

基于场景模型与统计学习的鲁棒行人检测算法

杨涛¹ 李静² 潘泉³ 张艳宁¹

摘要 提出一种基于场景模型和统计学习的行人检测算法. 针对训练行人检测器时面临的动态场景的复杂性和行人样本多样性等问题, 通过背景建模, 从场景的背景图像上提取有限的负样本用于训练, 大幅度提高了分类器的检测率, 同时降低了虚警; 提出一种快速弱分类器选择算法, 根据正、负样本特征大小的分布和期望的检测率, 直接求解特征大小的阈值范围, 能够满足在线训练和更新检测器的要求; 提出一种基于正样本错误率的训练算法, 先根据正样本加权错误率选择弱分类器, 快速提高检测率, 在训练结束后调整最终分类器的加权系数, 在保证检测率的同时尽可能降低虚警率. 实验中构建了一个试验视频数据库和行人样本库, 数据库包括雨、雪、阴影、季节变化、摄像机平移、旋转、缩放等情况, 并设计实现了一个实时行人检测系统 BMAT (Background modeling and Adaboost training), 实验结果证明了算法的有效性.

关键词 行人检测, 背景建模, 统计学习, 智能视频监控

DOI 10.3724/SP.J.1004.2010.00499

Scene Modeling and Statistical Learning Based Robust Pedestrian Detection Algorithm

YANG Tao¹ LI Jing² PAN Quan³ ZHANG Yan-Ning¹

Abstract A scene model and statistic learning based method for pedestrian detection in complicated real-world scenes is proposed. A unique characteristic of the algorithm is its ability to train a special cascade classifier dynamically for each individual scene. The benefit is that the classifier only focuses on the differences between the positive samples and the limited negative samples of each individual scene, thus greatly reduces the complexity of classification, and achieves robust detection result even with a few classifiers. A highly efficient weak classifier selection method and a novel boosting architecture are presented to speed up feature selection and classifier training. To evaluate the proposed algorithm, we captured pedestrian videos under different weathers, seasons and camera motions, and labeled 4300 positive samples. Moreover, a real-time pedestrian detection system named as background modeling and Adaboost training (BMAT) was developed, which produced fast and robust detection results as demonstrated by extensive experiments performed using video sequences under different environments.

Key words Pedestrian detection, background modeling, statistical learning, intelligent video surveillance

行人是视频监控系统关注的重要对象. 通过对行人的检测、跟踪、轨迹分析以及行为识别, 系统可实时检测异常事件并报警, 变被动监视为主动预警.

行人检测与定位的准确性将直接影响后续工作以至整个系统的性能. 由于中、远距离行人分辨率低, 特征信息不明显, 服饰、姿态各异, 模式多样, 再加之纷繁复杂的真实场景, 使得室外中、远距离下的行人检测成为一个极具挑战性的课题.

粗略地讲, 从视频中检测行人的解决思路可分为两类: 一类是采用背景建模^[1-3], 通过前景检测, 提取场景中的运动目标, 然后进一步提取特征, 将运动目标分为人、车、噪声干扰等, 这里, 背景模型的设计是研究的重点; 另一类是采用统计学习方式, 根据大量样本训练行人检测器.

1) 背景建模: 由于缺乏先验知识, 基于背景建模的方法仅将输入图像与背景的差异作为目标提取的手段, 这种方式使得人们很难设计一种鲁棒的背景模型, 同时解决真实环境中可能出现的干扰, 如雨、雪、风、波浪、树叶摇摆、光照突变、平台运动等. 在各类干扰下, 简单的背景模型抗干扰能力差, 而过度复杂的模型往往计算量大, 且对参数敏感, 鲁棒性差.

收稿日期 2008-10-22 录用日期 2009-01-21
Manuscript received October 22, 2008; accepted January 21, 2009

国家高技术研究发展计划 (863 计划) (2009AA01Z315), 国家自然科学基金 (60903126, 60872145, 60634030), 中国博士后科学基金 (20090451397), 教育部高等学校科技创新工程重大项目培育资金 (708085) 资助

Supported by National High Technology Research and Development Program of China (863 Program) (2009AA01Z315), National Natural Science Foundation of China (60903126, 60872145, 60634030), China Postdoctoral Science Foundation (20090451397), and Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (708085)

1. 西北工业大学计算机学院陕西省语音与图像处理重点实验室 西安 710129 2. 西安电子科技大学通信工程学院 西安 710071 3. 西北工业大学自动化学院 西安 710129

1. Shaanxi Key Laboratory of Speech and Image Information Processing, School of Computer Science, Northwestern Polytechnical University, Xi'an 710129 2. School of Telecommunications Engineering, Xidian University, Xi'an 710071 3. School of Automation, Northwestern Polytechnical University, Xi'an 710129

2) 统计学习: 行人检测本质上是一个分类问题, 受到统计学习算法在人脸检测中成功应用的鼓舞, 许多学者采用外观特征和统计学习相结合的思路训练行人检测器, 常用的外观特征包括目标的灰度、边缘^[4-7]、纹理、形状^[8-12]、梯度方向直方图^[13-16]、温度分布^[17] 等信息; 统计分类的方法包括神经网络、支持向量机^[18] 和 Adaboost^[5-6, 11, 19]. 但是, 相对于人脸而言, 将行人作为正样本训练分类器存在以下难点: 1) 行人的姿态、服装色彩各不相同 (见图 1(b)), 在特征空间中的模式不紧凑; 2) 分类器的性能直接受训练样本的影响, 而真实场景纷繁复杂 (见图 1(a)), 离线训练时有限的负样本无法涵盖所有场景.

上述两种思路分别从正、反两方面检测目标: 前者将监控场景的背景作为参考, 而后者是将待检测的行人作为参考, 检测都是建立在比较输入图像和参考图像之间的差异上. 实际上, 尽管真实场景的负样本是无穷的, 但是在应用时, 单个场景包含的负样本数量却十分有限. 在某一特定场景中, 检测器性能的好坏只取决于它能否准确地拒绝该场景内的负样本. 基于此, 本文将上述两种思路相结合, 提出一种新的基于背景建模与快速 Adaboost 训练的行人检测算法, 并研发了一个实时行人检测系统 BMAT (Background modeling and Adaboost

training). BMAT 采用离线采集的行人图片作为正样本, 用背景建模得到的场景参考背景生成负样本, 这种负样本产生方式充分利用了监控场景的固有信息, 大幅度减少了训练分类器所需的负样本数量, 提高了检测算法在该场景下的鲁棒性. 同时, 考虑到光照等因素造成背景的缓慢变化, 提出一种快速弱分类器选择算法和一种基于正样本错误率的分类器训练算法, 用于在线训练和更新检测器.

本文的后续部分安排如下: 第 1 节给出行人检测系统的总体介绍, 第 2 节具体描述行人检测器和相关快速特征选择与分类器训练算法, 第 3 节给出实验结果, 第 4 节是结论.

1 系统框架

图 2 给出 BMAT 系统流程图, 该系统主要包括两个模块: 1) 在线行人检测; 2) 在线检测器训练与更新. 输入为灰度化后的图像序列.

对于输入图像序列, 一方面, 在线行人检测模块 (图 2, 左侧) 采用训练好的行人检测器多尺度扫描输入图像, 然后对重叠较大的检测窗口合并, 作为本帧检测结果; 另一方面, 考虑到即使是同一场景, 其背景图像也会随着时间、光照变化而改变, 为了适应这种变化, 在线检测器训练与更新模块 (图 2, 右侧)



(a) 自建行人检测实验数据库示例

(a) Examples of testing video database and pedestrian sample



(b) 行人样本示例

(b) Examples of testing video database and pedestrian sample

图 1 测试数据与行人样本

Fig. 1 Testing video data and pedestrian samples

用行人检测器扫描背景建模生成的场景背景图像, 并将从背景图上误检为行人的子图像作为新的负样本, 当新负样本数目高于给定阈值, 则启动本文提出的快速弱分类器选择算法和基于正样本错误率的训练算法, 对现有的行人检测器进行在线更新.

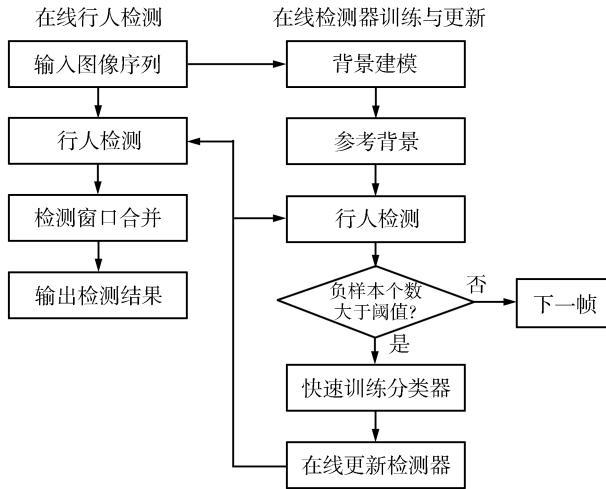


图2 基于场景模型与统计学习的行人检测框架

Fig.2 Scheme of scene modeling and statistic learning based pedestrian detection algorithm

本系统采用背景建模算法生成参考背景图像. 背景建模与前景减除算法是应用最为广泛的一类运动目标检测算法, 其基本思想是从输入图像和背景图像的差异中提取运动前景. 常见的背景建模算法包括中值滤波、线性预测、单高斯、混合高斯模型、核密度估计法、局部纹理分布等. 本系统在背景建模时选用基于像素灰度统计特性的混合高斯模型 (Gaussian mixture model, GMM)^[20]. GMM 模型对场景中枝叶晃动、光照变化等干扰具有较好的鲁棒性. 在该模型中, 图像上每一个像素的灰度值用 K 个高斯分布描述, K 值越大, 处理灰度变化的能力越强. 本文用 GMM 中权值最大的高斯分布的均值组成场景背景图.

综上所述, 在结构上, 本系统用背景图像产生负样本, 用离线采集的行人图像构建正样本, 降低了分类器训练难度; 在分类器训练上, 提出的快速训练算法使得行人检测器能够随着动态场景改变而自动更新.

2 行人检测算法

2.1 特征选取

特征对机器学习算法的最终学习结果有非常大的影响, 这主要体现在正确率和处理速度两个方面. 特征大小的计算复杂度是特征性能的一个重要的评

价因素. 过于复杂的特征计算必然会影响到检测系统的处理速度, 从而难以达到实时处理的要求. 矩形特征虽然只有水平和垂直两个方向, 但却可以通过各种组合有力地为机器学习提供丰富的图像描述, 而且所有的特征均可用积分图高效计算. 基于此, 本文选用 Viola^[6] 采用的四类对称矩形特征 (图 3(a)~3(d)) 和四类非对称特征用于特征提取.

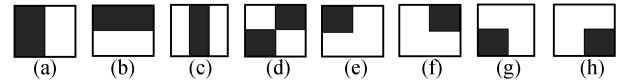


图3 本文选择的矩形特征

Fig.3 Rectangle features used in our method

在图 3 中, 图 3(a) 和图 3(b) 为 2-矩形特征, 图 3(c) 为 3-矩形特征, 图 3(d) 为 4-矩形特征. 2-矩形特征由两个大小和形状相同、并且是垂直或者水平相邻的矩形构成. 4-矩形特征中的类型特征是由四个大小和形状相同、并且分布在两条对角线上的矩形组成. 2-矩形特征和 4-矩形特征的大小定义为

$$f(r) = \left| \sum_{(x,y) \in S_B} i(x,y) - \sum_{(x,y) \in S_W} i(x,y) \right| \quad (1)$$

其中, S_B, S_W 分别表示矩形特征中的黑色区域和白色区域, $i(x,y)$ 表示矩形区域内像素点 (x,y) 的亮度值. 3-矩形的特征是由三个同样的并且相邻的矩形组成, 其特征的大小定义为

$$f(r) = \left| 2 \cdot \sum_{(x,y) \in S_B} i(x,y) - \sum_{(x,y) \in S_W} i(x,y) \right| \quad (2)$$

图 3(e)~3(h) 为非对称特征, 其特征的大小计算如下:

$$f(r) = \left| 3 \cdot \sum_{(x,y) \in S_B} i(x,y) - \sum_{(x,y) \in S_W} i(x,y) \right| \quad (3)$$

给定矩形特征在样本中的大小、位置和阈值, 就构成了一个简单的分类器. 每次开始训练前, 需要先对所有可能的分类器中选择出分类正确率大于随机猜测的弱分类器, 用于后继训练.

2.2 基于特征取值分布的快速弱分类器选择算法

Adaboost 算法在训练前需要先产生弱分类器集合. 直接搜索的方法选择弱分类器的过程如下:

1) 遍历全部 8 类矩形特征. 对每一类首先按比例增大矩形的尺寸, 然后遍历样本中可能的相对位置, 最后将矩形与可能的阈值进行组合, 得到不同大小、不同阈值和不同位置的分类器.

2) 遍历样本集合. 对每一个分类器, 求解分类错误率, 将错误率小于 50% 的分类器挑选为弱分类器. 由于上节中矩形特征的大小、位置和阈值等均可改变, 因此遍历样本集的计算量 C 约为

$$C = F \cdot \min(rw, rh) \cdot (rw \cdot rh) \cdot hT \cdot (n^+ + m^-) \quad (4)$$

其中, F 是矩形特征的类型数, rw, rh 是矩形特征的宽和高, $\min(rw, rh)$ 是矩形特征大小变化范围, hT 是量化后的阈值范围, n^+, m^- 是正、负样本数. 为了实现在线动态更新检测器, 必须加快弱分类器选择速度. 由于类型、大小和在样本中的相对位置是描述矩形特征的最基本要素, 因此式 (4) 第一部分 $F \cdot \min(rw, rh) \cdot (rw \cdot rh)$ 的计算量难以约减. 然而阈值搜索范围是由样本决定, 如果给定某一矩形特征, 只搜索可能的阈值搜索范围和部分样本, 则式 (4) 第二部分 $hT \cdot (n^+ + m^-)$ 的计算量将大幅度降低. 为此, 提出一种快速弱分类器选择算法, 该算法根据正、负样本特征的大小的分布和期望的检测率、漏检率, 直接求解特征大小的阈值范围.

用 $f = \{f(1), \dots, f(F)\}$ 表示特征集, $P_r^+(t)$ 和 $P_r^-(t)$ 代表 $f(r)$ 的特征大小在正、负样本集的概率密度:

$$\begin{aligned} P_r^+(t) &= \frac{k^+}{n^+} \\ P_r^-(t) &= \frac{k^-}{n^-} \end{aligned} \quad (5)$$

其中, k^+ 和 k^- 表示特征的大小为 t 的正、负样本个数. 特征大小的分布函数 $F_r^+(t)$ 和 $F_r^-(t)$ 为:

$$\begin{aligned} F_r^+(t) &= \sum_{T \leq t} P_r^+(T) \\ F_r^-(t) &= \sum_{T \leq t} P_r^-(T) \end{aligned} \quad (6)$$

给定特征在样本上的相对位置和尺寸后, 一个弱分类器 h_r 由选用的特征类型 $f(r)$ 、阈值 θ_r 以及不等号方向的符号 p_r 确定:

$$h_r = \begin{cases} 1, & \text{若 } p_r f(r) > p_r \theta_r \\ 0, & \text{否则} \end{cases} \quad (7)$$

根据弱分类器的定义, 选中的分类器应满足:

$$\frac{FNR \cdot n^+ + FPR \cdot n^-}{n^+ + n^-} \leq 0.5 \quad (8)$$

其中, 漏检率 $FNR =$ 错分为负的正样本数/正样本数, FNR 越低检测概率越高; 虚警率 $FPR =$ 错分为正的负样本数/负样本数. 由于采用级联分类器的结构, 每层训练的强分类器需要较高的检测率和一

定的虚警率. 因此, 在进行初始分类器选择时, 通过设置 FNR 和 FPR , 让选择结果偏重于正样本. 实验中采用的 $FNR = 0.3, FPR = 0.5$. 给定 FNR 和 FPR 后, 阈值搜索范围 $[ts_{\min}, ts_{\max}]$ 由训练样本特征取值的分布函数决定, 式 (9) 给出了选择弱分类器时直接确定阈值搜索范围 $[ts_{\min}, ts_{\max}]$ 的计算公式.

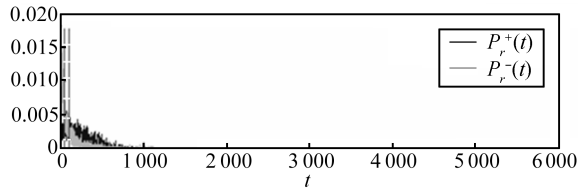
$$\begin{aligned} ts_{\min} &= \begin{cases} \min_t \{t | F_r^-(t) > 1 - FPR\}, & \text{若 } p_r = 1 \\ \min_t \{t | F_r^+(t) > 1 - FNR\}, & \text{若 } p_r = -1 \end{cases} \\ ts_{\max} &= \begin{cases} \min_t \{t | F_r^+(t) > FNR\}, & \text{若 } p_r = 1 \\ \min_t \{t | F_r^-(t) > FPR\}, & \text{若 } p_r = -1 \end{cases} \end{aligned} \quad (9)$$

以 $p_r = 1$ 为例, 由式 (7) 知分类器会将特征取值小于阈值 t 的样本判断为负样本, 根据分布函数的定义, $F_r^-(t)$ 表示负样本集合中特征取值小于 t 的样本比率, 即为正确分类的负样本比率, 因此, 分类器的阈值 t 需满足 $F_r^-(t) > 1 - FPR$, 则阈值搜索范围下限为 $ts_{\min} = \min_t \{t | F_r^-(t) > 1 - FPR\}$; 同理, $F_r^+(t)$ 表示阈值为 t 时错分为负的正样本比率, 该比率应低于 FNR , 则阈值搜索范围上限为 $ts_{\max} = \min_t \{t | F_r^+(t) > FNR\}$. 在遍历特征时, 如果某一矩形特征计算得到的 $ts_{\max} > ts_{\min}$, 则表示存在可能的阈值范围, 使得选择的分类器能够满足设定的漏检率和虚警率, 否则直接拒绝该特征.

图 4 给出一个阈值选取示意图. 对于一个已知类型、大小和在样本图像内相对位置的矩形特征, 通过一次遍历整个样本集, 可以求解该特征大小的概率密度 $P_r^+(t)$ (图 4(a), 图 4(b), 黑色曲线) 和 $P_r^-(t)$ (图 4(a), 图 4(b), 灰色曲线), 以及对应的分布函数 $F_r^+(t)$ 和 $F_r^-(t)$ (图 4(c)). 从图 4(a) 和图 4(b) 可以看出, 尽管整个特征大小的值域范围是 $[0, 6000]$, 但绝大多数特征取值都集中在 $[0, 1100]$ 范围内, 显然, 在搜索时无需遍历整个特征空间. 在得到分布函数后, 可以按照式 (9), 直接确定阈值搜索范围 $[ts_{\min}, ts_{\max}]$. 当 $p_m = 1$ 时, 得到的搜索范围为 $[48, 111]$ (图 4(c)), 仅相当于整个特征取值空间的 1.05%.

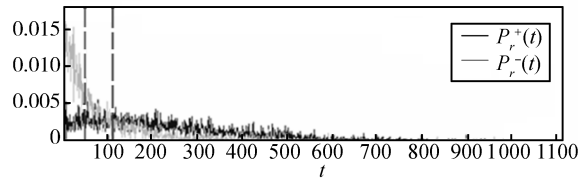
将搜索范围内任意阈值和矩形特征相组合, 都可以构成一个满足检测率和漏检率的弱分类器. 但是, 一方面考虑到 Adaboost 组合分类的特点, 一个高效的弱分类器集合里, 不同的弱分类器应该对不同类型的样本具有较高的分类能力; 另一方面, 弱分类器的数量和后继 Adaboost 训练速度密切相关. 因此, 本文在弱分类器选择时, 对于一个给定类型、

大小和位置的矩形特征, 在确定阈值搜索范围后, 只挑选一个最优阈值 θ_r 和该特征组合, 得到一个弱分类器.



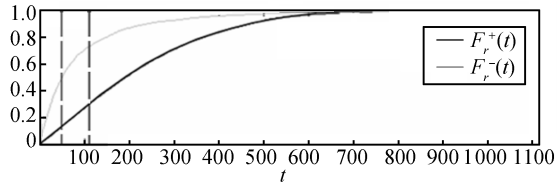
(a) 正、负样本特征大小的概率密度分布

(a) Probability density function of positive (black) and negative (gray) samples



(b) 正、负样本特征大小的概率密度分布

(b) Probability density function of positive (black) and negative (gray) samples



(c) 正、负样本分布函数 (正、负样本曲线分别用黑色和灰色表示, 虚线给出选取的阈值搜索范围 $[ts_{\min}, ts_{\max}]$.)

(c) Distribution function of positive (black) and negative (gray) samples (Note that dash lines display the searching scope $[ts_{\min}, ts_{\max}]$.)

图 4 阈值搜索范围示意图

Fig. 4 Searching scope of weak classifier's threshold

$$\theta_r = \begin{cases} \arg \min_t \left(\frac{F_r^+(t) \cdot n^+ + (1 - F_r^-(t)) \cdot n^-}{n^+ + n^-} \right), & \text{若 } p_r = 1 \text{ 且 } t \in [ts_{\min}, ts_{\max}] \\ \arg \min_t \left(\frac{1 - F_r^+(t) \cdot n^+ + (F_r^-(t)) \cdot n^-}{n^+ + n^-} \right), & \text{若 } p_r = -1 \text{ 且 } t \in [ts_{\min}, ts_{\max}] \end{cases} \quad (10)$$

式 (10) 中, 当不等号方向的系数 $p_r = 1$ 时, $F_r^+(t) \cdot n^+$ 表示被错误分类的正样本个数, $(1 - F_r^-(t)) \cdot n^-$ 为错误分类的负样本个数, 最优阈值 θ_r 使整体错误率达到最小. $p_r = -1$ 的情况与之类似.

上述算法能快速生成弱分类器集合, 由于一个

矩形特征仅选用一个最优阈值, 减少了集合内分类器个数, 且分类器间的耦合性降低, 有利于提高后继分类器训练效率.

2.3 基于正样本错误率的分类器训练算法

产生弱分类器集合后, 下一步就是用 Adaboost 方法训练分类器. 如前所述, 由于最终的行人检测器采用级联分类器的结构, 因此分类性能的好坏并不由某一级分类器性能决定, 而是由整个级联分类器共同决定. 文献 [6] 在 Adaboost 每一轮训练时选择令整体样本加权误差最小的分类器, 同时降低 FNR 和 FPR . 由于忽略了级联分类器的特性, 为了达到检测率的要求, 文献 [6] 中的算法需要进行多轮迭代, 不仅增加了训练时间, 且每层中弱分类器个数的增多会增加检测时的计算量, 影响系统的实时性.

为了充分发挥分级处理的优势, 本文在算法设计时并不要求级联分类器的每一级同时具有高检测率和低虚警率, 而是采用基于正样本错误率的训练方法, 在保证高检测率 (大于 99.8%) 的前提下, 放宽对虚警率的要求 (如虚警率小于 60% 即可).

实验结果表明, 该算法提高了对正样本检测率, 缩短了训练轮数, 使得每一级弱分类器的个数大幅度减少, 提高了在线检测的速度. 尽管该算法在每一级分类器训练时并不能够保证同时降低 FNR 和 FPR , 但是级联滤波可快速滤除被错误分类的负样本, 从而保证最终检测器的性能要求.

在训练完成后, 分类器由下式计算:

$$h(x) = \begin{cases} 1, & \sum_{t=1}^T \alpha_t h_t(x) \geq \mu \sum_{t=1}^T \alpha_t \\ 0, & \text{否则} \end{cases} \quad (11)$$

式中加权系数 μ 的大小对检测结果有着重要影响, 增加 μ 可使样本被分类为负样本的概率增加. 由于本算法在训练时不能保证同时降低 FPR , 为此, 在训练结束后, 通过增大式 (11) 的加权系数 μ , 在保持检测率满足要求的同时尽可能降低虚警率 FPR . 由于 $\sum_{t=1}^T \alpha_t > 0$, 则式 (11) 可改写为:

$$h(x) = \begin{cases} 1, & \psi \geq \mu \\ 0, & \text{否则} \end{cases}, \quad \psi = \frac{\sum_{t=1}^T \alpha_t h_t(x)}{\sum_{t=1}^T \alpha_t} \quad (12)$$

对所有的正样本遍历一次, 计算 ψ 取值的概率密度 $P(\psi)$ 和分布函数 $F(\psi)$:

$$P(\psi) = \frac{k^+(\psi)}{n^+}, \quad F(\psi) = \sum_{t \leq \psi} P(t) \quad (13)$$

式中 $k^+(\psi)$ 表示取值大小为 ψ 的正样本的个数. 为保证对正样本 99.8% 的检测率, 即 $FNR \leq 0.002$, 可根据分布函数 $F(\psi)$ 计算加权系数 μ 的最大值 μ_{\max} , 使得分类器在保证检测率的同时降低虚警率 FPR .

$$\mu_{\max} = \min_{\psi} \{\psi | F(\psi) \geq 0.002\} \quad (14)$$

综上所述, 本文分类器训练算法的主要步骤如下:

步骤 1. 已知训练集 $S = \{(x_1, y_1), \dots, (x_N, y_N)\}$, 其中, x_i 表示 24×32 的样本图像, $y_i \in Y = \{0, 1\}$.

步骤 2. 采用第 2.2 节算法, 快速构建弱分类器集合.

步骤 3. 初始化权重

$$w_{1,j} = \begin{cases} \frac{1}{2m}, & \text{若 } y_i = 1 \\ \frac{1}{2l}, & \text{若 } y_i = 0 \end{cases}$$

其中 m, l 是正、负样本的个数.

步骤 4. 循环 $t = 1, \dots, T$.

步骤 4.1. 归一化权重

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^N w_{t,j}}$$

对于每一个弱分类器 h_j , 计算正样本加权误判率 ε_j , $\varepsilon_j = \sum_{i=1}^m w_i |h_j(x_i) - y_i|$, 并选取具有最小正样本加权误判率的弱分类器 h_t ;

步骤 4.2. 更新样本权重: $w_{t+1,i} = w_{t,i} \beta^{1-e_i}$,

$$e_i = \begin{cases} 0, & \text{若 } h_t(x_i) = y_i \\ 1, & \text{若 } h_t(x_i) \neq y_i \end{cases}, \beta = \frac{\varepsilon_t}{1 - \varepsilon_t}$$

步骤 4.3. 如果当前选出的弱分类器组合的 $FNR < 0.2\%$, 满足检测率要求, 则提前终止循环.

步骤 5. 调整最终分类器的加权系数, 组合成强分类器.

步骤 5.1. 计算 ψ 取值的概率密度 $p(\psi)$ 和分布函数 $F(\psi)$:

$$\psi = \frac{\sum_{t=1}^T \alpha_t h_t(x)}{\sum_{t=1}^T \alpha_t}, p(\psi) = \frac{k^+(\psi)}{n^+},$$

$$F(\psi) = \sum_{T \leq \psi} p(T);$$

步骤 5.2. 根据分布函数 $F(\psi)$ 计算加权系数 μ 的最大值 μ_{\max} , 在保证检测率的同时降低虚警率 FPR : $\mu_{\max} = \min_{\psi} \{\psi | F(\psi) \geq 0.002\}$;

步骤 5.3. 生成最终的强分类器:

$$h(x) = \begin{cases} 1, & \text{若 } \sum_{t=1}^T \alpha_t h_t(x) > \mu_{\max} \sum_{t=1}^T \alpha_t \\ 0, & \text{否则} \end{cases}$$

其中, $\alpha_t = \log \frac{1}{\beta_t}$.

3 算法实现及结果分析

以本文算法为核心, 用 VC++ 开发了一个实时行人检测系统 BMAT. 实验中硬件系统采用 Intel Core2 CPU 1.66 GHz 处理器, 内存 2 G. 为了测试和改进算法, 实地拍摄了不同季节 (夏、秋、冬)、不同天气 (晴天、雨天、雪天)、不同场景条件、不同摄像机运动模式 (摄像机平移、旋转、缩放) 下中、远距离的行人视频, 作为测试数据库, 并从部分视频序列中手工分割 4 430 张行人图片, 标准化为 24×32 的行人正样本. 实验中共手工采集行人正样本 4 430 个, 图 1 (b) 给出部分样本示意图. 实验中, 系统对于 320×240 的图像序列检测速度为 7 帧/秒, 如用帧间差或背景减除等方法提取的运动区域作引导, 则检测速度为 15 帧/秒.

3.1 弱分类器选择计算量比较

表 1 给出一组三种弱分类器选择算法的计算量比较. 训练时正、负样本各有 4 430 个, 直接遍历整个特征取值空间需要耗费 8 小时 17 分 13 秒 (表 1 第 1 行), 根据特征取值的最大、最小值进行遍历的方法耗时为 1 小时 22 分 52 秒 (表 1 第 2 行), 本文提出的快速弱分类器选择算法仅花费 5 分 58 秒就完成弱分类器选择 (表 1 第 3 行), 共挑选了 937 个弱分类器用于分类器训练, 耗时仅为直接遍历的 1.2%. 采用基于正样本错误率的算法用 4 分 12 秒训练完成一层强分类器, 达到了预期的漏检率 FNR 和虚警率 FPR . 从弱分类器选择到训练结束共耗时 10 分 10 秒, 考虑到真实场景的变化较为缓慢, 算法能满足在线动态更新行人检测器的要求.

表 1 弱分类器选择算法计算量比较

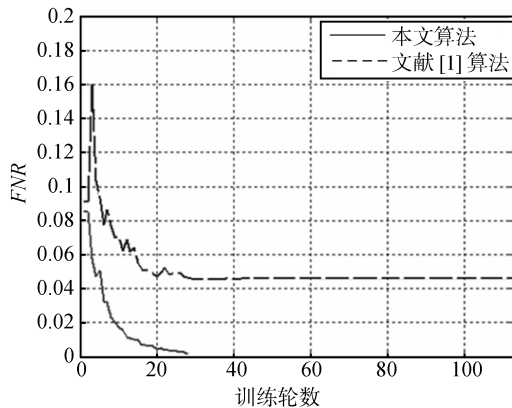
Table 1 Comparison of computational cost of weak classifier selection algorithms

弱分类器选择方法	计算时间	相对计算时间 (%)
直接遍历整个特征空间	8 小时 17 分 13 秒	100
最大、最小值遍历法	1 小时 22 分 52 秒	16.7
本文算法	5 分 58 秒	1.2

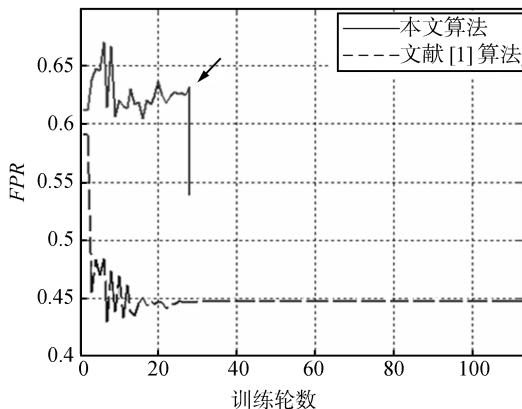
3.2 训练轮数比较

图 5 给出一组本文算法和文献 [6] 算法的比较结果. 在该组实验中, 样本集分别由 4 430 个正、负样本构成. 两种算法在训练时采用相同的弱分类器集合, 该集合由第 2.2 节算法选出. 从图 5 中可以看出, 在漏检率 FNR 方面, 本文算法仅用 28 个弱分类器就使 FNR 下降到 0.15% (图 5 (a), 实线), 而文

献 [6] 的算法使用了 100 个弱分类器尚未达到检测率的要求 (图 5 (a), 虚线). 在虚警率 FPR 方面, 通过在训练结束后用式 (14) 求解加权系数 μ_{\max} , 本文算法在保证 FNR 的同时, 将 FPR 从 63.3% 降低到 54.3%. 尽管此时 FPR 仍高于文献 [6] 的算法, 但考虑到最终的检测器采用的是级联分类的方式, 本文算法的训练结果已经达到级联分类器的设计要求, 即每一层的强分类器要保证低漏检率 ($FNR < 0.2\%$) 和一定的虚警率 ($FPR < 60\%$).



(a) FNR 曲线
(a) FNR curve



(b) FPR 曲线
(b) FPR curve

图 5 前景边缘点检测结果

Fig. 5 Comparison of FNR and FPR

3.3 固定摄像机室外中、远距离行人检测结果

图 6 给出一组在线学习、更新分类器的处理结果. 图 6 (a) 是输入图像, 图 6 (b) 是背景建模得到的背景图. 检测器更新前, 由于场景较复杂, 输入图像的检测结果中存在大量的虚警 (图 6 (c), 白色矩形框). 此时, 系统将背景图像上的检测结果 (图 6 (d)) 作为负样本, 快速训练新一层分类器. 图 6 (e) 给出

检测器更新后的处理结果, 可以看出, 绝大部分背景噪声被滤除. 本文采用的训练样本均为站立或行走的行人, 由于检测器对小范围尺度变化和位置平移不敏感, 当人体发生倾斜 (图 6 (f), 右下) 时, 超出样本覆盖的角度范围, 检测器将失效.

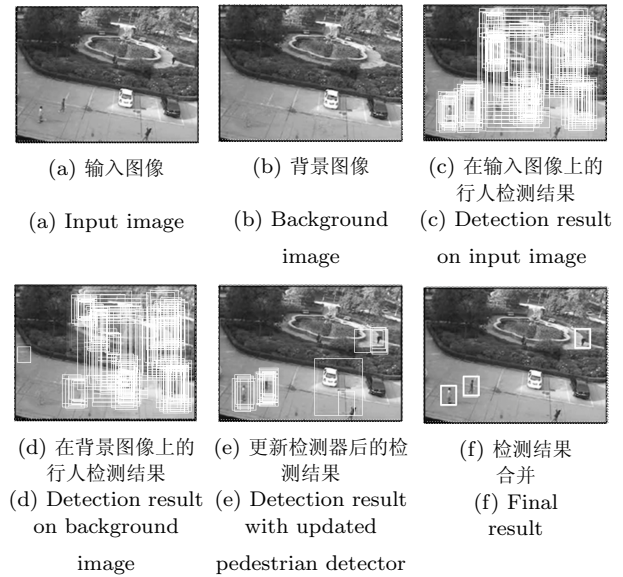


图 6 行人检测器学习与更新结果

Fig. 6 Online learning and updating results of pedestrian detector

图 7 给出一组在不同季节、不同天气条件下的行人检测结果. 图 7 (a) 中, 受光照影响, 目标周围存在明显的运动阴影. 由于运动阴影引发像素灰度变化, 使得它对基于背景建模的前景检测算法性能有较大影响. 但由于阴影和行人外观存在较大差异, 因此很容易从分类器的角度加以区分. 同理, 系统也能够区分场景中出现的运动车辆和行人 (图 7 (a)). 将这一特性和运动前景分割相结合能够对场景内的运动目标进行简单的分类. 图 7 (b) 和图 7 (c) 分别给出一组雪天和雨天的行人检测结果. 雪花、水面倒影这些影响运动检测算法的因素都得到了较好的解决. 但是, 由于本系统目前仅考虑检测问题, 没有对行人进行跟踪, 因此, 当两个目标距离比较靠近时, 检测窗口合并算法会将多个目标合并为一个 (图 7 (c), 第 3 列). 图 7 (d) 给出一组不同场景下的行人检测结果, 考虑到场景的复杂性和行人样本的多样性, 离线训练一个适用于所有场景的行人检测器将十分困难, 而本文提出的结合背景建模与在线学习的思路则取得了较好的结果.

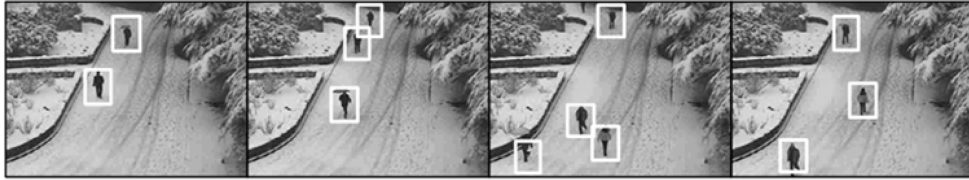
3.4 主动摄像机室外中、远距离行人检测结果

采用主动摄像机, 从大范围场景中主动检测感兴趣的目标是智能视频监控领域的重要研究方向, 有



(a) 运动阴影、车辆的检测结果

(a) Detection results under cast shadow and moving vehicle



(b) 雪天的检测结果

(b) Detection results under snow



(c) 雨天的检测结果

(c) Detection results under rain



(d) 多个不同场景下的检测结果

(d) Detection results under different real surveillance scenes

图 7 不同季节、天气、场景中的行人检测结果

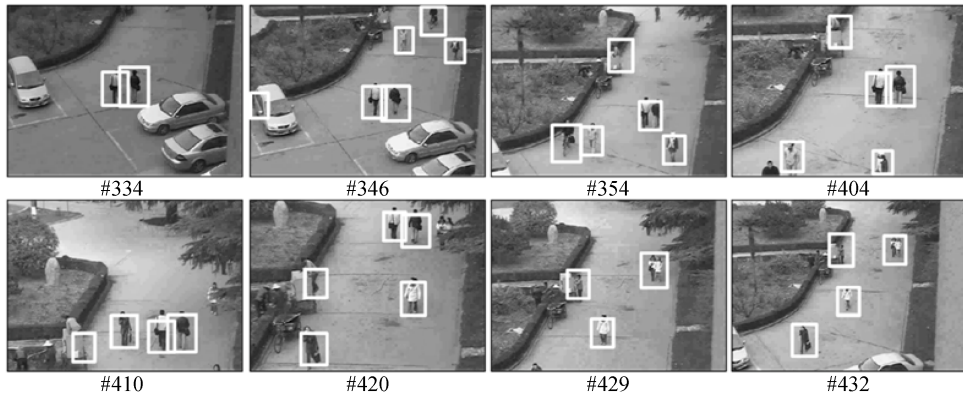
Fig. 7 Detection results under different seasons, weathers and scenes

着广阔的应用前景. 目前, 运动平台下的目标检测, 典型的解决思路是通过图像配准技术消除全局运动, 然后用背景建模的思路提取运动目标. 考虑到主动摄像机可能出现的旋转、缩放、亮度变化等因素, 图像配准结果往往存在误差, 从而导致提取的运动前景中会包含大量噪声. 由于行人检测器实质是一个分类器, 当训练结束后, 其检测结果与载台或摄像机运动无关. 基于此, 本文将行人检测器进一步推广到运动载台上的自动目标检测.

实验中, 先采用文献 [21] 中给出的基于不变特征的图像序列配准算法, 估计序列图像间变换模型参数, 镶嵌生成监控场景的全景图, 并用全景图中采

集的负样本训练了一个行人检测器. 图 8 给出该检测器在摄像机在场景内转动条件下的行人检测结果. 视频序列包含摄像机平移 (图 8(a), #334, #346)、缩放 (图 8(a), #354, #404, #410, #429) 等情况, 可以看出, 算法能够准确地检测出场景内的行人, 检测性能不受摄像机运动的干扰.

图 8(b) 给出一组主动摄像机对特定目标 (图 8(b), 黑色矩形框) 跟踪拍摄视频的行人检测结果. 从图中可以看出, 算法能在摄像机转动 (图 8(b), #528, #534, #536)、缩放 (图 8(b), #515, #518, #520) 以及背景动态变化等情况下稳定、准确地检测出感兴趣的目标, 具有良好的稳定性和扩展性. 需要指出



(a) 摄像机平移、旋转和缩放下行人检测结果

(a) Pedestrian detection results under camera pan, tilt and zoom



(b) 特定行人检测结果

(b) Detection results of a specific pedestrian

图 8 摄像机运动条件下的检测结果

Fig. 8 Pedestrian detection results with moving camera

的是,当场景内存在运动目标时,目前的拼接算法会将该目标引入全景图,从而被误检为负样本. 由于在训练时, BMAT 算法采用基于正样本错误率的 Adaboost 训练, 因此, 引入的负样本不会影响检测器对“训练正样本”的检测率, 但这部分“难分类样本”会引起最终级联分类器的级数的增加, 并会在一定程度降低分类器的泛化能力, 使对行人的检测率降低. 进一步的研究拟将背景建模与视频序列配准相结合, 在摄像机转动时, 用背景图像对全景图进行更新, 消除运动目标造成的干扰.

4 结论

针对复杂场景中的行人检测问题, 通过将背景建模与统计学习相结合, 提出一种基于背景建模与快速 Adaboost 训练的行人检测新算法. 算法从背景建模得到的场景参考背景上产生负样本, 充分利用了监控场景的固有信息, 克服了动态场景的复杂性和行人样本的多样性之间的矛盾, 大幅度减少了训练分类器所需的负样本数量, 提高了检测算法在

该场景下的鲁棒性. 在分类器训练上, 提出了一种快速弱分类器选择算法, 根据正、负样本特征取值的分布和期望的检测率, 直接求解特征取值的阈值范围, 能够满足在线训练和更新检测器的要求; 根据级联分类器的特点, 提出了一种基于正样本错误率的训练算法, 首先根据正样本加权错误率选择弱分类器, 使得组成强分类器所需的弱分类器个数大幅降低, 减少了 Adaboost 的训练轮数, 加快训练速度, 然后调整最终分类器的加权系数, 在保证检测率的同时降低虚警率. 实验结果证明了算法的有效性.

References

- 1 Collins R T. A System for Video Surveillance and Monitoring: VSAM, Final Report, Technical Report CMU-RI-TR-00-12, Carnegie Mellon University, USA, 2000
- 2 Brown L M. View independent vehicle/person classification. In: Proceedings of the 2nd International Workshop on Video Surveillance and Sensor Networks. New York, USA: ACM, 2004. 114–123
- 3 Wang L, Hu W M, Tan T N. Recent developments in human motion analysis. *Pattern Recognition*, 2003, **36**(3): 585–601

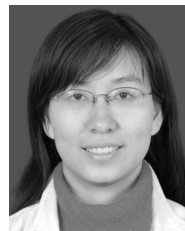
- 4 Gavrilu D M. Multi-cue pedestrian detection and tracking from a moving vehicle. *International Journal of Computer Vision*, 2007, **73**(1): 41–59
- 5 Viola P, Jones M, Snow D. Detecting pedestrians using patterns of motion and appearance. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Paris, France: IEEE, 2003. 734–741
- 6 Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the Computer Vision and Pattern Recognition. Kauai, USA: IEEE, 2001. 511–518
- 7 Mohan A, Papageorgiou C, Poggio T. Example-based object detection in images by components. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(4): 349–361
- 8 Ramoser H, Schlogl T, Belezna C, Winter M, Bischof H. Shape-based detection of humans for video surveillance applications. In: Proceedings of the International Conference on Image Processing. 2003. 1013–1016
- 9 Gavrilu D M, Giebel J. Shape-based pedestrian detection and tracking. *IEEE Intelligent Vehicle Symposium*. 2002. 8–14
- 10 Bertozzi M, Broggi A, Chapuis R. Shape-based pedestrian detection and localization. *IEEE Intelligent Transportation Systems*, 2003, **1**: 328–333
- 11 Sabzmejdani P, Mori G. Detecting pedestrians by learning shapelet features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007
- 12 Wu B, Nevatia R. Detection and segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. *International Journal of Computer Vision*, 2009, **82**(2): 185–204
- 13 Dalai N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 886–893
- 14 Leibe B, Seemann E, Schiele B. Pedestrian detection in crowded scenes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 878–885
- 15 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- 16 Mikolajczyk K, Schmid C, Zisserman A. Human detection based on a probabilistic assembly of robust part detectors. In: Proceedings of the 8th European Conference on Computer Vision. Prague, Czech Republic: 2004. 69–81
- 17 Bertozzi M, Broggi A, Fascioli A, Graf T, Meinecke M M. Pedestrian detection for driver assistance using multiresolution infrared vision. *IEEE Transactions on Vehicular Technology*, 2004, **53**(6): 1666–1678
- 18 Cheng H, Zheng N N, Qin J J. Pedestrian detection using sparse Gabor filter and support vector machine. In: Proceedings of IEEE Intelligent Vehicles Symposium. Vienna, Austria: IEEE, 2005. 583–587
- 19 Wu B, Nevatia R. Simultaneous object detection and segmentation by boosting local shape feature based classifier. In: Proceedings of the IEEE Conference on Computer Vision and Patter Recognition. Minneapolis, USA: IEEE, 2007
- 20 Stauffer C, Eric W, Grimson L. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, **22**(8): 747–757
- 21 Li J, Yang T, Pan Q, Cheng Y M, Hou J. A novel algorithm for speeding up keypoint detection and matching. *International Journal of Image and Graphics*, 2008, **8**(4): 643–661



杨 涛 西北工业大学计算机学院博士后。2008 年获西北工业大学控制理论与控制工程专业博士学位。主要研究方向为视频内容分析、视频目标检测与跟踪、图像配准。本文通信作者。

E-mail: yangtaonwpu@163.com

(**YANG Tao** Postdoctor at the School of Computer Science, Northwestern Polytechnical University. He received his Ph.D. degree from the School of Automation, Northwestern Polytechnical University in 2008. His research interest covers video content analysis, visual object detection and tracking, and image registration. Corresponding author of this paper.)



李 静 西安电子科技大学通信工程学院讲师。2008 年获西北工业大学控制理论与控制工程专业博士学位。主要研究方向为图像处理与模式识别、图像配准、文字定位与识别。

E-mail: jinglixid@mail.xidian.edu.cn

(**LI Jing** Lecturer at the School of Telecommunications Engineering, Xidian University. She received her Ph.D. degree from the School of Automation, Northwestern Polytechnical University in 2008. Her research interest covers image processing and pattern recognition, image registration, and text location and recognition.)



潘 泉 西北工业大学教授。主要研究方向为信息融合理论及应用、多目标跟踪、智能信息处理。

E-mail: quanpan@nwpu.edu.cn

(**PAN Quan** Professor at Northwestern Polytechnical University. His research interest covers theory and application of information fusion, multi-target tracking, and intelligence information processing.)



张艳宁 西北工业大学教授。主要研究方向为计算机视觉、图像与视频处理、智能信息处理、模式识别。

E-mail: ynzhang@nwpu.edu.cn

(**ZHANG Yan-Ning** Professor at Northwestern Polytechnical University. Her research interest covers computer vision, image and video signal processing, intelligence information processing, and pattern recognition.)