

6-1-2020

Multi-pose Pedestrian Detection Based on Posterior Multiple Sparse Dictionaries

Lingkang Gu

College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China;

Mingzheng Zhou

College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China;

Wang Jun

College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China;

Xiu Yu

College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China;

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

Multi-pose Pedestrian Detection Based on Posterior Multiple Sparse Dictionaries

Abstract

Abstract: In order to detect pedestrians effectively, a *multi-pose pedestrian detection method based on posterior multiple sparse dictionaries* was proposed. Through pre-learning multiple different sparse dictionaries, and sparse coding the image, statistics for each dictionary corresponds to sparse coding histogram as the pedestrian image feature descriptor. The common information of multiple sparse dictionary features of all positive samples was obtained, and the feature of a single pedestrian sample was weighted, and the features of a posteriori multiple sparse dictionary could be obtained. Then pedestrians of different poses and views were divided into subclasses with clustering algorithm. A classifier was trained for each subclass. A multi-pose-view ensemble classifier was trained to combine the output values of different subclass classifiers with an equally weighted sum rule. Experimental results on different datasets suggest that the proposed posterior feature is more than the classical sparse dictionary and other typical features. Compared with the existing methods, by combining the posterior feature and the multi-pose-view ensemble classifier, the proposed method improves the detection accuracy effectively.

Keywords

pedestrian detection, sparse dictionaries, feature extraction, posterior multiple sparse, multi-pose

Recommended Citation

Gu Ling kang, Zhou Mingzheng, Wang Jun, Xiu Yu. Multi-pose Pedestrian Detection Based on Posterior Multiple Sparse Dictionaries[J]. *Journal of System Simulation*, 2017, 29(2): 326-331.

基于后验多重稀疏字典的多姿态行人检测

谷灵康, 周鸣争, 汪军, 修宇

(安徽工程大学计算机与信息学院, 安徽 芜湖 241000)

摘要: 为了有效检测行人, 提出一种基于后验多重稀疏字典的多姿态行人检测方法。根据多个不同稀疏度字典对图像稀疏编码, 构成分布直方图作为行人图像多重稀疏字典特征; 统计全部正样本特征的特性信息, 对单个行人样本特征加权, 获得表征行人的后验多重稀疏字典特征; 根据视角和行人姿态的不同利用聚类方法划分子类, 且训练等权重加和方式的多姿态分类器。在多个数据集上的测试表明, 所提方法的特征维数低, 描述能力优, 有效提高了行人检测精度。

关键词: 行人检测; 稀疏字典; 特征提取; 后验多重稀疏; 多姿态

中图分类号: TP391.4 文献标识码: A 文章编号: 1004-731X (2017) 02-0326-06

DOI: 10.16182/j.issn1004731x.joss.201702012

Multi-pose Pedestrian Detection Based on Posterior Multiple Sparse Dictionaries

Gu Ling kang, Zhou Mingzheng, Wang Jun, Xiu Yu

(College of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China)

Abstract: In order to detect pedestrians effectively, a multi-pose pedestrian detection method based on posterior multiple sparse dictionaries was proposed. Through pre-learning multiple different sparse dictionaries, and sparse coding the image, statistics for each dictionary corresponds to sparse coding histogram as the pedestrian image feature descriptor. The common information of multiple sparse dictionary features of all positive samples was obtained, and the feature of a single pedestrian sample was weighted, and the features of a posteriori multiple sparse dictionary could be obtained. Then pedestrians of different poses and views were divided into subclasses with clustering algorithm. A classifier was trained for each subclass. A multi-pose-view ensemble classifier was trained to combine the output values of different subclass classifiers with an equally weighted sum rule. Experimental results on different datasets suggest that the proposed posterior feature is more than the classical sparse dictionary and other typical features. Compared with the existing methods, by combining the posterior feature and the multi-pose-view ensemble classifier, the proposed method improves the detection accuracy effectively.

Keywords: pedestrian detection; sparse dictionaries; feature extraction; posterior multiple sparse; multi-pose

引言

行人检测是智能视频监控、机器视觉、车辆辅



收稿日期: 2016-05-01 修回日期: 2016-07-14;
基金项目: 安徽省自然科学基金(KJ2015A311); 安徽省自然科学研究(TSKJ2014B11)
作者简介: 谷灵康(1981-), 男, 安徽临泉, 硕士, 讲师, 研究方向为行人检测; 周鸣争(1958-), 男, 安徽枞阳, 学士, 教授, 研究方向为图像处理等。

助驾驶等应用的第一步, 但由于行人自身的特征兼具刚性和柔性, 且易受姿态、遮挡和视角等因素对外观的影响, 使得行人检测成为计算机视觉研究的难点与热点^[1]。行人检测吸引了大批研究者, 取得了一定的研究成果^[2-4]。早期往往选用单一特征描述行人, 如 HOG^[5], Shapelet^[6], EOH^[7]等特征描述行人的轮廓特征, LBP 特征^[8]描述行人的纹理特

征, Haar 特征^[9]描述行人的频域信息, CSS 特征^[10]则是局部颜色自相似特征。单一特征仅从单个角度描述行人, 描述能力有限, 导致检测效果不理想。为了提高检测精度, 最近几年很多研究者提出多特征融合的方法^[2,4], 如 Wojek 等利用 HOG+Haar+光流 HOG 特征进行行人检测^[11]; Wu 和 Nevatia 则结合 Edgelet、HOG 和协方差矩阵构成行人检测框架^[12]; 而 Schwartz 等利用共生矩阵的 LBP 特征结合 HOG 特征实现行人检测^[13]。这样固然能提高检测的准确性, 但是需要花费大量的时间, 且由于行人姿态及视角的多样性, 对于现实应用还不是很理想。针对这些问题, 本文根据人体特性提出一种有效的行人检测方法。

对于图像中的隐含特征或局部性质(如角点、边缘等), 利用稀疏方法^[14]可以学习得到, 该方法能够有效地表达图像自身的信息。Ren 等在 HOG 模型的基础上, 通过稀疏编码和插值计算, 提出了稀疏编码直方图^[15], 该方法提升了检测精度, 但由于特征维数较高导致计算量较大。

由于视角和行人姿态的多样性, 在检测过程中常导致类内外观有很大差异, 为此一些文献把行人分成多个子类。如文献[16]按距离和视角的前、后、左、右把行人分为十二类, 并利用训练好的级联分类器检测行人; 文献[17]则将行人样本划分为前、后、左、右及混合五个子类; 文献[18]按照行人不同姿态和视角, 将行人分成 9 个子类。上述文献都是根据研究者的主观看法进行行人子类划分的, 而对于行人子类的自动划分则研究的较少。对于子类自动划分方法比较经典的是基于 LLE 和 K-means 聚类提出的子类自动划分方法^[19], 但是这种方法计算量较大。

针对上述问题, 本文提出一种基于后验多重稀疏字典的多姿态行人检测方法。该方法先获取能够表征行人的后验多重稀疏字典特征; 再利用聚类方法, 自动按行人姿态和视角划分子类, 并构建分类器, 从而有效地提高了检测精度。

1 基于后验多重稀疏字典的特征描述

后验多重稀疏字典行人检测方法是建立在稀疏编码的基础上, 依此获得能够表征行人的后验多重稀疏字典特征。

1.1 稀疏编码

稀疏编码的目标就是找到一组只有很少的几个非零系数项的线性组合, 并用较少的几个基向量将信号表示出来, 故称之为稀疏。给定集合 $D = \{\phi_i, i=1, 2, \dots, k\}$, $D \in \mathbb{R}^{m \times k}$, D 是原子信号字典, $k > m$, ϕ_i 是基向量, 是字典的第 i 个原子, 则稀疏编码模型可以表示为:

$$Y = \sum_{i=1}^k \alpha_i \phi_i \quad (1)$$

式中: α_i 为 Y 在基向量 ϕ_i 上的系数。

对输入信号的学习是稀疏表示的第一步, 其中 $Y \in \mathbb{R}^{m \times n}$ 表示输入的训练信号, $D \in \mathbb{R}^{m \times k}$ 表示字典, $X \in \mathbb{R}^{k \times n}$ 表示稀疏表示系数矩阵, 则稀疏表示优化为:

$$\min_{D, X} \|Y - DX\|_F^2 \quad \forall i, \|x_i\|_0 \leq K \quad (2)$$

式中: x_i 是 X 的第 i 列向量, 为 ℓ^0 范式, K 是非零分量数目上限。经典的字典学习方法有很多, 常用的有基追踪、最优方向方法等, 本文采用源于奇异值分解、期望条件最大算法-KSVD 算法^[20], 该算法的优点在于可以降低计算量且可以避免陷入局部最小。

通过统计被激活单元构成分布直方图来形成图像特征, 其记录了被激活单元的分布特征, 在一定程度上能够描述图像特征。

1.2 多重稀疏字典直方图及后验特征描述

单个稀疏字典可能只能得到有限的被激活单元, 导致有用的特征信息的遗漏, 不能有效描述图像特征。因此考虑预先学习多个不同稀疏度字典, 组合多个不同稀疏度字典直方图, 以此形成多重稀疏字典直方图。

图 1 为图像的多重稀疏字典直方图, 统计

了多个不同稀疏字典的特征单元，直方图特征信息显示了图像在多个稀疏字典中的特征分布差异性。

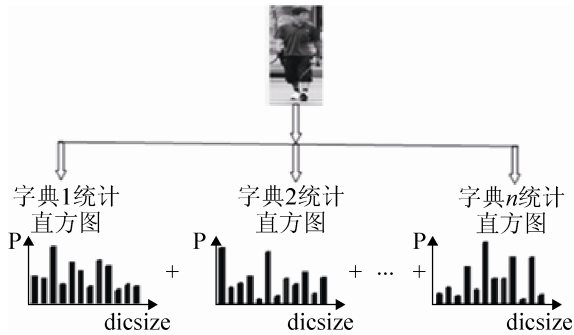


图1 多重稀疏字典直方图
Fig. 1 Multiple sparse dictionary

多重稀疏字典直方图提取特征描述如下：

$$m = (\text{His}D_1, \text{His}D_2 \cdots \text{His}D_c) \quad (3)$$

式中： $\text{His}D_i (i=1, 2, \dots, c, c$ 为稀疏字典个数)表示第*i*个稀疏字典直方图，为了保证其稀疏性，取 $1 < c \leq 10$ 。 $\text{His}D$ 表示如下：

$$\text{His}D = \sum_{i=1}^k \sum_{j=1}^n r(\phi_j) \quad (4)$$

式中： ϕ_j 表示字典的第*j* ($1 \leq j \leq n$, n 为输入信号的总数，即块个数)个输入信号的第*i*个基向量； k 为字典列向量个数，即基向量数目，而 $r(\phi_i)$ 为：

$$r(\phi_i) = \begin{cases} 0, & \text{if } \alpha_i = 0 \\ 1, & \text{otherwise} \end{cases} \quad (5)$$

式中： α_i 为基向量 ϕ_i 的系数。

后验多重稀疏字典特征提取方法如下：首先，计算每个正样本的多重稀疏字典直方图特征 $m_i (i=1, 2, \dots, c)$ ，并加和特征对应元素，得到训练正样本的特征和值，记为先验特征 M ：

$$M = \sum_{i=1}^c m_i \quad (6)$$

利用归一化方法对式(6)进行归一化处理，记为，其表达了行人样本的共性信息。

然后用特征对单个样本特征 m_i 做加权处理，得到单个样本的后验多重稀疏字典特征，其第*j*个元素值为：

$$\bar{m}_{i,j} = \bar{M}_j * m_{i,j} \quad (7)$$

2 多姿态-视角集成分类器

由于拍摄视角及行人姿态的多样性，想找到一种有效的方法能够识别各种姿态和视角的行人很难。如前文所述，一些文献按照一定的规律，为了降低行人姿态的类内差异，把行人划分为多个子类，取得了良好的效果。本文在文献[19]研究的基础上，利用K均值聚类方法对不同姿态和视角的行人做子类划分，并集成各子类分类器。

运用K均值聚类方法对多姿态多视角行人自动聚类，由于K均值聚类算法须事先指定类别个数K，而个数的多少直接影响集成分类器性能，根据下一节的实验分析，选取K值为5在自制训练数据集上进行聚类，如图2所示，5种不同形状的点代表了聚类结果的分布，图中也给出了各子类的代表样本，可以看出，算法大致将行人分为了正面视角(Front)、背面视角(Back)、左侧面视角(Left)、右侧面视角(Right)和腿大角度分开(Big)的5个行人子类。

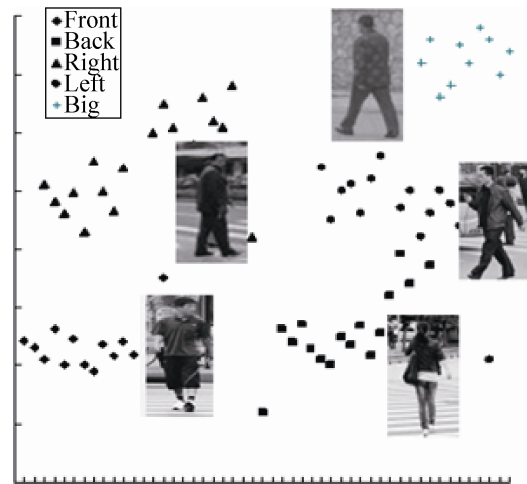


图2 行人聚类图
Fig. 2 Pedestrian clustering

根据聚类结果可看出，这种聚类考虑了视角以及行人步态的大小，这样就可以降低行人类内的差异性。

实验中对每一类样本分别提取后验特征后，训练K个交叉核SVM^[21]子类分类器，但由于视角和

姿态是连续变化的量, 聚类得到的有限姿态-视角划分是离散的, 因此, 利用式(8)将各子分类器集成, 构成统一的多姿态行人检测器。

式(8)中 b 为阈值、第 j 个子分类器输出值为 S_j 。集成后, 对于一个待识别行人样本, 可能介于多个有限的姿态视角之间, 导致同时被多个子分类器所识别, 被越多的子分类器识别的样本是行人的可能性越大。

$$O_T = \sum_{j=1}^k S_j + b \quad (8)$$

3 仿真实验及分析

本文算法程序在 Intel CPU 双核 2 GHz, 64 位 Windows 7 系统, 4G 内存的计算机上运行。实验样本之一来源于自拍的行人视频数据库, 该视频库包括多种场所(教学楼、停车场、道路等)、多种天气(晴天、雨天、雾天)和多种姿态及多视角等复杂情况下的行人视频。含有行人的图称为正样本, 共建正样本 4 600 个, 其中 4 000 个用来训练, 600 个用来验证。为了验证本文所述算法检测效果的通用性, 在目前常用的行人数据集^[22]中, 引入 2 个国际上常用的数据集: INRIA 行人数据集和 NICTA 行人数据集^[23]。后面的实验数据图是在 INRIA 行人数据集上获得的, 而自拍行人数据集和 NICTA 行人数据集是作为进一步验证用的。

稀疏度是影响字典直方图的重要参数, 不同稀疏度的字典对应的统计特征信息会有一些的差异。

多重稀疏字典直方图实验中选取不同的稀疏度字典, 稀疏度分别是: sparsity=1~3(sp3); sparsity=1~5(sp5); sparsity=1~7(sp7); sparsity=1~9(sp9), 实验结果如图 3 所示。

由图 3 可知, 在稀疏度较少(如 sp3)时, 其表现效果较差, 随着稀疏度的不断增加, 检测效果不断提升, 但是当稀疏度由 sp7 到 sp9 时, 实验效果已经非常接近。为了统计更加全面的特征分布信息而不断增加稀疏度, 虽然实验效果会不断提升, 但是随着选取的稀疏度不断增加, 效果将

趋于饱和。

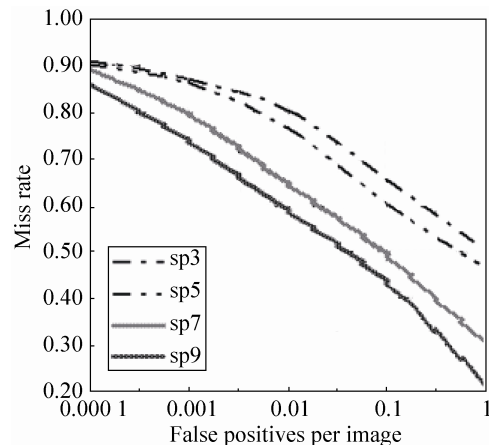


图 3 不同稀疏度字典的 DET 曲线
Fig. 3 DET curves of different sparsity dictionary

选取不同大小的块(patch), 在稀疏表示中对所学习到的字典特征, 也会产生影响。在实验过程中分别选取大小为 5×5、7×7 和 9×9 的块进行试验, 测试结果如图 4 所示。由图 4 可知, 在块大小为 7×7 时得到的检测效果最优; 而当用 5×5 大小的块进行实验时, 效果要低于其他大小的检测效果, 主要是因为较小的块在稀疏表示中学习到的字典特征不全面; 如果选取的块较大(如 9×9)也会降低检测效果。

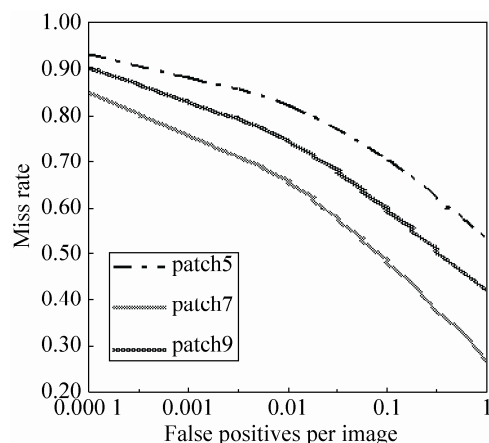


图 4 不同大小块的 DET 曲线
Fig. 4 DET curves of different size blocks

字典大小影响着特征维数和统计特征特性, 选取大小从 50~150 的字典进行实验, 结果如图 5 所示。由图 5 可知, 随着字典增大, 检测效果会有所提升, 但是当大小为 150 时, 其表现效果反而略低

于大小为 100 的字典。究其原因是因为较小的字典学习到的特征会非常有限,不能很好地表现图像特征;而为了更好地表现图像特征就会选择较大的字典,但是随着字典大小的增加,对应的特征分布越来越离散,这样又会导致效果有所下降,所以选取适当的字典大小是非常必要的。

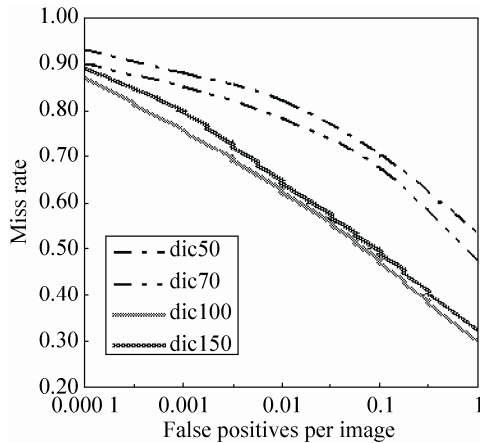


图5 不同大小字典的 DET 曲线

Fig. 5 DET curves of different size dictionaries

不同子类个数 K 对检测性能也会有影响,为了分析其影响,设定子类个数分别为: 3, 5, 7, 然后根据上一节介绍的方法聚类,训练集成分类器,图6给出了在测试集上的检测曲线。

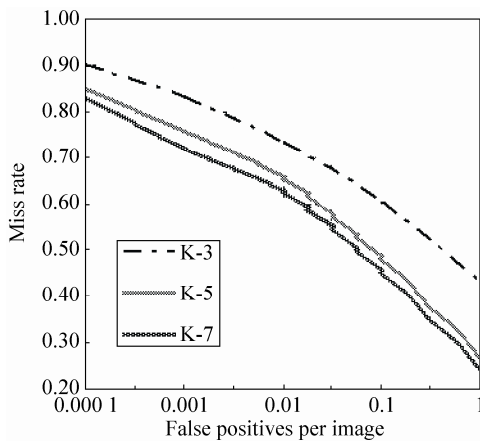


图6 不同子类个数的性能比较

Fig. 6 Performance comparison of different subclasses

从图6可以看出,子类个数为7的集成检测器性能比子类个数为3的性能有很大的提高,但是,与子类个数为5的性能相近。考虑到时间因素,本

文选定 K 值为 5。

将本文算法与现有经典算法文献[11-13, 19]的方法进行比较。文献[11]利用 HOG+Haar+光流 HOG 特征进行行人检测;而文献[12]是结合 Edgelet、HOG 和协方差矩阵。文献[13]结合共生矩阵的 LBP 特征和 HOG 特征实现行人检测。文献[19]能自动划分行人子类。实验结果如图7所示,本文与文献[19]所述算法在检测性能上比其他算法较优,且都是多姿态-视角子类自动划分,从而说明本文所述方法的有效性。

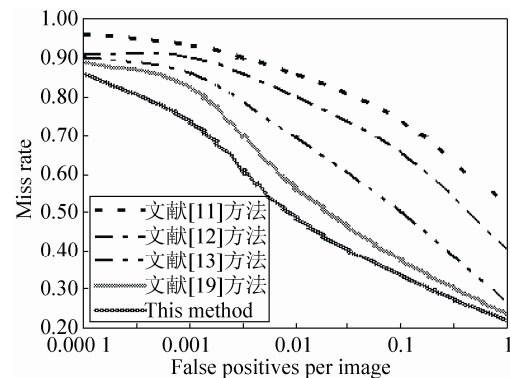


图7 和现有方法的比较

Fig. 7 Comparison with existing methods

4 结论

本文提出一种基于后验多重稀疏字典的多姿态行人检测方法,先统计全部正样本的多重稀疏字典特征共性信息,然后对单个行人样本特征进行加权,获得能够表征行人的后验多重稀疏字典特征;再利用聚类方法对不同姿态和视角的行人做子类划分,训练等权重加和方式的多姿态视角集成分类器。本文所述特征算法的描述能力优于经典稀疏字典和其它典型特征,且维度较低,通过将所提出的特征与多姿态视角集成分类器结合,与现有方法在测试集上进行实验,结果表明本文提出方法有效地提高了检测精度、运算较快能满足实时要求,但是前期训练所需时间较长,这也是后期需要改进的地方。另外文中还对影响实验检测效果的一些参数(如稀疏度、字典大小、块及子类个数)进行了分析研究,为后续的进一步深入研究打下基础。

参考文献:

- [1] Pedagadi S, Orwell J, Velastin S, et al. Local fisher discriminant analysis for pedestrian re-identification [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, USA. USA: IEEE, 2013: 3318-3325.
- [2] Hua Chunsheng, Makihara Yasushi, Yasushi Yagi, et al. Onboard monocular pedestrian detection by combining spatio-temporal hog with structure from motion algorithm [J]. *Machine Vision and Applications* (S0932-8092), 2015, 26(2): 161-183.
- [3] Junjie Yan, Bin Yang, Zhen Lei, et al. Adaptive structural model for video based pedestrian detection [J]. *Lecture Notes in Computer Science* (S0302-9743), 2014, 9003: 211-226.
- [4] Si Wu, Robert Laganière, Pierre Payeur. Improving pedestrian detection with selective gradient self-similarity feature [J]. *Pattern Recognition* (S0031-3203), 2015, 48(8): 2364-2376.
- [5] N Dalal, B Triggs. Histograms of oriented gradients for human detection [C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005: 886-893.
- [6] P Sabzmeydani, G Mori. Detecting pedestrians by learning shapelet features [C]// IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, MN, USA: IEEE, 2007: 1-8.
- [7] D Gerónimo, A López, D Ponsa, et al. Haar wavelets and edge orientation histograms for on-board pedestrian detection [C]// *Pattern Recognition and Image Analysis*. Germany: Springer Berlin Heidelberg, 2007: 418-425.
- [8] Y Mu, S Yan, Y Liu, et al. Discriminative local binary patterns for human detection in personal album [C]// IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK, USA: IEEE, 2008: 1-8.
- [9] C Papageorgiou, T Poggio. A trainable system for object detection [J]. *International Journal of Computer Vision* (S0920-5691), 2000, 38(1): 15-33.
- [10] S Walk, N Majer, K Schindler, et al. New features and insights for pedestrian detection [C]// IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010: 1030-1037.
- [11] Wojek C, Walk S, Schiele B. Multi-cue onboard pedestrian detection [C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2009: 794-801.
- [12] Wu B, Nevatia R. Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection [C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2008: 1-8.
- [13] Schwartz W R, Kembhavi A, Harwood D, et al. Human detection using partial least squares analysis [C]// 2009 IEEE 12th international conference on Computer vision. USA: IEEE, 2009: 24-31.
- [14] Rubinstein R, Bruckstein A M, Elad M. Dictionaries for sparse representation modeling [J]. *Proceedings of the IEEE* (S0018-9219), 2010, 98(6): 1045-1057.
- [15] Ren X, Ramaman D. Histograms of sparse codes for object detection [C]// 2013 IEEE Conference on Computer Vision and Pattern Recognition. USA: IEEE, 2013: 3246-3253.
- [16] K Guto, K Kidono, Y Kimura, et al. Pedestrian Detection and Direction Estimation by Cascade Detector with Multi-classifiers Utilizing Feature Interaction Descriptor [C]// IEEE Intelligent Vehicles Symposium. Germany: IEEE, 2011: 224-229.
- [17] A Prioletti, A Mogelmoose, P Grisleri, et al. Part-based pedestrian detection and feature-based tracking for driver assistance: real-time, robust algorithms, and evaluation [J]. *IEEE Transactions on Intelligent Transportation Systems* (S1524-9050), 2013, 14(3): 1346-1359.
- [18] A Shashua, Y Gdalyahu, G Hayun. Pedestrian detection for driving assistance systems: Single-frame classification and system level performance [C]// IEEE Intelligent Vehicles Symposium. Italy: IEEE, 2004: 1-6.
- [19] Q Ye, J Liang, J Jiao. Pedestrian detection in video images via error correcting output code classification of manifold subclasses [C]// IEEE International Conference on Intelligent Transportation Systems. Anchorage, AK, USA: IEEE, 2012: 193-202.
- [20] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Processing* (S1053-587X), 2006, 54(11): 4311-4322.
- [21] 孙锐, 侯能干, 陈军. 基于特征融合和交叉核 SVM 的快速行人检测方法 [J]. *光电工程*, 2014, 41(2): 53-62. (Sun Rui, Hou Nenggan, Chen Jun. Fast pedestrian detection method based on features fusion and intersection kernel SVM [J]. *Opto-Electronic Engineering*, 2014, 41(2): 53-62.)
- [22] Zhang Chunfeng, Song Jiatao, Wang Wanliang. Survey on pedestrian detection technology [J]. *Video Engineering* (S1002-8692), 2014, 38(3): 157-162.
- [23] Overett G, Petersson L, Brewer L, et al. A new pedestrian dataset for supervised learning [C]// Proc. 2008 IEEE Intelligent Vehicles Symposium. Eindhoven, Holland: IEEE Press, 2008: 373-378.