

基于 DCT 变换的多通道特征的行人检测

刘春阳 吴泽民 胡磊 刘熹

(解放军理工大学通信工程学院 南京 210007)

摘要 在行人检测中,针对目前多通道检测算法特征利用不充分的问题,提出一种基于 DCT 变换的多通道特征级联的行人检测算法。通过一种 2 层卷积网络模型将图像信息 DCT 变换后的数据进行整理,形成新的频域通道特征,该通道能描述行人的复杂纹理特征。结合梯度方向直方图特征、颜色空间特征和 DCT 频域特征,基于 Adaboost 算法训练了低开销的多通道特征行人检测器。在典型的公开行人库上的实验结果表明,该方法能提高检测的性能,在较低误检率时效果更加显著。

关键词 行人检测, DCT 变换, 多通道特征

中图法分类号 TP391 **文献标识码** A

Pedestrian Detection Based on DCT of Multi-channel Feature

LIU Chun-yang WU Ze-min HU Lei LIU Xi

(College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China)

Abstract In pedestrian detection, multi-channel feature detection has the defect of incomplete using features. In this paper, an algorithm of multi-channel feature detection based on discrete cosine transform (DCT) was proposed. We used a two-layer convolution network for arranging the image information after DCT to build a new channel in the frequency domain. This channel can describe complex textures about pedestrian. Combined with the features of the histogram of gradient and the color space as well as the DCT frequency domain, a low cost multi-channel pedestrian detector has been trained based on the Adaboost algorithm. Experiments on popular pedestrian databases show that the proposed method improves the accuracy of detection, and the effect is remarkable in low false positive per image.

Keywords Pedestrian detection, DCT, Multi-channel feature

1 引言

计算机视觉是一门研究图像和视频获取处理、分析和理解的科学,它的目标是使机器具有像人一样的视觉感知和理解能力^[1]。行人检测是计算机视觉领域中多项研究的前提,其具体定义为:判断输入图片(或视频帧)是否包含行人,如果有则给出位置信息^[2]。行人检测在行人跟踪、步态识别、行为分析和基于相机的辅助驾驶等^[3]领域应用广泛,具有非常重要的研究意义,是目前计算机视觉领域的研究热点和前沿方向。

目前基于图像的行人检测方法主要有两大类:第一类是基于卷积神经网络的深度学习模型,该模型利用多层传感器模拟人类的视觉处理机制,对大量的图像信息进行特征学习,其中最具代表性的模型有 AlexNet^[4], R-CNN^[5], YOLO^[6]等。该方法避免了特征的设计和标定,减少了手工设计和信息标注的时间,学习过程更加符合人类对于图像信息提取的步骤。但是该方法存在训练数据多,训练耗时长,硬件要求高等缺点。第二类是基于特征提取的方法,通过人工设计较好的样本特征(包括边缘、形状和变换等特征)来描述行人,其中最具代表性的特征有方向梯度直方图(Histogram of Oriented Gradient, HOG)^[7]特征、Haar^[8]小波特征、颜色的自相关特征

和形状轮廓模板特征等。该方法的检测性能在很大程度上依赖于提取的人工特征是否能够充分表征行人信息^[9]。本领域的技术权威 Dollar 教授等于 2009 年提出的积分通道特征(Integral Channel Features, ICF)^[10]对后续的行人检测研究工作影响巨大。在近些年的研究中,许多基于多通道特征的改进算法不断提高行人检测的性能,展现出多通道特征在行人检测领域的巨大潜力。一般来说,特征选取得越多,其涵盖的信息就越丰富,目标检测的识别效果就越好。因而,当下许多行人检测算法提出将不同的特征作为单独的通道,将多个不同的特征通道进行融合,综合地反映出行人的不同特征,以达到提高行人检测的准确度的目的^[11]。

本文对现有的多通道的行人检测算法进行了改进,采用了 HOG 特征、颜色特征和 DCT 频域特征,并将这 3 类特征作为不同类型的通道进行组合,取得了较好的检测效果。实验的仿真结果表明,相比于目前的多通道行人检测算法,联合了 DCT 频域特征的多通道联合算法的性能提高显著。

2 基于多通道的行人检测

一幅图像对应的通道可以看作是原始图像的一种映射,该通道既保持了原始图像的结构又反映了该图像的某些特

本文受国家自然科学基金(61501509)资助。

刘春阳(1993-),男,硕士,主要研究方向为计算机视觉、机器学习, E-mail: plaust_liu@163.com; 吴泽民(1973-),男,教授,硕士生导师,主要研究方向为数据融合、视觉信息处理; 胡磊(1985-),男,讲师,主要研究方向为压缩感知、视频信息处理; 刘熹(1972-),男,副教授,主要研究方向为数据链信息处理。

征^[12]。如式(1)所示,通道的输入一般为原始图像的像素值,通道的输出为对应通道映射的响应值。

$$C = \Omega(I) \quad (1)$$

其中, C 为通道, Ω 为该通道的某种映射关系, I 为原始图像。通道一般具有平移不变性,能够使用滑动窗口法进行快速检测。本文依据3种不同类型的映射变化,分别提取了颜色特征、梯度直方图特征和DCT频域特征,共形成了14个表征不同特征的通道。

2.1 方向梯度直方图特征通道

经典的方向梯度直方图算法的步骤如图1所示,步骤包括:颜色空间归一化、计算梯度信息、统计梯度方向、块内归一化梯度直方图等。

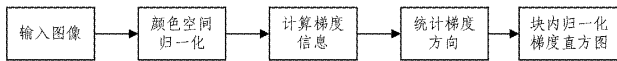


图1 方向梯度直方图通道提取流程

2.1.1 颜色空间归一化

为了减少光照因素的影响,首先对整个图像进行归一化。

Gamma压缩公式:

$$I(x, y) = I(x, y)^g \quad (2)$$

其中, g 代表Gamma压缩系数,一般取 $g=1/2$ 。

2.1.2 计算图像梯度信息

图像中像素点 (x, y) 的梯度如式(3)所示:

$$\begin{aligned} G_x(x, y) &= H(x+1, y) - H(x-1, y) \\ G_y(x, y) &= H(x, y+1) - H(x, y-1) \end{aligned} \quad (3)$$

其中, $G_x(x, y), G_y(x, y), H(x, y)$ 分别表示输入图像中的像素点 (x, y) 处的水平方向梯度、垂直方向梯度和像素值。像素点 (x, y) 处的梯度幅值和梯度方向分别如式(4)所示:

$$\begin{aligned} G(x, y) &= \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \\ \alpha(x, y) &= \tan^{-1} \left(\frac{G_y(x, y)}{G_x(x, y)} \right) \end{aligned} \quad (4)$$

2.1.3 级联不同方向的特征向量

将一幅图像分为许多不重叠的块,每个块为 8×8 像素。为提高计算速度,减少信息冗余,本文将每个块内的所有像素的梯度方向和幅值离散化到6个直方图区间,每个直方图区间表征该方向区域的梯度特征。将一幅图像中每个块的特征值进行级联,可组成图像在某个方向区域的梯度信息图,共提取6个不同方向梯度信息的特征图。

2.1.4 构建梯度特征通道

将6个不同方向的梯度信息作为映射通道,生成6个方向梯度通道,同时,将每个像素点对应的梯度强度作为一个单独的映射通道。方向梯度直方图特征通道共包括6个梯度方向通道和1个梯度强度通道,如图2所示。



图2 实例图像与基于方向梯度直方图的特征通道

2.2 颜色通道特征

颜色特征是计算机视觉领域中最常见的一种特征提取方法。在行人检测中,颜色特征不仅能够表征目标本身的大小、尺寸、运动方向等属性,而且计算方法简单,易于实现,具有良好的鲁棒性,能够很好地区分目标和背景。在行人检测领域中常用的颜色空间包括RGB,HSV,HIS,CIELUV等。本文选用CIELUV颜色空间作为颜色通道,首先将RGB颜色空间转换为CIELUV颜色空间,分别建立L,U,V颜色分量和通道输出的映射关系,形成3个不同的颜色通道,如图3所示。



图3 实例图像与基于颜色空间的特征通道

2.3 DCT频域特征通道

离散余弦变换经常在信号处理和图像处理中使用,一般用于对信号和图像进行有损数据压缩,具有能量集中的特性。图像信号的能量都集中在离散余弦变换后的低频部分^[13],边缘和纹理信息则出现在高频部分,通过相应的组合和数学计算能够突出高频部分的边缘信息同时降低背景的干扰,提高检测的性能。本文对图像进行DCT变换,提取频域中最能表现行人特征的低频分量,重新组合形成4个DCT频域特征通道。其具体步骤详见第3节。图4(a)为原始图像,图4(b)为4个DCT频域特征的通道。



图4 实例图像与基于DCT变换的通道

3 DCT频域特征的形成方法

3.1 DCT变换和DCT频率分量的选择

JPEG是图像有损压缩最流行的一种编码方式,通常将图像划分为大小相等的非重叠子块,然后对每个块进行DCT变换,根据所得的变换后的频域进行压缩、量化和编码,具有良好的去相关性^[14]。本文受JPEG编码的启发,以 4×4 像素为子块对图像进行DCT变换。经过DCT变换,其左上角为低频分量,右下角为高频分量,低频系数体现了图像中目标的轮廓和灰度分布特性,高频系数体现了目标形状的细节信息。

我们对DCT变换后的频率分量进行分析,选择能够更好地描述行人特征的频率分量。同时,只选择重要的频率分量可以有效降低整个系统的复杂度,并去除噪声等无关信息的影响,达到提高检测精度、降低计算复杂度和编码难度的目的^[15]。

根据理论分析及相关的实验对比,最终选择了图5中灰色子块所示的9个低频分量。通过这9个分量的变化和组合能够有效地描述行人的特征。

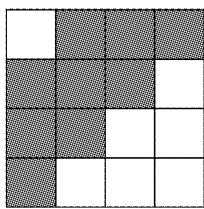


图 5 选择的 DCT 低频分量

3.2 频率分量的量化

研究表明,人类视觉系统中存在对部分频率更加敏感的现象^[16],因此 DCT 频率分量在被应用前需要做相应的量化处理。JPEG 标准量化表通过不同的感知因子来量化相应 DCT 频率分量的幅度,可以有效避免图像失真。考虑到人类视觉系统对频率分量的敏感程度不同(见图 6(a)),我们参考了 JPEG 压缩中采用的 8×8 的 DCT 标准量化表^[17],通过池化得到一种符合本文行人检测使用的 4×4 的 DCT 标准量化表(见图 6(b))。

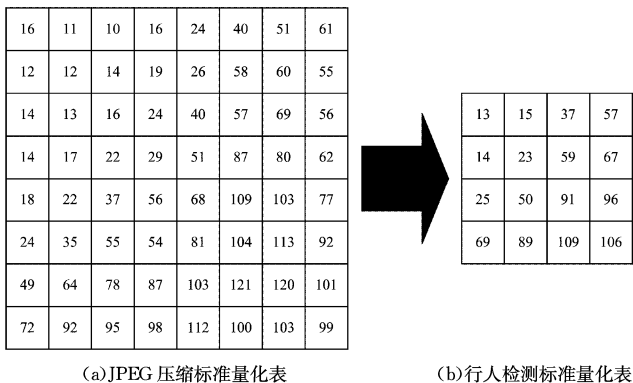


图 6

平均池化是指在池化过程中对池化域内所有值求和,并取其平均数作为子采样的值,本文采用平均池化的方法,池化域大小为 4×4。平均池化模型公式如式(5)所示:

$$P_{ij} = \frac{1}{2^2} \left(\sum_{2i-1}^{2i} \sum_{2j-1}^{2j} F \right) \quad (5)$$

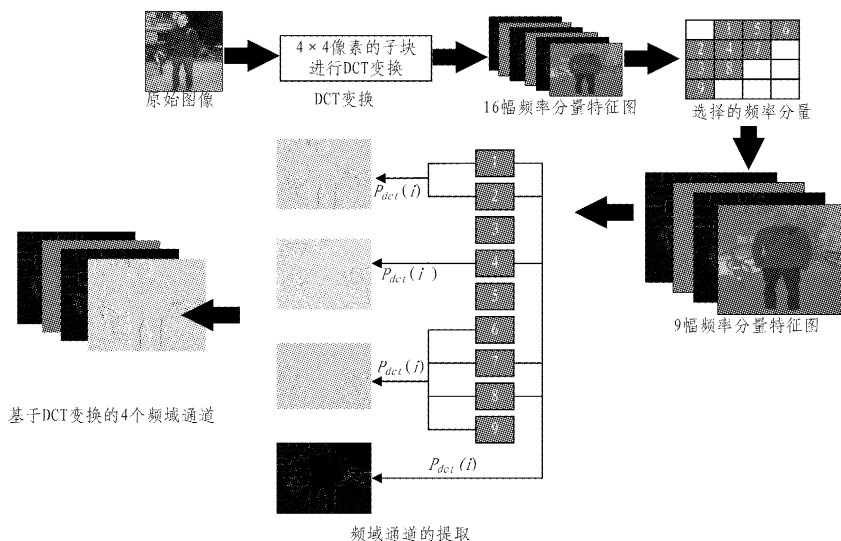


图 7 DCT 频域特征的提取方法

两层网络之间采用半连接的方式,具体如表 2 所列,其中√代表两层网络的链接点,将输入层中的 9 幅特征图与输出层的 4 个输出通道相连接。两层网络连接处的权

其中, F 为 8×8 的 JPEG 压缩标准量化矩阵中的分量, P 代表 DCT 行人检测标准量化表, P_{ij} 代表第 i 行中第 j 列的分量。如 P_{11} 代表第 1 行中第 1 列的分量。

根据式(6)将低频分量对应的量化值转换为通道计算中的量化权重,重要的频率分量(即量化表中的较小值)对应较大量化权重,不太重要的频率分量(即量化表中的较大值)对应较小的量化权重。每个量化权重的定义如下:

$$P_{dct}(i) = \begin{cases} \frac{1/q_i}{\sum_{j \in F} 1/q_j}, & i \in F \\ 0, & i \notin F \end{cases} \quad (6)$$

其中, q_i 和 q_j 为图 6(b)中行人检测标准量化表中的量化值, F 为图 5 中选取的 DCT 频率分量。 $P_{dct}(i)$ 代表通道中的低频分量的量化权重。所得到的量化权重能够在通道提取中更好地体现各低频分量对行人特征提取的不同贡献。所得的行人量化权重表如表 1 所列。

表 1 行人检测量化权重表

0	0.210	0.085	0.055
0.225	0.137	0.053	0
0.126	0.063	0	0
0.046	0	0	0

3.3 频域通道的提取

根据 3.1 节选取的 DCT 变换后的低频分量,提取每个子块中的不同低频分量,形成不同低频分量下包含不同边缘和纹理信息的特征图像,共得到了 9 幅包含不同低频信息的特征图像。考虑到不同特征图所表征的图像频域信息不同,参考卷积神经网络 LeNet-5^[18]中利用半连接的操作融合不同特征信息的图像,本文构建了一个简单的 2 层卷积网络模型,分为输入层和输出层。输出层与输入层以一种半连接的结构相连接,所产生的输出特征图能够更好地表征各频域融合后的图像特征。

如图 7 所示,输入层为 3.1 节中 9 个低频分量所构成的特征图,输出层形成 4 个单独的特征通道。

重为表 1 所列的行人检测量化权重 $P_{dct}(i)$ 。例如输出层中的第 1 个通道的特征值,由第 1 张和第 2 张特征图的特征值相连接。

表 2 输入通道与输出通道的半卷积连接表

	1	2	3	4	5	6	7	8	9
1	✓	✓							
2			✓	✓	✓				
3						✓	✓	✓	✓
4	✓	✓	✓		✓				✓

根据式(7)即可得到 4 个基于 DCT 变换的通道特征值。

$$C(j) = \sum_{i \in G_j} P_{det}(i) * f(i) \tag{7}$$

其中, G_j 代表输出层的第 j 通道中相连接的所有特征图, $P_{det}(i)$ 为连接处的权值, $f(i)$ 代表输入层中第 i 幅图像的特征值。

4 算法性能分析

4.1 数据库及实验环境

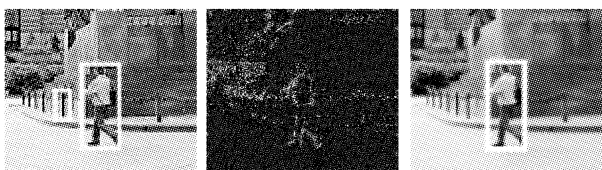
实验的数据库为 INRIA 人体检测数据库和 Caltech 人体检测数据库。INRIA 数据库包括数据集和测试集。其中训练集包含 614 张正样本图片和 1218 张负样本图片。训练集从含有行人的 614 张正样本图片中标注了 1237 个行人, 这些人体大多数是站立的, 但朝向各异、背景多变。Caltech 人体检测数据库是一个更大的行人检测数据库, 采用车载摄像头拍摄, 标注了约 250000 帧(约 137min), 350000 个矩形框, 2300 个行人, 包含测试集与训练集。这两个数据集都是现阶段较为通用的行人检测数据库, 能够较好地反映算法的性能。本文实验的硬件环境为一台 Inter Core i3, 2.40GHz 的双核处理器的电脑, 操作环境为 Windows7 操作系统下的 Matlab 2012R 平台。

4.2 实验步骤

基于多通道特征的行人检测算法通常分为训练和检测两个部分。训练流程包括载入训练集、提取多通道特征和训练分类器。分类器的训练通过 AdaBoost 实现, 共进行 3 轮训练。本文以 INRIA 图库训练为例, 在分类器训练的第一轮过程中, 首先将 614 张图片中标注的 1237 个正样本行人图片和从 512 个负样本图片中随机截取的共计 5000 张负样本图片投入到分类器中进行训练, 得到包含颜色、梯度信息和 DCT 频域 3 种不同类型共 14 个通道的初始检测器。然后, 用初始分类器在所有的负样本图片上进行行人检测, 检测出来的矩形区域自然都是分类错误的负样本, 我们称之为难例, 并将所有的难例作为负样本投入到分类器中, 进行第二轮训练。最后, 在第三轮使用新的分类器再次对所有负样本进行检测, 将生成的新的难例作为负样本投入分类器进行训练, 生成最终分类器。

4.3 实验结果及分析

在实验过程中, 本文通过添加 DCT 变换后的通道, 能够明显地提高检测的性能。对造成干扰的图像信息进行模糊处理, 以达到提高检测精度的目的。



(a)未添加 DCT 的检测 (b)基于 DCT 的通道 (c)添加 DCT 后的检测

图 8 测试图像对比示例

如图 8(a)所示, 误检窗口为左侧的一个类人柱状体。添加 DCT 通道后, 如图 8(b)所示, 能够弱化误检物体的边缘信息, 保持行人的特征, 达到抑制误检窗口的目的。如图 8(c)所示, 该误检窗口得到了抑制, 减少了误检个数, 提高了检测率。

我们采用漏检率-误检率(miss rate-false positive per image)曲线^[9]来评价本文算法的性能。该性能指标为目前行人检测领域的通用指标, 通过分类器漏检率和误检率的关系曲线, 可以清晰客观地反映行人检测过程中的综合性能。如图 9 所示在 INRIA 图库中, 本文提出的基于 DCT 变换的多通道算法在检测精度上优于传统的通道检测算法, 平均漏检率为 13.29%, 达到了目前传统算法中最低的平均漏检率。因此, 所提方法的性能优势明显, 特别是在较低误检率的情况下效果显著。

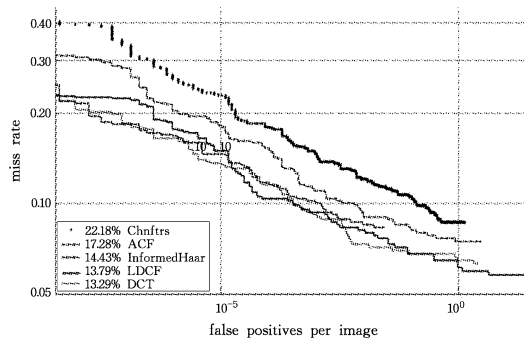


图 9 INRIA 图库的性能曲线

为了更全面地反映检测的性能, 我们在 Caltech 图库中进行了同样的实验, 实验结果如图 10 所示。可以看出, 在 Caltech 图库中, 其检测性能也大大优于其他几种算法, 与传统算法中最好的 LDCF 算法相比, 也只有约 1% 的性能差距。

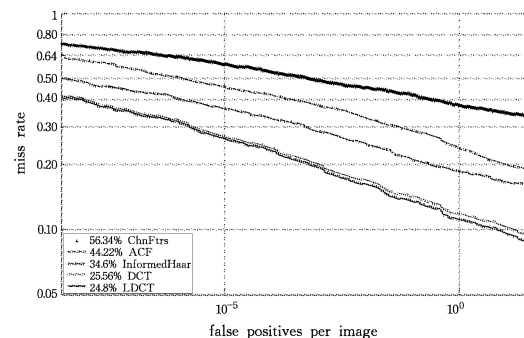


图 10 Caltech 图库的性能曲线

表 3 列出了添加 DCT 变换后的通道的检测性能与未添加 DCT 检测通道的性能在 INRIA 图库中不同的 FPPI 情况下具体行人检测的漏检率。可以发现, 在较低的 FPPI 的情况下, 其漏检率均为最低, 且下降明显, 说明在行人检测精度要求较高时本算法的性能提升明显, 整体的检测性能与原始方法相比提高明显。

表 3 不同误检率下的漏检率/%

漏检率	FPPI=			
	平均	10^{-2}	10^{-1}	10^0
ChnFtrs	22.18	40.19	23.37	9.18
ACF	17.28	31.09	18.17	8.32
InformeHaar	14.43	24.69	16.08	\
LDCF	13.79	22.87	15.56	6.83
DCT	13.29	22.57	13.32	7.25

- [6] 冯立男, 戚正伟. 嵌入式三维场景漫游系统的设计与实现[J]. 计算机工程, 2009, 35(18): 257-259.
- [7] 官酷杰. 基于 OpenGL ES 的移动平台图形渲染引擎研究与实现[D]. 北京: 北京交通大学, 2010.
- [8] 徐彩虹, 刘志, 潘翔, 等. 一种基于实例学习的三维模型检索匹配方法[J]. 浙江工业大学学报, 2012, 40(3): 326-330.
- [9] 吕书明, 张明磊, 孙树立. 基于简化和细分技术的三角形网格拓扑优化方法[J]. 计算机辅助设计与图形学学报, 2014, 26(8): 1225-1231.
- [10] HOPPE H. Progressive meshes[C]//Conference on Computer Graphics and Interactive Techniques. New York, USA: ACM, 1996: 99-108.
- [11] GARLAND M. Surface simplification using quadric error metrics[C]//Conference on Computer Graphics & Interactive Techniques. New York, USA: ACM, 1997: 209-216.
- [12] 于海军, 马纯永, 张涛, 等. 基于图像空间的快速碰撞检测算法[J]. 计算机应用, 2013, 33(2): 530-533.
- [13] 何伟, 李勇, 苏虎. 碰撞检测中的包围盒方法[J]. 重庆工学院学报(自然科学版), 2007, 21(23): 148-152.
- [14] 徐峰, 陈敏智. 虚拟校园三维仿真系统的设计及实现[J]. 浙江工业大学学报, 2007, 35(2): 155-158.
- [15] 卢威, 曾定浩, 潘金贵. 支持外观属性保持的三维网格模型简化[J]. 软件学报, 2009, 20(3): 713-723.
- [16] LLER T, TRUMBORE B. Fast, minimum storage ray/triangle intersection[C]//ACM SIGGRAPH. New York, USA: ACM, 2005: 21-28.
- [17] 夏怒. 三维场景中的景物拾取技术[J]. 电脑知识与技术: 学术交流, 2008, 4(1): 196, 199.

(上接第 224 页)

结束语 本文着重研究了基于多特征通道的行人检测算法, 深入分析了算法原理, 并且加入了基于 DCT 变换后的特征, 通过级联方向梯度直方图特征、颜色空间特征和基于 DCT 的特征通道对数据集进行训练和检测, 行人检测的精度明显提高。实验结果表明: 加入了基于 DCT 变换后的特征通道, 在误检率为 10^{-1} 时, 漏检率下降了 2% 以上, 且平均漏检率也优于已知的传统算法。目前所提算法还有进一步优化的空间, 可以通过减轻运算量降低复杂度, 缩短检测时间。

参 考 文 献

- [1] IKEUCHI K. Computer Vision: A Reference Guide[M]. Springer Publishing Company, Incorporated, 2014.
- [2] 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.
- [3] CAO J, PANG Y, LI X. Pedestrian Detection Inspired by Appearance Constancy and Shape Symmetry[J]. IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society, 2016, 25(12): 5538.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2): 2012.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]//CVPR'14. 2014: 580-587.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//CVPR'16. 2016: 779-788.
- [7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE Conference on Computer Vision & Pattern Recognition. 2005: 886-893.
- [8] ZHANG S, BAUCKHAGE C, CREMERS A B. Informed Haar-Like Features Improve Pedestrian Detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2014: 947-954.
- [9] WANG X G. Deep learning in image recognition[J]. Communications of the CCF, 2015, 11(8): 15-23.
- [10] DOLLÁR P, TU Z, PERONA P, et al. Integral Channel Features[C]//British Machine Vision Conference, BMVC 2009. London, UK, 2009.
- [11] ZHANG S, BENESON R, OMRAN M, et al. How Far are We from Solving Pedestrian Detection? [C]//CVPR'16. 2016: 1259-1267.
- [12] DOLLAR P, APPEL R, BELONGIE S, et al. Fast Feature Pyramids for Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(8): 1532-1545.
- [13] FARINELLA G M, RAV D, et al. Representing scenes for real-time context classification on mobile devices[J]. Pattern Recognition, 2015, 48(4): 1086-1100.
- [14] WALLACE G K. The JPEG still picture compression standard[J]. Communications of the ACM, 1991, 34(4): 30-44.
- [15] STURGES P, ALAHARI K, LADICKY L, et al. Combining Appearance and Structure from Motion Features for Road Scene Understanding[C]//British Machine Vision Conference, BMVC 2009. London, UK, 2009.
- [16] BATTIATO S, MANCUSO M, BOSCO A, et al. Psychovisual and Statistical Optimization of Quantization Tables for DCT Compression Engines[C]//International Conference on Image Analysis and Processing. IEEE Computer Society, 2001: 602-606.
- [17] PETERSON H A, PENG H, MORGAN J H, et al. Quantization of color image components in the DCT domain[C]//Proceedings of SPIE-The International Society for Optical Engineering. 1991: 210-222.
- [18] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2232.