

基于稀疏表达和机器学习的行人检测技术研究

王 坚 兰 天

(中央财经大学信息学院 北京 100081)

摘 要 针对行人检测技术在智能交通系统中的应用,为了提高行人检测方法的有效性、实时性和准确性,将稀疏表达应用到图像的特征压缩中,提出一种基于 HOG 和 LTP 特征训练 SVM 分类器进行行人检测的方法。基于 HOG 和 LTP 特征训练 SVM 分类器进行行人检测的方法有效地结合了图像的梯度特征和纹理特征,利用稀疏表达进行特征数据的压缩可以有效地加速算法。实验结果表明,提出的算法具有精度高、速度快等优点。

关键词 稀疏表达,行人检测,LTP,HOG,SVM,图像处理

中图法分类号 TP391.9 文献标识码 A

Study on Pedestrian Detection Based on Sparse Representation and Machine Learning

WANG Jian LAN Tian

(School of Information, Central University of Finance and Economics, Beijing 100081, China)

Abstract According to the application of pedestrian detection technology in the intelligent transportation system, in order to improve the efficiency, real-time and accuracy of pedestrian detection method, in this paper, the sparse representation was applied to the feature compression of the image, and a new method of pedestrian detection based on HOG and LTP feature training SVM classifier was proposed. Training SVM classifier for pedestrian detection based on the characteristics of HOG and LTP effectively combines the image gradient feature and texture features and takes advantage of the sparse expression on data compression which can effectively speed up the algorithm. Experimental results show that the proposed algorithm has the advantages of high precision and speed.

Keywords Sparse representation, Pedestrian detection, LTP, HOG, SVM, Image processing

近年来,随着图像处理技术的飞速发展,行人检测逐渐成为一个热门的研究课题。行人检测以其良好的应用前景和重要的理论意义被越来越多的研究者所重视^[1]。直接从图像的像素点中找到行人基本不可能,所以从最初的将 Haar 特征和灰度特征应用于行人检测发展到最近应用比较多的利用 HOG 特征检测行人,再发展到将各种特征综合应用于检测行人,虽然检测方法各有不同,但是行人检测的基本原理不变,都是先提取图像的特征,然后根据特征进行行人检测。而如何提取特征、如何表征特征、如何利用特征建立分类器、如何评估分类器的好坏等就需要应用很多其他领域的知识。本文结合已有的行人检测方法,提出一种改进的低秩分解(Go De-compressing)^[2]的方法进行图像前景提取,基于 HOG^[3]特征和 LTP^[4]特征建立 SVM 分类器进行行人检测,此方法满足准确性、鲁棒性和高准确性的要求,而且很好地结合应用了稀疏表达和机器学习的知识。

1 图像前景提取的方法研究

本文提出的方法是将一段时间内的视频图像作为向量(实验中取 200 帧)存入一个矩阵,然后利用改进低秩分解方

法分解得到的背景部分 L (低秩部分)、前景部分 S (稀疏部分)、噪声部分 N 。低秩部分对应的是图像背景,而前景部分由尖峰噪声来表示^[5],行人在图像中的位置也可以理解为尖峰噪声,这样排除了背景,只需要对尖峰噪声进行行人检测,从而可以提高速度和检测精度。

低秩分解可以高效和可靠地从 X 中估计低秩部分 L 、稀疏部分 S 和噪声部分 G ,即: $X=L+S+G$ 。其原理是:首先用 $X-L$ 去逼近 S ,得到 S 的优化值,然后用 $X-S$ 去逼近 L ,得到 L 的进一步优化值^[6]。由于低秩部分可以通过 SVD 方法求得,但是速度很慢,因此需要加速算法。加速是通过 BRP (Bilateral Random Projection) 和指数修正实现的。

算法流程如下:

输入: X, r, q, ϵ, t

输出: L, S

初始化: $L_0 := X, S_0 := 0, t := 0$

While $\|X - L_t - S_t\|_F^2 / \|X\|_F^2 > \epsilon$ do

$t := t + 1$;

$\tilde{L} = [(X - S_{t-1})(X - S_{t-1})^T]^q (X - S_{t-1})$

$Y_1 = \tilde{L} A_1, A_2 = Y_1$

本文受中央财经大学重点学科建设项目资助。

王 坚(1975—),男,博士,讲师,主要研究方向为模式识别、数据挖掘, E-mail: wanderingful@126.com; 兰 天(1987—),男,硕士,工程师,主要研究方向为数据挖掘。

$$Y_2 = \tilde{L}^T Y_1 = Q_2 R_2, Y_1 = \tilde{L} Y_2 = Q_1 R_1;$$

If $\text{rank}(A_2^T Y_1) < r$ then, $r = \text{rank}(A_2^T Y_1)$, go to the first step

End;

k largest entries of; end while;

当目标值 $\|X - L - S\|_F^2$ 收敛于局部极小值, 则 L 和 S 线性收敛于局部最优。在 $X = L + S + N$ 中, 解出低秩部分 L 、稀疏矩阵 S 、噪声部分 N , 如图 1 所示。对应于图像中就是 L 为背景部分, S 为前景部分, N 为噪声部分。

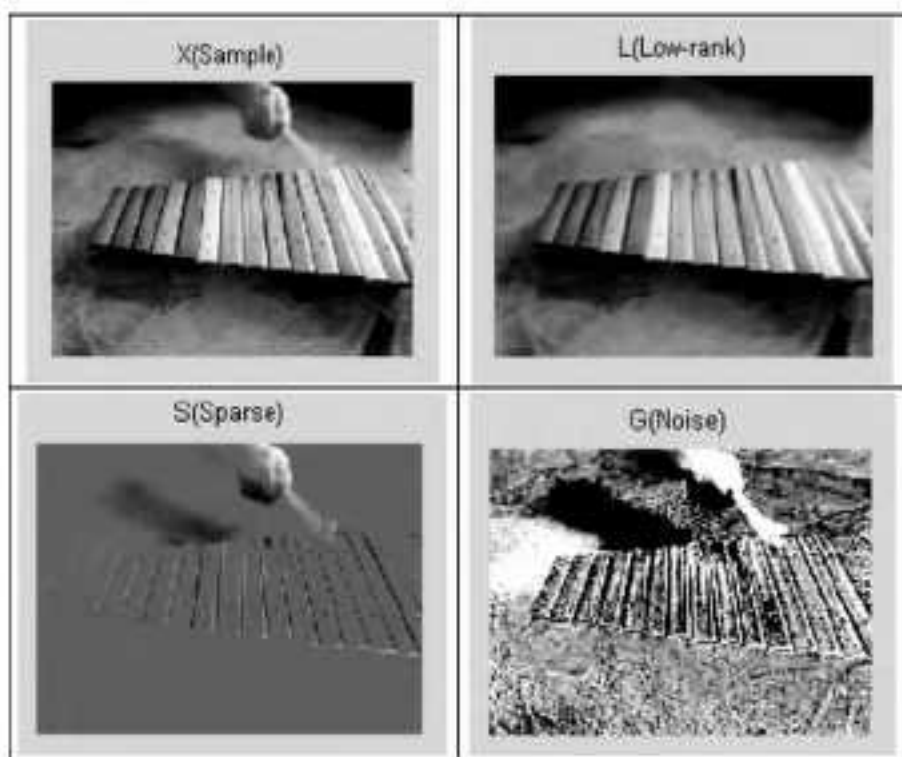


图 1 低秩分解完成对图像低秩部分的提取

2 基于稀疏表达的图像特征数据压缩

2.1 信号的稀疏表达

稀疏表达可以理解为一个原始信号用一组字典元素的线性组合进行表示。典型的稀疏表达式为: $x = D \times a$, 其中 x 为原始信号, D 是字典, a 就是原始信号在字典上的表达。作为稀疏表达的一个分支, 压缩感知向传统的 Nyquist-Shannon 抽样定理发起了挑战, 因为它可以极大地降低采样频率。将一段时间内的信号中的每一个作为一个向量输入到矩阵中, 进行矩阵分解, 可以得到其低秩部分、稀疏部分和噪声部分。

一般对于自然信号 x , 其本身并不是稀疏的, 但是其在某种稀疏基上可能是稀疏的, 所以需要在某种稀疏基上对 x 进行稀疏表示, 如图 2 所示, $x = \Psi s$, Ψ 为稀疏基矩阵, s 为稀疏系数 (s 只有 K 个是非零值 ($K \ll N$))。压缩感知方程可以表示为: $y = \Phi x = \Phi \Psi s = \Theta s$ 。将原来的测量矩阵 Φ 变换为 $\Theta = \Phi \Psi$ (传感矩阵), 解出 s 的逼近值 s' , 则原信号 $x' = \Psi s'$ 。压缩感知的具体步骤主要可以分为 3 个部分: 信号的稀疏表达、信号的观测矩阵和信号的重构算法。



图 2 压缩感知的过程

2.2 图像特征数据压缩

由于提取到的图像特征数据的维数过高、数量过大, 不利于存储和进一步的机器学习, 因此需要进行相应的压缩。本文采用 Kaihua Zhang 等^[7]采用的思路, 即应用稀疏表达。采用了随机测量矩阵, 其不仅满足 RIP, 而且满足高速型的要求。

最典型的随机测量矩阵是随机高斯矩阵 $R \in R_{n \times m}$, 其中 $r_{ij} \in N(0, 1)$ 。但是由于矩阵是稠密的, 当 m 比较大时, 内存

和计算量仍然很大。因此采用了一个非常稀疏的测量矩阵, 矩阵的数值取值由概率参数 s 确定, 输入为:

$$r_{ij} = \sqrt{s} \times \begin{cases} 1 \\ 0 \\ -1 \end{cases} \quad (1)$$

分别以概率 $\frac{1}{2s}$, $1 - \frac{1}{s}$ 和 $\frac{1}{2s}$ 进行取值。

Achlioptas 证明了上述类型的矩阵在 $s=2$ 或者 3 时满足 RIP。由于只需要一个单位的随机生成器, 因此矩阵比较容易计算。更重要的是, 当 $s=3$ 时, 矩阵非常稀疏, 其 $2/3$ 的计算都可以避免。本文采用 $s = m/4$ ($x \in R_m$) 得到稀疏矩阵。对于每一行 R , 只需要计算其中的 c ($c \leq 4$) 个输入。因此, 计算复杂度仅为 $O(cn)$, 是非常低的。而且只需要存储 R 的非零输入, 这使得内存要求非常低。

3 基于机器学习的行人检测

3.1 LTP 特征

实验中使用 LTP 的两种不同编码模式统计直方图, 采用 8×8 大小的一个 BLOCK 区域进行 LTP 特征的统计。图 3 是从行人训练图像中提取的利用 Upper Pattern 进行编码的行人的 LTP 特征。图 3(b) 是对原图整个图像进行 LTP 特征采用 Upper Pattern 进行编码的统计直方图的结果。

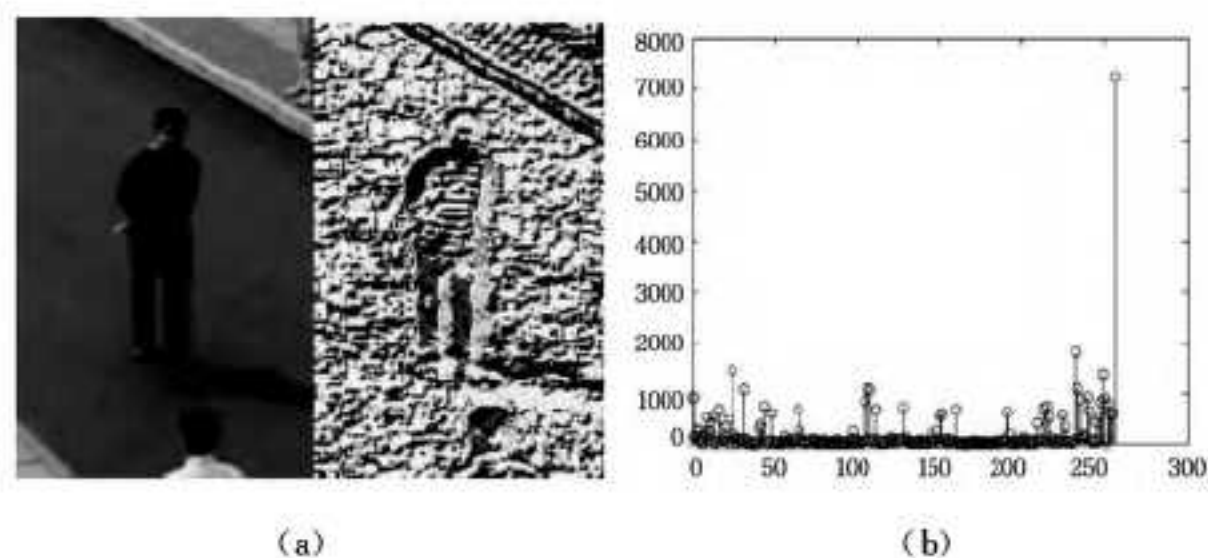


图 3 LTP 对应的特征图和统计直方图

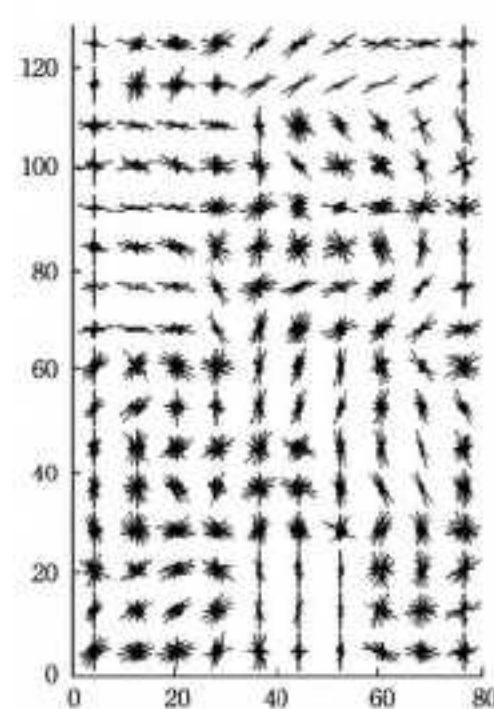
一个 Block 含有 256 个数据, 这些数据是区域内的 LTP 特征的统计直方图。实验中的图像为 128×64 , 步长采用一个 Block 大小, 则总的特征数为 $15 \times 7 \times 256 = 26880$ 。图像的纹理特征还有其他的表达方式, 但是 LTP 可以从像素点与周围像素点的灰度值的关系上进行计算, 从两个方向进行编码, 特征的细致程度更高。但是其弊端是得到的特征维数过高, 需要进行降维。

3.2 HOG 特征

相对于其它描述子, 图像的几何和光学不变性 (除非物体的方向改变) 没有在 HOG 的提取过程中被破坏。因此 HOG 描述子特别适合人的检测。图 4(a) 是行人检测训练图片中的行人, 图 4(b) 是其相应的 HOG 特征。由图 4 可以直观地看出 HOG 特征与行人整体相对应的关系, 在进一步的操作中可以进一步优化并提取其中的行人的明显特征, 并将其应用到实际的行人检测中。HOG 特征中没有旋转和尺度不变性, 因此计算量小; 然而 SIFT 特征中每个 feature 都需要用 128 维的向量进行描述, 因此在计算 SIFT 时所使用的计算量相对较大。



(a)



(b)

图4 原图和其对应的HOG特征图

3.3 训练SVM分类器

将提取到的HOG特征数据作为SVM分类器的输入,训练分类器,寻找一个最优的可以得到最好分类效果的分器,也就是SVM中的最优超平面。本文使用的是台湾林智仁的libsvm软件包^[8],以HOG特征和两LTP特征作为输入,训练行人检测的分类器。在有足够的内存空间的情况下,负类样本准备得越多越好。但是由于实验中的电脑内存有限,没有采用过多的负类样本,样本集为1000个正样本和3313个负样本。本方案中采用的是从每个负样本中随机抽取3个位置的图作为负类。由于有足够多的负样本,因此只取了3个。

3.4 行人检测过程

将表征图像梯度特征的HOG特征和图像纹理特征的LTP特征相结合,发挥二者的优势,并利用稀疏表达的方法进行特征数据的压缩,利用二者相结合训练的SVM分类器进行了行人检测的测试。检测时,首先对图像进行预处理,例如图像仿射纠正几何形变、图像去噪、曝光过度图像的滤除,然后进行前景提取,因为行人在图像中是以前景的形式出现,所以可以得到行人的候选区域,再对输入图片进行尺度缩放,对于每一层图像,采用大小为 128×64 的窗口进行滑动计算,提取出每个滑动窗口的HOG特征和LTP特征,并利用SVM分类器进行判别,可以得到该窗口中是否有目标。如果当前窗口符合行人的HOG特征和LTP特征,就将其存储作为目标区域,如果没有,则检测下一个窗口。

4 实验结果与分析

在提取到HOG特征和LTP特征之后,利用其训练带有高斯核的SVM分类器可以很好地检测行人。下面是用得到的HOG特征和LTP特征训练SVM分类器得到的实验结果,硬件平台为CPU为Intel Pentium Duo,主频为2.0GHz,内存为4GB。利用HOG特征和LTP特征相结合的方法训练的SVM分类器在本文3.3中所述样本集(1000/3313)上的分类结果分别如表1、表2所列。

表1 利用SVM分类器对测试集1进行行人检测的效果

利用测试集1	Accuracy
线性核函数	96.4062
多项式核函数(degree为3)	96.48
RBF核函数	85.138

表2 利用SVM分类器对INRIA测试集进行行人检测的效果

利用INRIA测试集	Accuracy
线性核函数	97.1664

利用多项式核函数得到的SVM分类器在INRIA数据集上获得的检测准确率为97.1664,其训练时间比最快速度的Bo Wu算法缩短了92%,训练时间大大缩短。该方法可以兼顾图像的梯度特征和纹理特征,可以从两个方面对图像进行描述,因此可以得到图像更全面的信息。由于采用了这两种特征相结合的方法,因此得到的行人检测效果相对于其它方法也有了一定的提升。

结束语 本文利用基于HOG特征和LTP特征相结合训练的SVM分类器进行行人检测,该方法可以应用于静止摄像头的监控场景中,将前景图像、背景图像和噪声分离,理论意义和实际应用价值都比较高。基于HOG和LTP特征相结合的SVM分类器进行行人检测部分则可以应用于行人检测系统中,如交通监控、商场或者广场的监控等。前景提取方法可以很好地抵抗噪声,并将图像的3个组成部分:低秩部分、稀疏部分和噪声部分进行有效地分离。实验结果表明提出的算法有精度高、速度快等优点。然而,实验中的前景提取和行人检测是相分离的,如何将二者进行结合将是下一步的研究方向。

参考文献

- [1] Zhou T, Tao D. Randomized low-rank & sparse matrix decomposition in noisy case[C]// International Conference on Machine Learning. 2011, 3:2
- [2] Zhang K, Zhang L, Yang M H. Real-time compressive tracking [M]// Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012:864-877
- [3] Belezni C, Schreiber D, Rauter M. Pedestrian detection using GPU-accelerated multiple cue computation [C]// 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2011:58-65
- [4] Yan J, Lei Z, Yi D, et al. Multi-pedestrian detection in crowded scenes: A global view[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012: 3124-3129
- [5] 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
- [6] Li Y, Lu W, Wang S, et al. Local Haar-like features in edge maps for pedestrian detection[C]// 2011 4th International Congress on Image and Signal Processing (CISP). IEEE, 2011, 3: 1424-1427
- [7] Baltieri D, Vezzani R, Cucchiara R. People orientation recognition by mixtures of wrapped distributions on random trees[M]// Computer Vision-ECCV 2012. Springer Berlin Heidelberg, 2012:270-283
- [8] Ouyang W, Wang X. A discriminative deep model for pedestrian detection with occlusion handling[C]// 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012:3258-3265