

基于特征融合的行人检测技术^①

徐琳, 张明

(上海海事大学 信息工程学院, 上海 201306)

摘要: 首先研究了三种不同的特征算子在基于图像行人检测中的应用. 他们分别是: 梯度直方图(Histogram of Oriented Gradient, HOG)、局部三值模式特征(Local Ternary Patterns, LTP)以及改进了的局部三值模式特征(Sqrt Local Ternary Patterns, S-LTP). 对以上三种特征算子进行了实验比较, 最后将 HOG 和 S-LTP 算子融合得到 HOG+S-LTP 的基于多特征的行人检测算子, 利用 SVM 分离器在 INRIA 人体库上进行了实验, 实验表明, 融合后的特征显著地提高了行人检测率, 同时也满足实时性要求.

关键词: 行人检测; 梯度直方图(HOG); LTP; 支持向量机(SVM)

Pedestrian Detection Technology Based on Features Fusion

XU Lin, ZHANG Ming

(School of Information Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: First of all, we studied three different types of feature operators in the application of image-based pedestrian detection. They are Histogram of Oriented Gradient(HOG), Local Ternary Patterns(LTP) and the Sqrt Local Ternary Patterns(S-LTP). Through conducting experiments to compare these three feature operators, we combined the algorithms of HOG and S-LTP and proposed a new human detection feature operator named HOG+S-LTP. Finally, by using a linear SVM as the classifier, we compared the new feature with the other single features in INRIA person dataset. Results of experiments showed that the new combined feature operator significantly improves the pedestrian detection rate, and it also meets real-time requirements.

Key words: human detection; HOG; LTP; SVM

行人检测技术即判断一幅图像中是否包含行人以及将其中的行人部分定位出来, 其可广泛应用于行人流量统计、机器人开发、车辆辅助驾驶系统、视频监控等领域. 由于人体姿态各异, 服装变化多端, 行人检测技术已成为计算机视觉领域的中一个既具有研究价值又极具挑战性的热门课题.

随着最近几年的研究进展, 对于行人检测形成了两种主要的研究方法: 一种是基于部件特征的方法, 即将人体分为头部、躯干、四肢三部分, 采用模板匹配的方法检测每个部分从而找到整个行人; 第二种是基于全局特征的方法, 也是近些年的主流方法. 这种方法采用统计得到行人特征, 再利用模式识别中的方

法(SVM, AdaBoost 等)进行分类, 图 1 描述了该方法的一般流程.

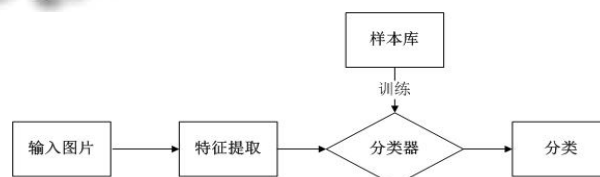


图 1 行人检测流程

特征提取方面主要采取边缘特征、统计特征、形状特征或者变换特征等各类静态特征来描述行人, 最具代表性的特征有 HOG 特征、Haar 小波特征、Edgelet

^① 收稿时间:2015-02-02;收到修改稿时间:2015-04-17

特征、Shapelet 特征等. 2005 年 Dalal 和 Triggs 提出梯度直方图(Histogram of Oriented Gradient, HOG)^[1]概念, 并将其应用于行人检测领域, 在 MIT 数据库^[2](该行人数据库背景简单, 识别难度小)上获得了将近 100% 的检测率; 在包含光照、视角和背景变化的 INRIA 行人数据库^[1](该行人数据库背景、行人姿态复杂, 相对于 MIT 数据库识别难度很大)上也取得了 90% 左右的检测率. HOG 特征能够很好地描述局部梯度信息, 具有不错的检测效率, 已成为当前行人检测技术的主流方法. 随后很多学者对 HOG 特征方法进行了改进, 使其得到了一定的改善, 比较著名的有 2006 年 Zhu^[3]等人提出积分直方图的概念, 通过计算多尺度 HOG 特征, 并且结合 AdaBoost 算子构建了级联分类器, 得到了一个快速的行人检测系统; Qu 等^[4]学者提出了无背景 HOG 特征的方法, 不仅消除了一般情况下背景元素对 HOG 特征的影响, 而且增加了 HOG 特征提取的速度, 通过实验的方法证明了在处理视频中的行人检测时, 其效率高于传统 HOG 特征的方法; 2009 年 Wang 等人^[5]将应用于人脸识别领域中的局部二值特征(LBP)与 HOG 特征相结合, 得到一种基于 HOG-LBP 特征的行人检测方法, 在一定程度上克服了 HOG 特征的缺陷, 并较大地提高了行人检测的准确度. 对于分类器的选取, 主要包括支持向量机(SVM)、AdaBoost 等进行分类训练. 本文将对 HOG 特征、LTP 特征以及对 LTP 改进后的 S-LTP 进行比较, 最后将 HOG 和 S-LTP 相融合, 实验证明, 融合后的特征应用于行人检测得到的正确率更高, 时间效率上也满足实时性要求.

1 特征提取

特征的选取是行人检测技术重要的环节, 特征提取的好坏直接影响下一步的检测效果, 单一特征虽然检测效率高, 但获取的特征信息有限, 无法较完整地描述整个行人特征, 若把两个甚至多个单一特征相融合, 将大大提高检测准确性.

1.1 HOG 特征

HOG 特征能够很好的刻画出图像的边缘特性, 而且对小的位置偏移和光照不敏感, 其已成为使用最为广泛的行人特征描述子. Dalal 将一张图像用一个 16×16 的 Block、采用 8 个像素为步长进行遍历, 再将每一个 Block 分为 4 个 8×8 的 Cell, 在 $0^\circ \sim 180^\circ$ 之间以 20° 为步长统计每个 Cell 中各个像素的梯度值, 得到一

个 9 维的梯度方向直方图; 那么 4 个 Cell 将串联为一个 36 维的梯度方向直方图; 接着以 L2 范数进行归一化; 最后一步把整张图片的所有 Block 的梯度直方图特征串联, 得到样本图像的 HOG 特征描述. 在计算梯度值时, Dalal 通过比较不同的梯度掩码, 实验发现一维梯度算子 $[-1, 0, 1]$ 计算得出的效果最为理想. 对于图像中的某个点 (X, Y) , 其梯度值定义为公式(1):

$$R(X, Y) = \sqrt{(I(X+1, Y) - I(X-1, Y))^2 + (I(X, Y-1) - I(X, Y+1))^2} \quad (1)$$

梯度方向为公式(2):

$$\text{Ang}(X, Y) = \arccos(I(X+1, Y) - I(X-1, Y) / R) \quad (2)$$

Dalal 比较了不同的单元格(Cell)、区间(Bin)、块(Block)大小对实验结果的影响, 最后得到采用 6×6 像素的 Cell, 3×3 个 Cell 组成的 Block, $[0, \pi]$ 的方向上划分成 9 个区间所得到的检测效果最好. 本文在提取 HOG 特征时使用了这类参数.

1.2 LTP 特征和改进的 S-LTP 特征

局部三值模式特征(Local Ternary Patterns, LTP)继承了局部二值模式特征(Local Binary Patterns, LBP)^[6]图像纹理描述能力强、计算速度快等优点, 并且对噪声相对不敏感, 判别能力更强. LBP 特征是由 Timo Ojala 等提出的一种刻画纹理信息的描述算法, 一开始主要应用于人脸检测中. 其定义为在 3×3 窗口内, 将中心点像素的灰度值作为阈值, 把周围 8 邻域像素的灰度值与该阈值相比较, 小于该阈值的像素点的位置标记为 0, 否则标记为 1. 经过这样的处理之后, 中心点周围 8 邻域像素点就可以得到一个 8 位的二进制编码, 再赋予不同位置不同的权重就可以将该二进制编码转换为一个整数, 就可以得到该中心点像素的 LBP 值. 如图 2 所示.

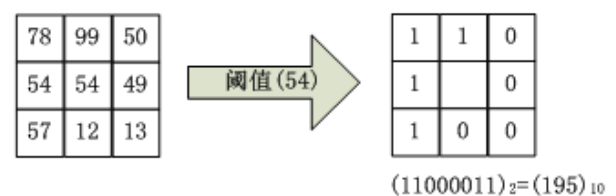


图 2 LBP 编码示意图

参考文献[7]提出把 LBP 的二值编码方式改进为三值编码, 提出了局部三值模式(LTP)的概念. 由于在提取 LBP 特征时直接将中心点与周围的其它像素点相比较, 导致对噪声比较敏感, 所以提取 LTP 特征时在中心点像素值附近设置一个范围 $(-t, +t)$, 当中心像素点

的灰度值与周围像素点灰度值相比较的结果处于该范围时, 则把周围该像素点的值量化为 0, 小于 -t 则量化为 -1, 否则量化为 +1. 其计算公式如公式(3)所示:

$$s(m, i, t) = \begin{cases} 1, & m - i > t \\ 0, & |m - i| < t \\ -1, & m - i < -t \end{cases} \quad (3)$$

其中 i 表示中心像素点灰度值, μ 表示邻域像素点灰度值, $s(\mu, i, t)$ 为量化后的值. 这样 LTP 编码能够对噪声有一定的鲁棒性. 当阈值 t 设置为 5 时(该阈值由实验经验所得, 当 t 设置为 5 时实验效果最佳, 后面的实验将采用该阈值), LTP 特征编码如图 3 所示.

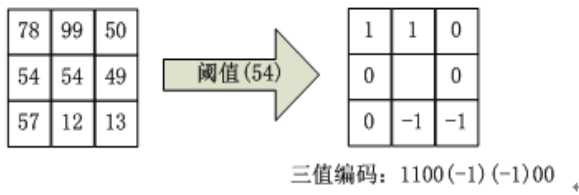


图 3 LTP 编码示意图

为使 LTP 特征算子计算简单, 可以将其编码过程分解成图 4 所示的两个正负 LBP 编码, 先将两个部分分别计算, 再把它们合并起来就可以得到最后的值.

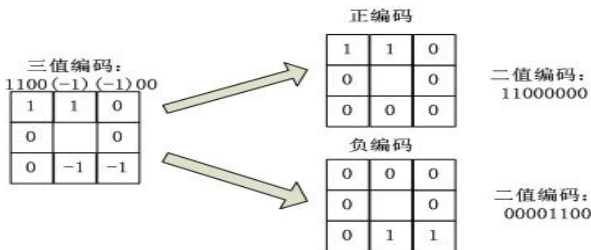


图 4 LTP 正负编码示意图

本文在提取图像的 LTP 特征时, 首先将彩色图像处理成灰度图像, 再对其进行预处理, 最后按照以上介绍的方法提取 LTP 特征. 另外又对提取到的 LTP 特征值进行开二次方运算得到 S-LTP 特征作实验对比. 具体步骤如下:

步骤 1: 伽马校正(Gamma Correction), 对图像明亮区域作压缩处理, 以提高暗部区域的对比度, 也就是减少光照对图片的影响.

步骤 2: 高斯差异滤波(Difference of Gaussian Filtering, DoG)^[7], 上一步处理并不能彻底消除阴影部分对整个图像灰度信息的影响, 通过高通滤波原理能够去除部分干扰信息, 低通滤波可以去除空间高频信

息, 减少衍射及噪声对图像信息识别的影响.

步骤 3: 对比度均衡化, 把整个图像重新归一化到一个适度的灰度值区间内. 采用公式(4)和公式(5)对图像进行处理:

$$I(x, y) \leftarrow \frac{I(x, y)}{(\text{mean}(|Ix', y'|^a))^{\tau/a}} \quad (4)$$

$$I(x, y) \leftarrow \frac{I(x, y)}{(\text{mean}(\min(\tau, |Ix', y'|^a))^{\tau/a}} \quad (5)$$

其中 τ 是阈值, 用来去除归一化后图像中比较大的数值; α 是压缩因子, 用来减少图像中像素点数值较大地的影响. 在后续的实验, 这里设定 $\tau=10, \alpha=0.1$.

步骤 4: 求取 LTP 特征或者改进后的 S-LTP 特征, 通过各个步骤对图像进行预处理之后, 就可以采用之前介绍的 LTP 特征计算方式来提取. 在提取 LTP 特征实验中我们发现归一化处理后所得到的 LTP 特征值在 0 到 1 之间, 且数值很小, 因此对于图像纹理信息的描述能力不够强. 从而我们对这些小数作开二次方根处理将那些接近 0 的微小特征值放大, 以提高那些特征值在整体特征值中的比重. 得到的特征就是 S-LTP 特征.

1.3 特征融合方法的选取

对于使用 SVM 作为分类器的特征融合方法主要分为两种(如图 5 所示): 方法一是对两种特征各自训练一个分类器, 然后采用加权累积方式形成一个最终分类器; 方法二是将两种方法得到的特征向量串联成一个综合的特征, 再用分类器进行训练. 本文对这两种融合方法进行了实验对比, 得出方法二要优于方法一, 因此在后面的实验阶段将采用方法二来对两种特征进行融合. 首先按照 2.1 节介绍的提取方法来提取 HOG 特征得到相应的特征向量, 再通过 2.2 介绍的方法步骤计算得到 S-LTP 特征向量, 最后将以上两种特征向量串联起来进行分类训练.

2 分类器SVM的使用

支持向量机(SVM)是由 Cortes 和 Vapnik^[8]提出来的, 其将原始目标函数准变为对偶目标函数的二次规划问题, 保证了解的唯一性与全局最优性, 能够很好地解决传统方法中局部最小和过学习等问题; SVM 在解决非线性、小样本、高维模式识别问题中也有很多特有的优势. 其基本思想如图 6 所示.

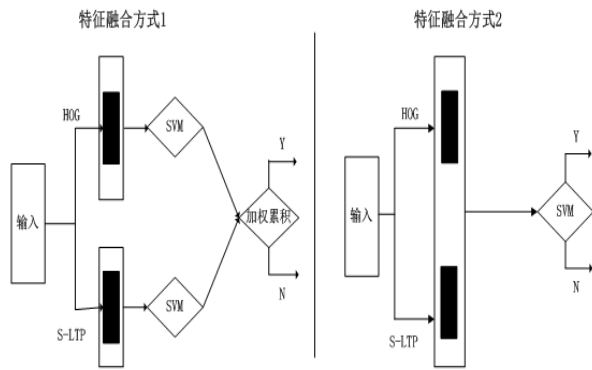


图 5 两种特征融合方式

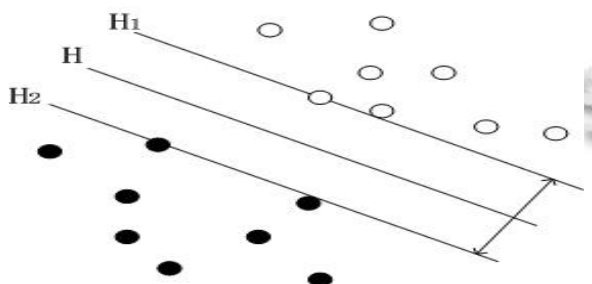


图 6 最优超平面分类示意图

其中直线 H 表示两类样本的分界线，直线 H1、H2 分别表示过各类中离分类线最近且平行于分类线的直线。SVM 不仅能够准确地分开这两类样本，而且可以使直线 H1 和 H2 间的距离最大，这样既保证经验风险最小(训练误差为 0)，又使置信范围最小，从而真实风险最小。公式(6)定义了超平面：

$$F(x) = \beta^T x + \beta^T x_0 \quad (6)$$

其中 β 叫做权重向量， $\beta^T x_0$ 叫做偏置。最优超平面有无数种表达方式，即通过任意的缩放 β 和 $\beta^T x_0$ ，习惯上用公式(7)来表达最优超平面：

$$|\beta^T x + \beta^T x_0| = 1 \quad (7)$$

式中 x 表示离超平面最近的那些点。使用 SVM 训练样本的过程如下：

- (1)提取正负样本特征；
- (2)投入 SVM 分类器训练，得到 model；
- (3)通过 model 生成检测子；
- (4)采用检测子检测负样本，得到 hardexample；
- (5)提取 hardexample 中的特征并将步骤一中的特征一起训练，得到最终检测子。

3 实验结果与分析

实验数据库采用 INRIA 人体数据库，其中涵盖了

64*128 大小的正负图像集。此样本库中行人的姿态、服装、视角和光照等因素变化很大，是一个难度相对很高的数据库。本实验采用 2416 个正样本和 5000 个负样本作为训练集，使用不同于训练集的 1126 个正样本和 5000 个负样本作为测试集，如表 1 所示。将 SVM 作为分类器，并使用 OpenCV 提供的接口，采用 VS2010 编译器。实验比较了以上介绍的单一特征以及融合后多特征的检测效果，为使结果具有可比性，将分类器 SVM 的阈值和迭代次数等参数保持一致。

表 1 训练、测试样本数

| | 正样本 | 负样本 |
|-----|------|------|
| 训练集 | 2416 | 5000 |
| 测试集 | 1126 | 5000 |

表 2 各特征行人检测性能比较

| | HOG | LTP | S-LTP | HOG+S-LTP |
|------------|------|------|-------|-----------|
| 召回率(%) | 82.5 | 88.6 | 90.7 | 93.5 |
| 准确率(%) | 91.2 | 95.4 | 96.8 | 98.7 |
| 平均检测时间(ms) | 642 | 549 | 604 | 928 |

实验结果采用召回率和准确率以及时间效率上来比较分类的性能。从表 2 实验数据可以看出，在召回率方面，LTP 和 S-LTP 特征分别比 HOG 特征高出 6.1% 和 8.2%；在准确率方面，LTP 和 S-LTP 特征比 HOG 特征分别高出 4.2% 和 5.6%，主要是因为 HOG 特征对图像的光照变化特别敏感，这两种方法能够很好的描述图像的纹理特征，克服了光照变化的影响；S-LTP 特征相比于 LTP 特征在准确率上略有提高是由于在对 LTP 归一化后的特征值开平方处理后，增大了接近于零的那些特征值，从而提高了图像纹理的描述能力，减少了错检率；而融合后的 HOG+S-LTP 特征在以上两个指标中检测效果最好；从时间效率上来看，LTP 特征方法用时最少，改进后的 S-LTP 方法用时略增于 LTP，但相比于 HOG 方法，用时都要少；而融合后的 HOG+S-LTP 方法由于特征向量维数的增加，平均用时增加至 928ms，但依旧能够满足实时性要求。

图 7 是各个特征在行人数据库上检测结果的 ROC 曲线，从这幅图可以看到，基于 HOG+S-LTP 特征的 ROC 曲线最靠近右上角，说明了其性能优于另外三个基于单特征检测方法。改进后的 S-LTP 特征检测效果优于 LTP 特征，是因为对归一化后的 LTP 特征值开二次方后，增加了图像纹理特征的描述能力。

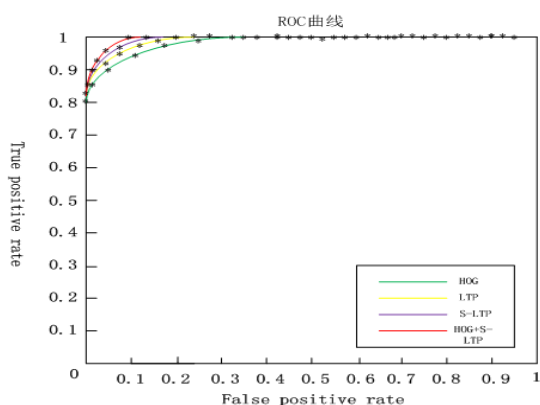


图 7 基于各个特征的行人检测 ROC 曲线

本文特征融合后的 HOG+S-LTP 方法, 不仅适用于背景简单的图像, 对于背景复杂且出现多个行人的图像也能表现出很好检测效果, 如图 8 中第一行和第二行所示。



图 8 实验部分效果图

图 8 中的第一行和第二行图像均采用了本文的 HOG+S-LTP 方法, 其中第一行图像背景单一, 行人明显, 该方法基本能够精确检测出行人. 第二行图像中虽然出现多个行人且背景复杂, 但该方法依然表现出很强的鲁棒性. 第三行使用与第二行同样的图像, 但其采用了论文基于多特征的行人检测技术研究^[9]中的方法, 通过二、三两行的实验结果可以发现, 本文所提出的融合方法检测效果更佳, 第三行文献^[9]所提出的方法在检测效果中出现了漏检以及误检现象。

5 结束语

本文介绍了 HOG 特征、LTP 特征以及改进后的 S-LTP 特征, LTP 特征解决了 HOG 无法刻画图像纹理特征的缺点, 提出了 S-LTP 的概念. 实验证明改进后的 S-LTP 方法进一步提高了行人检测性能, 由于在提取 LTP 特征时归一化处理后的特征值在 0 到 1 之间, 将 0 到 1 之间的小数作开二次方根处理, 增大了接近 0 的那些特征值, 从而能够更好地描述纹理特征. 将其与 HOG 特征融合后的方法更是提高了行人检测准确率, 虽然时间效率略有降低, 但依旧可以满足实时性要求。

参考文献

- 1 Triggs DN. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition*, 2005, (1), 886–893.
- 2 Papageorgiou C, Poggio T. A trainable system for object detection. *International Journal of Computer Vision*, 2000, 38(1): 15–33.
- 3 Zhu Q, Avidan S, Yeh MC, et al. Fast human detection using a cascade of histograms of oriented gradients. *Computer Vision and Pattern Recognition*, 2006, (2): 1491–1498.
- 4 Qu J, Liu Z. Non-background HOG for pedestrian video detection. *Proc. 8th IEEE International Conference on Natural Computation*. Chongqing, China. IEEE Press. 2012. 535–539.
- 5 Wang XY, Han TX, Yan SC. An HOG-LBP human detector with partial occlusion handling. *IEEE International Conference on Computer Vision*. Kyoto, Japan. 2009. 32–39.
- 6 Tuzel O, Porikli FM, Mcerr P. Pedestrian detection via classification on Riemannian manifolds. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2008, 30 (10): 1713–1727
- 7 Tan XY, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. on Image Processing*, 2010, 19(6): 1635–1650.
- 8 Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, 1995, 20(3): 273–297.
- 9 张春风. 基于多特征的行人检测技术研究[学位论文]. 杭州, 浙江工业大学, 2013.