窗口固定大小为 64×128 , cell 大小为 16×16 像素, block 由 2×2 个 cell 组成, 提取的 PHOG 特征为 3780 维。再结合 Objectness 特征, 得到 O-PHOG 特征, 再经过 PCA 降维, 得到 196 维的 PHOG-PCA 特征向量。

为验证本文 PHOG 特征对窗口内行人的描述能力, 将

PHOG 特征、PHOG-PCA 特征和其他几种常用行人特征进行检测试验, 实验数据如表 2 所示。

表 2 各种特征的行人识别率试验结果 %

特征/实验次数	1	2	3	4	5	Average
LBP	63.8	64.6	63.7	61.2	63.9	63.44
Haar	59.1	61.7	63.8	60.4	60.7	61.14
HOG	75.4	76.3	76.9	77.1	78.1	76.76
LBP-HOG	84.7	84.2	83.8	84.3	84.5	84.30
PHOG	87.9	88.0	87.8	87.4	87.6	87.74
PHOG-PCA	87.6	86.9	86.7	86.8	86.8	86.96

由表 2 可知, 本文基于多种特征分别进行了 5 次实验, 并进行了均值统计。相比于传统的 HOG 特征, LBP 特征和 Haar 特征训练得到的分类器识别率较低, 识别率均值分别为 63.44%, 61.14%。HOG 特征的识别率最高为 78.1%, 均值为 76.76%。LBP-HOG 的联合特征是的检测率较传统 HOG 有一定程度的提高。本文提出的 PHOG 特征识别率高于传统的 HOG 特征, 5 次实验中识别率最高的为 86.8%, 均值也为 86.69%。PHOG-PCA 特征和 PHOG 数值相差不大。

为反映本文特征与漏检率的关系, 将实验数据统计如图 10, 该图反映了行人检测漏检率 (Miss rate) 和每个窗口的误检率 (False positive per window) 之间的关系。由该图可看出当误检率一定时, 本文 PHOG 特征、PHOG-PCA 特征对窗口内目标的描述力最强, 所以漏检率比其余特征都低。

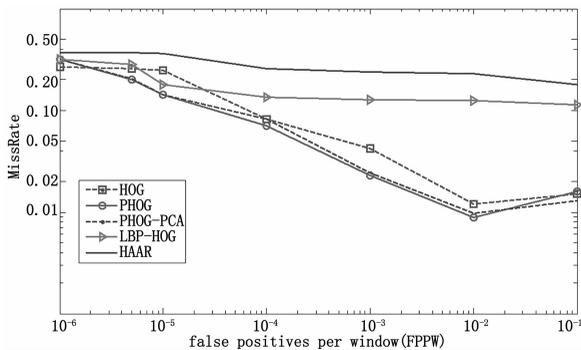


图 10 各特征检测的漏检率

为验证本文提出的特征在整幅图片中的漏检率, 进行了实验与数据统计, 如图 11。观察图 11, PHOG 特征和 PHOG-PCA 特征的漏检率整体偏低。相比于传统 HOG 检测, 在 FPPI 为 10-2 时, 本文算法将漏检率从 35% 降到了 22%。分析原因是 PHOG 特征不仅增强了对行人目标的描述, 而且也突出了背景的特征, 进而增大了两者间的差异, 使分类器更易识别目标。实验检测结果如图 12 所示。

本文提出的基于 PHOG-PCA 特征的行人检测实验结果, 如图 12 所示。由该图可以看出, 该算法能准确地检测出图片中的行人, 几乎没有漏检。

为直观显示本文算法检测速度, 将 INRIA 测试集尺寸

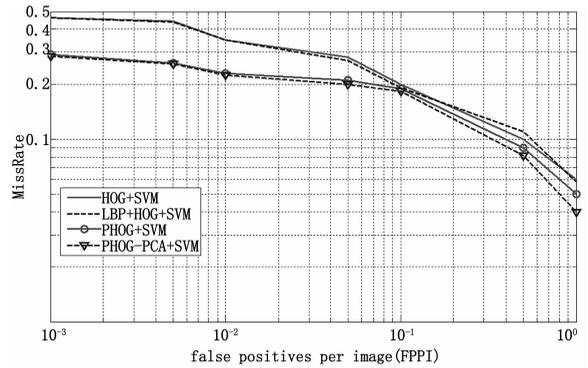


图 11 各特征实验对比



图 12 基于 PHOG-PCA 特征的检测实验结果图

缩减为 192×256, 在该测试集上进行检测, 平均检测时间统计见表 3。由表 3 可看出, 本文提出的算法检测速度较快, 与其余算法时间相差较多, 存在很大优势。

表 3 各算法检测时间比较

Method	PHOG-PCA	PHOG	HOG	HOG-LBP	HAAR	RCNN
Time/ms	45	100	102	60	75	130

5 结论

本文提出了一种基于 PHOG 特征的行人检测算法, 针对传统 HOG 特征的行人检测中存在较高漏检和检测速度慢的问题, 提出了解决方法。首先提出了 PHOG 特征, 该特征增强了对目标和背景的描述, 有效降低了漏检率。再通过构建特征金字塔及 PCA 降维, 减少了计算量, 进而加快了行人检测的速度。从实验结果可以看出, 本文算法在行人检测中有较大优势。本文下一步的研究计划是研究行人检测中存在的遮挡问题, 进一步提高行人检测的准确率。

参考文献:

[1] 王国华, 刘 琼, 庄家俊. 基于局部特征的车载红外行人检测方法研究 [J]. 电子学报, 2015, 43 (7): 1444-1448.

[2] 黄志杰, 杨广柱. 基于机器视觉的车辆行人检测系统技术综述 [J]. 数字技术与应用, 2017 (2): 115-116.

[3] 孙育霖. 智能视频监控中行人检测及外貌特征分析关键技术 [D]. 济南: 山东大学, 2016.