

组合 SVM 分类器在行人检测中的研究

邹冲^{1,2} 蔡敦波^{1,2} 刘莹^{1,2} 赵娜^{1,2} 赵彤洲^{1,2,3}

(武汉工程大学智能机器人湖北省重点实验室 武汉 430205)¹

(武汉工程大学计算机科学与工程学院 武汉 430205)² (华中科技大学自动化学院 武汉 430074)³

摘要 在基于 HOG 特征的 SVM 行人检测算法的基础上,提出了组合分类器的改进算法。该算法首先采用多尺度滑动窗口提取 HOG 特征,并对单个 SVM 分别进行训练,再将训练好的 SVM 分别采用串联、并联结构形成新分类器后对行人进行检测。为解决用多尺度滑动窗口提取特征时产生的目标候选区域重叠问题,采用非极大值抑制算法对重叠区域进行融合,进而得到准确候选区。实验表明,组合的 SVM 分类器可以有效降低误检率和漏检率。

关键词 行人检测, HOG, SVM, NMS, 组合分类器

中图分类号 TP391.41 文献标识码 A

Research of Combination SVM Classifier in Pedestrian Detection

ZOU Chong^{1,2} CAI Dun-bo^{1,2} LIU Ying^{1,2} ZHAO Na^{1,2} ZHAO Tong-zhou^{1,2,3}

(Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430205, China)¹

(School of Computer Science & Engineering, Wuhan Institute of Technology, Wuhan 430205, China)²

(School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China)³

Abstract On the basis of histogram of oriented gradient and support vector machine(HOG-SVM)algorithm, this paper proposed an improved algorithm for combination classifiers. Firstly, This algorithm uses multi-scale sliding windows to extract the HOG features and trains SVM separately. Then, the trained SVM which is formed to a new classifier in series or parallel is used to detect pedestrian. In order to solve the problem that the target area is overlapped when features are extracted in multi-scale sliding windows, the non-maximum suppression (NMS) algorithm is used to fuse the rectangles and to get exact candidate region. Experiments show that combined SVM classifiers can effectively reduce the false detection rate and missed rate.

Keywords Pedestrian detection, Histogram of oriented gradient(HOG), Support vector machine(SVM), Non-maximum suppression(NMS), Combination classifiers

行人检测是指从图片或者视频中将复杂场景中的行人目标从背景中分割出来并精确定位。行人检测技术被广泛应用于基于航拍图像的行人检测^[1]、家庭服务机器人^[2]、Google 街景中行人消除^[3]等领域,且随着信息智能化的发展,对行人检测的实时性及准确率有更高的要求。基于手工标注的传统方法无法满足海量数据的行人检测要求,在依靠计算机视觉技术检测时,由于受到场景复杂、行人的非刚性特征等客观因素的制约,在检测准确率上未能取得令人满意的效果。因此,行人检测一直以来都是计算机视觉领域的难点和研究重点。

Oren^[4]最早提出了“Haar 小波特征+支持向量机”的行人检测方法。虽然目前许多行人特征的检测性能优于小波特征的检测性能,但 Oren 的主要贡献在于将机器学习的思想引入行人检测中^[5]。为了更好地描述行人轮廓特征, Dalal 等^[6]提出了基于梯度方向直方统计图特征(Histogram of Oriented

Gradient, HOG)的行人检测方法,该方法的提出将行人检测技术的研究推向了新的高潮,并成为行人检测领域的里程碑,以至于后来很多行人检测算法都是基于 HOG+SVM 思想做出的改进。然而, Dalal 提出 HOG+SVM 行人检测方法是基于线性核 SVM 实现的,尽管该算法的速度可以满足实时性要求,但其识别率往往不如人意。为了提高 SVM 分类器的性能, Maji 等^[7]提出了基于直方统计图交叉核支持向量机(Histogram Intersection Kernel SVM, HIKSVM)的行人检测方法,使得 HIKSVM 在检测性能上接近于非线性 SVM,同时检测速度接近于线性 SVM。然而使用单一分类器往往容易出现误检(本该是背景区域检测为行人)、漏检(本该是行人区域检测为背景)等情况,而且错误在后继处理中很难得到纠正。后来, Dolla P 等人^[8]提出 ACF(Aggregated Channel Features)算法,将图片中的 HOG+LUV(一种颜色空间标

本文受国家自然科学基金资助项目(61103136), 武汉工程大学创新基金资助项目(CX2015057)资助。

邹冲(1992—),男,硕士生,主要研究方向为机器视觉、智能计算, E-mail: witzouchong@126.com; 蔡敦波(1981—),男,博士,副教授,主要研究方向为智能规划与自动推理、机器视觉(通信作者); 刘莹(1994—),女,硕士生,主要研究方向为模式识别、数据挖掘; 赵娜(1992—),女,硕士生,主要研究方向为机器视觉、智能计算; 赵彤洲(1970—),女,硕士,副教授,主要研究方向为模式识别、图像处理。

准)特征带入级联 boosting 分类器中做分级过滤,有效避免了因使用单一分类器而产生误检、漏检等情况,提高了算法的识别率。

尽管不断有新的方法被提出,但行人检测的识别率与识别速度一直没有达到令人满意的效果。由于行人具有非刚性特征^[9],在行人检测过程中常受到衣着颜色、尺寸大小、姿态多样化、复杂背景等影响^[10],提取特征分布不紧凑而导致训练后分类器的性能不稳定,因此检测过程中经常出现漏检、误检等情况。而在一次检测中仅仅使用单一分类器将很难适用各类检测情况,本文提出一种组合 SVM 分类器方法,其可在一定程度上提高检测率。算法首先对单一分类器 SVM1 与 SVM2 分别训练,然后再进行组合训练。当 SVM1 和 SVM2 都对同一个目标区域做检测时,即为并联组合;当 SVM1 和 SVM2 对同一个目标区域采用分级过滤检测时,即为串联组合。实验表明,组合的 SVM 分类器可以有效降低误检率和漏检率。

1 多尺度滑动窗口的 HOG 特征提取

HOG 特征由局部区域的梯度方向直方图构成,是典型的行人检测特征。由于图像局部目标的外观特征能够很好地被梯度或边缘方向密度分布描述,因此 HOG 特征常被用于描述人体轮廓。

1.1 空间颜色标准化

在提取 HOG 特征之前,为减小光照因素对行人检测的影响,提高检测器的鲁棒性,需要对颜色空间归一化。经验证^[6],平方根标准化能够有效抑制噪声的影响。平方根压缩公式为 $I(x,y) \leftarrow I(x,y)^{0.5}$,其中 $I(x,y)$ 表示图像 $I(x,y)$ 在点 (x,y) 处的像素值。

1.2 计算图像梯度

图像梯度计算是为了获取轮廓信息,同时也可以进一步弱化光照的影响^[11]。Dalal 证明使用简单一维离散微分模板 $[-1,0,1]$ 最有效,根据该模板,可以计算出该像素在水平方向梯度与垂直方向梯度的幅值,再根据水平方向梯度幅值 $G_x(x,y)$ 与垂直方向梯度幅值 $G_y(x,y)$ 计算在点 (x,y) 处的梯度幅值 $G(x,y)$ 以及梯度方向 $\alpha(x,y)$,如式(1)所示:

$$\begin{aligned} G(x,y) &= \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \\ \alpha(x,y) &= \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \end{aligned} \quad (1)$$

1.3 统计空间和方向上权重获取 HOG 特征

以滑动窗口的方法统计每一块(block)中的 HOG 特征,并将获取到的 HOG 特征做串联组合后作为该块的特征向量,再将特征向量做归一化处理(弱化 HOG 特征对光照强弱、边缘变化的敏感性)。归一化过程如式(2)所示:

$$G(x,y) \leftarrow G(x,y) / \sqrt{G(x,y)^2 + \epsilon^2} \quad (2)$$

其中, ϵ 是一个趋于 0 的常数,用于防止除数为 0。

将归一化后的 HOG 特征向量送入 SVM 分类器计算置信度,根据置信度大小及正负号来判定该区域是否存在行人。

提取 HOG 特征后,图片显示及 HOG 特征统计图如图 1 所示。

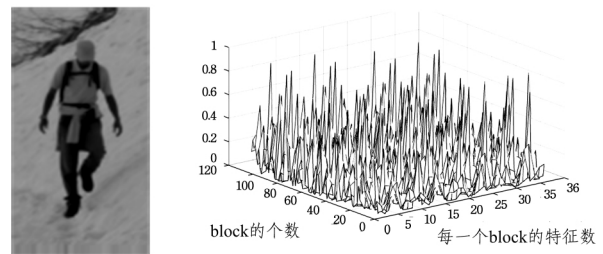
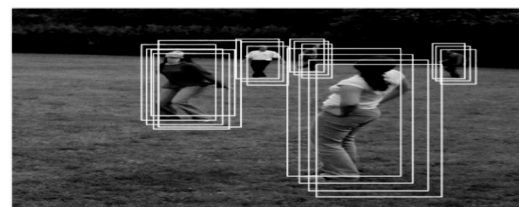


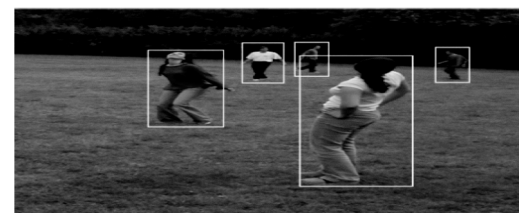
图 1 64 * 128 图像的 HOG 特征统计结果图

2 非极大值抑制法的候选区域融合

由于拍摄距离、角度以及人身高等因素的影响,行人在图片中以大小不一的姿态呈现。采用滑动窗口在单一尺寸图片中做滑动检测,往往因滑动窗口尺寸与检测目标不匹配而出现漏检的情况。为了提高算法的识别率,采用多尺度滑动窗口策略(滑动窗口尺寸不变,检测图片以一定比例缩小)进行目标检测,这在降低了漏检率的同时,也导致了一个目标被多个窗口标定,如图 2(a)所示。为了消除多余的标定窗口,本文采用非极大值抑制算法(Non-Maximum Suppression, NMS)对检测出的重叠窗口做矩形融合。提取检测区域中置信度最大的窗口,抑制与其重叠的窗口^[12]。采用非极大值抑制算法做矩形融合后的结果如图 2(b)所示。



(a) 目标被多个矩形标定



(b) 非极大值抑制算法对多矩形融合

图 2 非极大值抑制算法的矩形融合图

3 串-并联组合 SVM 分类器

3.1 SVM 的基本原理

训练 SVM 分类器的过程其实是寻找最优分割超平面的过程,在保证分割正负样本的前提下,使得支持向量到分割超平面的几何间距最大^[13]。设分割超平面为 $g(x) = \omega \cdot x + b$,则支持向量到分割超平面的几何间隔为:

$$\delta_i = \frac{y_i(\omega \cdot x_i + b)}{\|\omega\|} \quad (3)$$

其中, y_i 表示样本所属类别。寻求 δ 最大,就是寻求 $\|\omega\|$ 最小的过程。因此,线性可分 SVM 最优化问题可归结为:在满足不等式 $y_i(\omega \cdot x_i + b) - 1 \geq 0$ 的前提下,寻求最小 $\frac{1}{2} \|\omega\|^2$ 。

此后,再将提取的 HOG 特征向量带入 SVM 分类器中计算置信度,计算方式如图 3 所示。

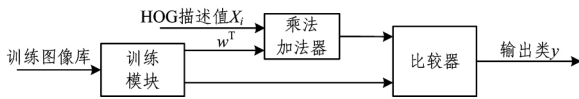


图3 SVM分类器原理图

3.2 确定分类决策准则

由于训练样本类型及训练出来的单个分割超平面的限制,行人检测过程中经常出现漏检、误检等情况。为改善分类器性能,本文采用 SVM 串并联组合策略,将两个 SVM 分类

器重新组合形成新的分类器,并计算出每个候选区域在 SVM1,SVM2 中的置信度 \$p_1\$ 和 \$p_2\$。为了提高算法的识别率,降低漏检率与召回率,确定算法分类决策准则为:

$$(p_1 + p_2) \geq 1.6 \wedge (p_1 > 0.6) \wedge (p_2 > 0.6) \quad (4)$$

其中,置信度 \$(p_i > 0.6) (i=1,2)\$ 的设定方式是为了防止分类器在过滤候选区域时出现漏检,而置信度 \$p_1 + p_2 > 1.6\$ 是为了防止出现误检。通过两个分类器同时过滤后,剩余矩形框即判定为行人区域,检测模型如图 4 所示。

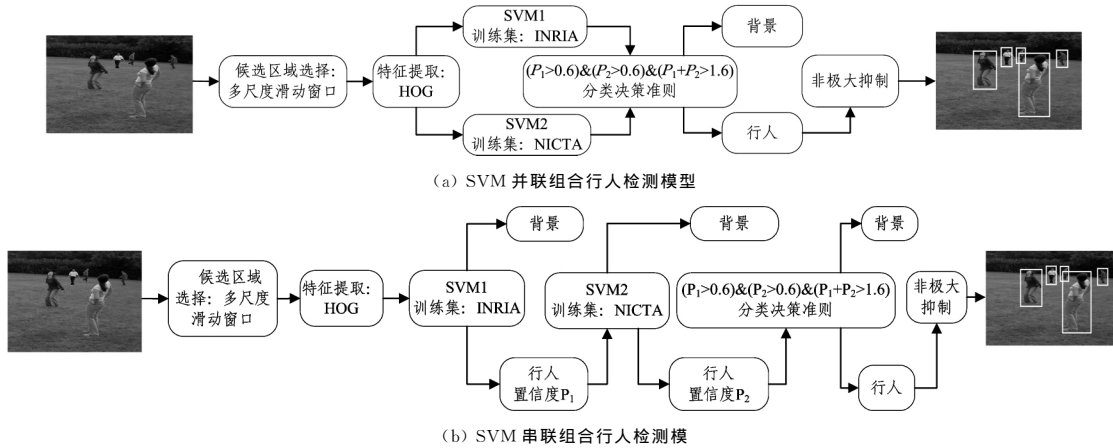


图4 SVM组合检测流程图

4 实验结果及分析

本文实验数据分别取自 INRIA 和 NICTA 行人数据集。其中,提取 INRIA 行人数据集正、负样本各 2000 张,用于训练分类器 SVM1;提取 NICTA 行人数据集正、负样本各 2000 张,用于训练分类器 SVM2。在分类器并联组合方式中,先将提取的 HOG 特征向量同时带入 SVM1 分类器与 SVM2 分类器中计算置信度 \$p_1\$ (由分类器 SVM1 计算得到)与 \$p_2\$ (由分类器 SVM2 计算得到),再通过图 4(a)模型中的判定准则判定该区域中是否存在行人;在分类器串联组合方式中,先将提取的 HOG 特征向量带入 SVM1 分类器中计算置信度 \$p_1\$,过滤后的候选区域 HOG 特征带入 SVM2 分类器中计算置信度 \$p_2\$,再通过图 4(b)模型中的判定准则判定该区域中是否存在行人。

实验分别通过采用单个 SVM 和 SVM 的串联、并联组合方式,作用在多人数据集(图片大小为 \$640 \times 480\$,图片数量为 300,图片中行人共 613 个)进行对比实验,其中正检数(正确检测到的行人数)、漏检数以及误检数的数据统计情况如表 1 所列。

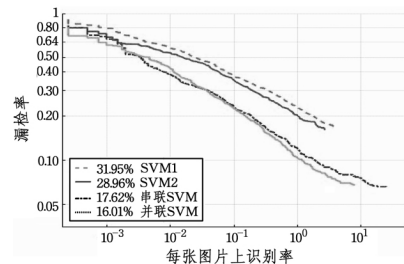
表1 4种检测模型在多人数据库中的检测结果对比表

分类器	图片数	正检数	误检数	漏检数	准确率 /%	召回率 /%	检测速度 /fps
SVM1	300	458	215	155	68.05	74.71	1.16
SVM2	300	471	192	142	71.04	76.84	1.24
串联 SVM	300	533	114	80	82.38	86.95	1.37
并联 SVM	300	535	102	78	83.99	87.28	2.39

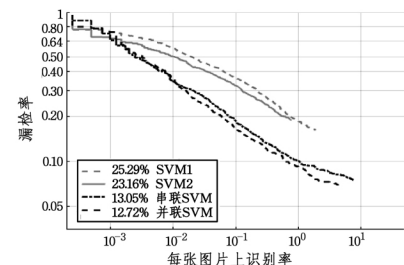
由表 1 可知,使用单一线性 SVM1 分类器,检测准确率为 68.05%,召回率为 74.71%;单一使用 SVM2 分类器,检测准确率为 71.04%,召回率为 76.84%。然而将 SVM1 与 SVM2 串联组合成新的分类器,检测准确率为 82.38%,召回

率为 86.95%;并联组合成新的分类器,准确率为 83.99%,召回率为 87.28%。不管是并联组合还是串组合,其识别率都比使用单一分类器高出 10% 以上。

为了说明 4 种分类器的检测情况,分别绘制了误检率结果对比图与漏检率结果对比图做综合分析,如图 5 所示。



(a) 误检率结果对比图



(b) 漏检率结果对比图

图5 4种分类器在多人数据库的检测结果

通过图 5 分析可知,在误检率和漏检率上,不管是采用 SVM 串联组合还是采用 SVM 并联组合,识别结果的总体下降趋势都明显优于使用单一 SVM 的识别结果。

综合分析可知,采用串联组合 SVM 分类器时识别效果最佳。并联组合 SVM 分类器虽然在准确率与召回率上与串联组合 SVM 分类器相当,但是在平均用时上却明显多于采用

串联组合 SVM 分类器。其原因是并联组合 SVM 分类器在检测行人的过程中,实际检测窗口数为 $2 \times (\frac{width-64}{8}) \times (\frac{height-64}{8})$,而在串联组合中,实际检测窗口数为 $(\frac{width-64}{8}) \times (\frac{height-64}{8}) + num$,其中 num 为第一道检测模型过滤后的候选区域窗口数目,由于经过第一道检测过滤后,大量背景区域被排除,在数量级上 num 远小于 $(\frac{width-64}{8}) \times (\frac{height-64}{8})$,因此才会出现串联组合在识别率上与并联组合相当,但是在检测速度上明显优于并联组合。综合考虑,采用串联组合的 SVM 分类器,在不明显降低检测速度情况下,可以有效降低误检率、漏检率。

结束语 本文比较了单一 SVM 分类器与多个 SVM 分类器组合方式在行人检测中的性能。实验证明,组合 SVM 分类器可以有效降低误检率和漏检率,其中串联 SVM 分类器在降低误检率和漏检率的同时,仍保持着高效的检测速度。本文主要对采用单一分类器的行人检测方式进行了改进,在特征提取方式上仅采用单一 HOG 特征,该特征注重检测区域轮廓信息而忽视了其他特征信息,而行人特征具有多样性(如纹理信息、颜色信息、尺度信息等),单独使用一种 HOG 特征对候选区域判定会造成一定的影响。因此,改进特征提取方式,提高特征多样性,采用多特征融合分类方式是我们未来的研究内容。

参考文献

- [1] FLORES A, BELONGIE S. Removing pedestrians from Google street view images[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, San Francisco: IEEE Press, 2010: 53-58.
- [2] ANDRILUKA M, SCHNITZSPAN P, MEYER J, et al. Vision based victim detection from unmanned aerial vehicles[C]// Proceedings of IEEE Conference on Intelligent Robots and Systems. Taipei: IEEE Press, 2010: 1740-1747.
- [3] REILLY V, SOLMAZ B, SHAHM. Geometric constraints for human detection in aerial imagery[C]// Proceedings of Springer Conference on Computer Vision. Heraklion: Springer Press, 2010: 252-265.
- [4] OREN M, PAPAGEORGIOU C, SINHA P, et al. Pedestrian detection using wavelet templates[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Bombay: IEEE Press, 1997: 193-199.
- [5] 王坚, 兰天. 基于稀疏表达和机器学习的行人检测技术研究[J]. 计算机科学, 2016, 43(s1): 207-209.
- [6] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego: IEEE Press, 2005: 886-893.
- [7] MAJI S, BERG A C, MALIK J. Classification using intersection kernel support vector machines is efficient[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: IEEE Press, 2008: 1-8.
- [8] DOLLAR P, APPER R, BELONGIES S, et al. Fast Feature Pyramids for Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(8): 1532-1545.
- [9] 苏松志. 行人检测: 理论与实践[M]. 厦门: 厦门大学出版社, 2016: 21-112.
- [10] 苏松志, 李绍滋, 陈淑媛, 等. 行人检测技术综述[J]. 电子学报, 2012, 40(4): 814-820.
- [11] 甘玲, 邹宽中, 刘肖. 基于 PCA 降维的多特征级联的行人检测[J]. 计算机科学, 2016, 43(6): 308-311.
- [12] 方智文, 曹治国, 肖阳. 基于多尺度局部极值和边缘检测的目标性算法[J]. 信号处理, 2016, 32(8): 911-921.
- [13] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 95-136
- [14] 廖斌, 苏涛, 刘斌. 基于多尺度分解的邻域随机查找快速图像修复[J]. 电子与信息学报, 2015, 37(9): 2097-2102.
- [15] ZHANG H, DAI S. Image Inpainting Based on Wavelet Decomposition[J]. Procedia Engineering, 2012, 29(4): 3674-3678.
- [16] 张东, 唐向宏, 张少鹏, 等. 小波变换与纹理合成相结合的图像修复[J]. 中国图象图形学报, 2015, 20(7): 882-894.
- [17] 石宏理, 胡波. 双树复小波变换及其应用综述[J]. 信息与电子工程, 2007, 5(3): 229-234.
- [18] KINGSBURY N. The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement[C]// Signal Processing Conference. IEEE, 1998: 319-322.
- [19] KINGSBURY N. Complex Wavelets for Shift Invariant Analysis and Filtering of Signals[J]. Applied & Computational Harmonic Analysis, 2001, 10(3): 234-253.
- [20] SELESNICK I W, Baraniuk R G, Kingsbury N G. The dual-tree complex wavelet transform[J]. IEEE Signal Processing Magazine, 2005, 22(6): 123-151.
- [21] 吴一全, 宋昱. 基于双树复小波域 HMT 模型的煤燃烧火焰图像去噪[J]. 华南理工大学学报(自然科学版), 2014, 42(1): 59-65.

(上接第 182 页)

- [8] WEXLER Y, SHECHTMAN E, IRANI M. Space-time completion of video[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(3): 463-476.
- [9] BUGEAU A, BERTALMIÓ M, CASELLES V, et al. A comprehensive framework for image inpainting[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2010, 19(10): 2634-2645.
- [10] 孟春芝, 何凯, 焦青兰. 自适应样本块大小的图像修复方法[J]. 中国图象图形学报, 2012, 17(3): 337-341.
- [11] 李志丹, 和红杰, 尹忠科. 基于块结构稀疏度的自适应图像修复算法[J]. 电子学报, 2013, 41(3): 549-554.
- [12] RAN L, MENG X. Fast Seam Carving Using Gaussian Pyramid [C]// 2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC). IEEE, 2014: 59-63.
- [13] REN S, LEI J, ZHANG T, et al. Research of High Performance Information Hiding Scheme Based on Gaussian Pyramid and CARDBAL2 multi-wavelet for Secret Communication[J]. International Journal of Applied Mathematics & Statistics™, 2014,