

基于非局部稀疏特征的行人检测方法

彭怡书, 颜云辉, 赵久梁, 张 尧

(东北大学 机械工程与自动化学院, 辽宁 沈阳 110819)

摘 要: 利用周围邻域信息约束进行加权稀疏表示以达到行人检测的目的. 采用 Fisher 判别字典学习的方法, 得到一个能够更好地提取图像的具有更强辨别性稀疏特征的字典, 利用图像中周围信息约束, 求得该字典表示下的稀疏特征, 并根据对当前图像块的稀疏表示残差进行分类. INRIA 数据库的实验表明非局部稀疏特征具有明显的区分能力. 同时, 对行人目标进行邻域约束, 能够有效地表示出同目标区域的稀疏特征.

关 键 词: 行人检测; 非局部; 稀疏表示; 判别字典; 优化配矿

中图分类号: TP 391.41

文献标志码: A

文章编号: 1005-3026(2015)04-0465-04

Pedestrian Detection Based on Nonlocal Sparse Feature

PENG Yi-shu, YAN Yun-hui, ZHAO Jiu-liang, ZHANG Yao

(School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China.

Corresponding author: PENG Yi-shu, E-mail: yshu_peng@163.com)

Abstract: By using the constraints around the neighborhoods for weighted sparse representation, the pedestrian detection problem was solved. A dictionary with a strong extracting discriminate and sparse features power was obtained by using the Fisher discriminant dictionary learning method. With the constraint of the neighborhoods, the image patch was represented as a sparse feature via the dictionary. By computing the representation of the residuals and comparing the residuals with a threshold, the patch label was determined to finish the classification task. The experiments on INRIA person datasets showed that non-local sparse feature has an obvious power of discrimination. The constraint of the neighborhoods makes the sparse feature represented effectively.

Key words: pedestrian detection; nonlocal; sparse representation; discriminate dictionary; optimization ore matching

物体检测是计算机视觉领域最为重要的问题之一, 其中, 行人检测由于其在视频监控、智能交通等领域的重要性, 成为物体检测问题中与人们的生活密切相关的问题. 随着方向梯度直方图技术在行人检测中的应用及推广, 基于方向梯度直方图特征描述进行的行人检测技术^[1]得到了很多学者的关注.

针对行人检测问题, 常见的解决方法分为特征提取和分类器设计两部分. 特征提取主要是选择一个较好地能够描述对象特征的算子, 如形状、颜色、轮廓等基本特征及近年来非常受欢迎的方向梯度直方图特征, 在行人检测问题中发挥了重

要的作用^[2]. 当获取较好的特征描述算子后, 基于 SVM^[3], Adaboost^[4], Parts^[5] 等分类方法判断处理的对象是否属于目标类.

近年来, 稀疏表示理论在计算机视觉领域发挥了重要的作用, 如图像去噪^[6]、图像识别^[7]、图像超分辨^[8] 等问题. 特别是人脸识别问题中, 人脸图像样本经过稀疏表示后的系数被证明具有特征描述的功能, 并且在简单的分类器下, 可以很好地进行分类任务. 基于此, 本方法主要使用稀疏表示作为特征描述算子, 并考虑行人检测问题的特点, 引入邻域信息的约束, 得到非局部稀疏表示算子, 采用稀疏分类的策略, 在 INRIA 行人数据库

上进行实验.

1 非局部稀疏特征提取

稀疏表示是指信号可用一个过完备字典进行稀疏表示,即该系数具有少数的非零元素值,或大量的元素值较小.基于稀疏表示的分类模型假设共有 n 个类,每类共有 m 个样本组成了该类的训练样本集.对于未知类别的测试样本 \mathbf{y} ,其应当可以很好地用训练样本中同类别的样本进行线性表示.那么 \mathbf{y} 可以用训练样本集 \mathbf{Y} 进行稀疏表示:

$$\mathbf{y} = \mathbf{Y}\boldsymbol{\beta}, \text{ s. t. } \|\boldsymbol{\beta}\|_0 < \zeta. \quad (1)$$

式中: $\|\cdot\|_0$ 表示向量的零范数; ζ 是一个较小的正数.得到测试样本的稀疏系数后,通过计算表示的残差,即可确定测试样本的类别:

$$\text{label}(\mathbf{y}) = \arg \min_i \|\mathbf{y} - \mathbf{Y}\boldsymbol{\beta}_i\|_2^2. \quad (2)$$

式中, $\boldsymbol{\beta}_i$ 表示样本对应第 i 类的稀疏表示系数.该模型通过引入一个判别字典,表示出样本的稀疏系数,进而利用残差进行分类.该分类模型简单有效,在人脸识别等分类问题中取得了很好的效果.

在行人检测问题中,图片信息量较大,具有冗余信息,而方向梯度直方图特征是基于方向梯度进行特征提取,不能很好地描述图像中更为丰富的特征.本方法对图像中每个像素点进行稀疏特征的描述,为了更好地表示出对象的稀疏特征,本方法会采用 Fisher 判别字典学习的方法得到一个具有启发式的字典,该字典学习方法在文献[7]中有详细介绍,这里不再赘述.

此外,由于单独的行人在图像中都是一片联通区域,在进行稀疏特征描述时,引入非局部信息,对当前点周围的信息进行加权及稀疏特征的提取.如图1所示,像素点 a_i 的非局部信息为

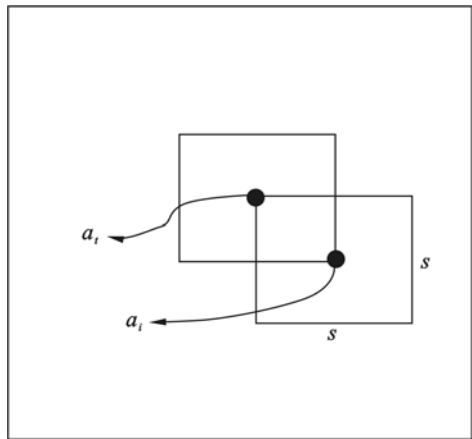


图1 像素点 a 的非局部信息示意图

Fig. 1 Explanatory views of nonlocal information of pixel a

$G(a_i)$,以 a_i 为中心的 $s \times s$ 的窗口的图像块为 $P(a_i)$,则有

$$G(a_i) = \sum_{t \neq i} \omega_t P(a_t) + \gamma P(a_i). \quad (3)$$

式中,

$$\sum_{t \neq i} \omega_t = 1 - \gamma, \omega_t^{-1} \propto \|a_t - a_i\|_2. \quad (4)$$

利用 Fisher 判别字典 \mathbf{D} 表示以像素点 a 为中心图像块的非局部信息,通过式(5)得到其非局部稀疏特征 $\hat{\mathbf{x}}$,

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_0, \text{ s. t. } G(a) = \mathbf{D}\mathbf{x}. \quad (5)$$

式(5)的求解方法很多,最常见方法有正交匹配追踪^[9]等,但求解效率较低.由于在进行非局部稀疏特征求解的时候,采用了有监督的 Fisher 判别字典^[10],已经具有潜在的稀疏表示能力.因此,本方法采用了迭代算法进行稀疏系数的求解,能够保证在很有限的迭代次数下得到一个满意的稀疏系数,表示当前像素的非局部稀疏特征.

2 分类器设计与行人检测

2.1 分类器选择

每个像素的非局部稀疏特征维数由学习获得的 Fisher 判别字典的尺寸决定,由于学习字典的时候启发式地考虑到类别信息,得到的特征可分为两部分,分别对应行人和非行人的编码系数.但本方法并不严格地依赖判别字典的学习效果,这是因为行人检测问题中感兴趣目标具有多种表现状态,不是简单的训练样本集可以拟合的.直接采用稀疏表示模型中的分类策略,分类的能力将会受到影响.为此,本方法首先计算图像块 $P(a_i)$ 的非局部信息 $G(a_i)$ 对应两类别的重构残差 $T(a_{ij})$,其中, $j=1,2$.

$$T(a_{ij}) = \|G(a_i) - \mathbf{D}\delta(j)\mathbf{x}\|_2^2. \quad (6)$$

式中, $\delta(j)$ 是示零函数,表示对应 \mathbf{D} 的第 j 部分元素为 0. 该图像块的类别通过式(7)判断:

$$\text{label}(P(a_i)) = \begin{cases} 1, T(a_{i1}) > \rho T(a_{i2}); \\ 2, T(a_{i2}) > \rho T(a_{i1}). \end{cases} \quad (7)$$

其中,阈值 ρ 用以筛选可以通过 Fisher 字典表示类别特征的图像块,如果不满足两个条件,该图像块的类别用 0 标记.这种松约束可以避免训练样本的过拟合问题,确定行人区域时,可以引入少量标记为 0 的区域即可.

2.2 行人区域检测过程

通过上述分析,本方法将待检测的图像进行

分块,并进行非局部信息的稀疏编码,进而判断图像块的类别,从而在图片中标注出行人区域,达到检测的目的. 首先,给定待检测图像 I ,将图像分块并计算块 a_i 的非局部信息 $G(a_i)$. 然后,利用 Fisher 字典 D 计算对应块的非局部信息 $G(a_i)$ 的稀疏特征 x_i . 最后,根据非局部稀疏特征确定该图像块是否属于行人区域.

此外,在对一个图像块进行稀疏表示之前,需要进行必要的预处理. 常见的图像预处理有去噪、边缘提取、增强等. 由于基于稀疏表示的框架往往是对图像的稀疏特征进行提取,而这种具有区分性的特征在浅层图像理解时表现出一定的梯度信息和边缘轮廓信息. 因此,本方法对图像块进行了类似操作,更侧重彰显图像块所具有的边缘能量. 这里,拉普拉斯算子被用作提取图像块的边缘信息.

3 实验与分析

本实验在 MatlabR2012b 环境下运行,使用 INRIA 行人数据集,该数据集主要标注了图片中大于 100 像素的直立行人. 由于在行人检测过程中,主要感兴趣的对分类有关键作用的是边缘信息,在进行训练前,对样本进行预处理,能够保留更为明显的特征. 图 2 是图像预处理后的结果. 由图 2 可知,图像的边缘信息被很好地表现处理,平

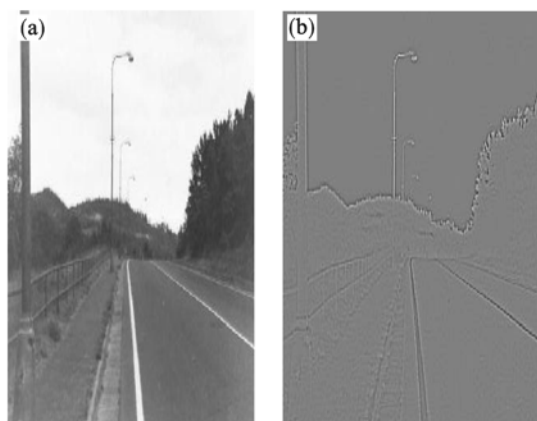


图 2 图像预处理结果
Fig. 2 Results of image preprocessing
(a)—原图; (b)—预处理图.

滑区域的信息被弱化. 这点并不会直接导致方法的性能退化,因为稀疏表示中的字典原子也往往是一些边缘的组合. 经过预处理,图像块的特征更加明显,更有利于稀疏特征的表达.

首先,采用 Fisher 字典学习的方法对训练样本进行训练,图 3 是训练样本集中的部分正负样本. 其中,上方是含有行人的正样本图片,下方是不包含行人的负样本,实验中,统一到相同的尺寸. 待检测图片经过预处理后,分块并提取块的非局部信息,实验中,分块的窗口大小并不固定,这里 $s = 13$. 非局部信息提取过程中 $\gamma = 0.68$,标记图像块类别的阈值 $\rho = 1.3$.



图 3 Fisher 判别字典学习使用的样本集
Fig. 3 Sample sets used in Fisher discriminate dictionary learning

稀疏特征作为冗余信息的压缩特征,常被使用在分类问题中,在基于 KSVD 字典学习的分类

问题中也发挥了重要的作用,此外,行人检测问题中的 SVM 分类器对较高维的特征具有很好的效

果. 为此, 本实验将与这两种方法进行对比. 本实验中, Fisher 字典的尺寸为 300, SVM 方法的检测结果为 88%, 基于 KSVD 字典学习方法进行检测的结果为 89%, 而本方法检测结果为 91%. 此外, 本方法与经典的 KSVD 字典学习方法均需要训练字典, 图 4 是每个字典中不同基元个数的检测结果. 其中, KSVD 字典学习方法是每类独立学习子字典, 然后将两个类别的子字典合并为一个大数据对样本进行稀疏表示. 从图中可以看出, 当基元个数为 300 时, 具有最好的检测结果.

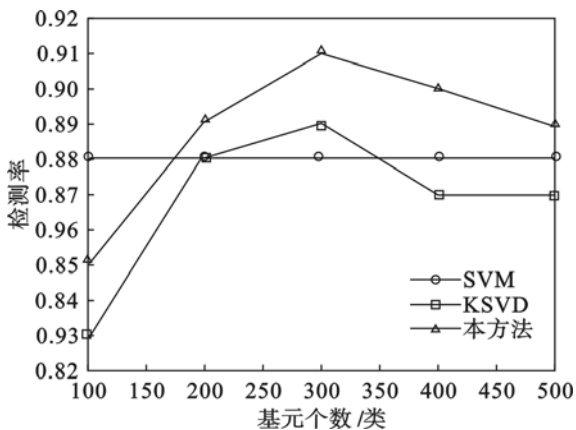


图 4 不同基元个数的检测方法实验结果对比

Fig. 4 Comparison of different methods with different basis numbers

4 结 论

1) 图像经过稀疏表示的特征具有明显的区分能力, 不仅表现在像素意义上的冗余信息压缩, 通过一定的预处理, 能够很好地表示不同类别之间的差异性, 应用于行人检测问题中.

2) 非局部稀疏表示的方法在提取稀疏特征的同时, 对行人目标进行邻域约束, 能够有效地表示出同目标区域的稀疏特征.

参考文献:

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, 2005: 886 – 893.
- [2] 苏志松, 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.
(Su Zhi-song, Li Shao-zi, Chen Shu-yuan, et al. A survey on pedestrian detection [J]. *Acta Electronica Sinica*, 2012, 40(4): 814 – 820.)
- [3] Oren M, Papageorgiou C, Sinha P. Pedestrian detection using wavelet templates [C]// IEEE Conference on Computer Vision and Pattern Recognition. Puerto Rico, 1997: 193 – 199.
- [4] Viola P, Jones M, Snow D. Detecting pedestrians using patterns of motion and appearance [C]// IEEE International Conference on Computer Vision. Nice, 2003: 734 – 741.
- [5] Xu J L, Vazquez D, Lopez A M, et al. Learning a part-based pedestrian detector in a virtual world [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(5): 2121 – 2131.
- [6] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries [J]. *IEEE Transactions on Image Processing*, 2006, 15(12): 3736 – 3745.
- [7] Wright J, Ganesh A, Yang A, et al. Robust face recognition via sparse representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210 – 227.
- [8] Yang J C, Wright J, Huang T S, et al. Image super-resolution via sparse representation [J]. *IEEE Transactions on Image Processing*, 2010, 19(11): 2861 – 2873.
- [9] Pati Y C, Rezaiifar R, Krishnaprasad P S. Orthogonal matching pursuit; recursive function approximation with applications to wavelet decomposition [C]// The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers. Pacific Grove, 1993: 40 – 43.
- [10] Yang M, Zhang L, Feng X C, et al. Fisher discrimination dictionary learning for sparse representation [C]// IEEE Conference on Computer Vision. Barcelona, 2011: 543 – 550.