

文章编号 1004-924X(2018)07-1802-11

## 采用人体树图与混合粒子群聚类的行人检测

孟晓燕\*, 段建民, 刘 丹

(北京工业大学 信息学部, 北京 100124)

**摘要:**为提升辅助驾驶系统的可靠性及安全系数,实现更高精度的行人检测,基于人体树图模型提出了一种改进的离线训练、在线检测的行人检测方法。首先,定义人体部件间的共生关系,得到对应父子部件对,结合 K-means 算法对其位置关系进行聚类获得部件类型。为兼顾类内紧密性与类间分离性,采用 MSE 和 DBI 构建具有两阶段适应度函数的混合粒子群聚类算法,在有效估计各部件最优聚类中心数量的同时,消除随机初始化对聚类准确率造成的影响。其次,将优化聚类得到的部件类型作为隐藏变量,通过求解隐结构 SVM 获取改进后的人体检测模型。最后,通过动态规划算法求解状态转移方程,在多个尺度上有效估计人体部件位置及检测包围盒,并结合非极大值抑制思想得到最终的行人检测结果。实验结果表明,所提方法在检测性能上明显优于 5 种行人检测方法,并且相比于原始 Pose-original 方法,在 INRIA 和 ETH 数据集上的丢失率分别下降了 8.14% 和 5.05%。实验证明该方法检测性能良好且具有较高的准确性和鲁棒性。

**关键词:**行人检测;人体树图模型;K-means 聚类;粒子群算法

**中图分类号:**TP391 **文献标识码:**A **doi:**10.3788/OPE.20182607.1802

## Pedestrian detection based on tree-structured graphical model of the human body and hybrid particle swarm clustering

MENG Xiao-yan\*, DUAN Jian-min, LIU Dan

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

\* Corresponding author, E-mail: 529108419@qq.com

**Abstract:** In order to improve the reliability and safety factor of driver assistance systems, and achieve pedestrian detection with a higher precision, an improved pedestrian detection method based on a tree-structured graphical model of the human body is proposed, and it consists of an offline training part and an online detection part. First, the corresponding parent-child parts are obtained by defining the symbiotic relationship between human parts, and then the K-means algorithm is applied to the location relationship between part pairs to acquire part types via clustering. For the purpose of taking both intra-class tightness and inter-class differences into account, a hybrid particle swarm optimization algorithm is built with a two-phase fitness function via introducing MSE and DBI. It is not only effective in estimating the number of optimal cluster centers, but also in eliminating the effect of random initialization on the clustering accuracy. Then, the part type obtained using the optimized

收稿日期:2017-11-20;修订日期:2018-01-23.

基金项目:北京市属高等学校人才强教计划资助项目(No. 038000543117004);北京市自然科学基金资助项目(No. JJ002790200802)

clustering method is considered as the latent variable. The pedestrian detection model is obtained through solving the latent structural SVM problem. Finally, we estimate the position of human parts and the detection bounding box on multiple scales based on solving the state equation via a dynamic programming algorithm, and obtain the final pedestrian detection result through incorporating the idea of non-maximum suppression. Experimental results indicate that the performance of the proposed algorithm is superior to those of five other pedestrian detection algorithms. In particular, on the INRIA and ETH databases, the loss rate of the proposed algorithm decreased by 8.14% and 5.05%, respectively, compared with that of the pose-original method. Experimental results show that the proposed algorithm has good performance and high accuracy and robustness.

**Key words:** pedestrian detection; tree-structured graphical model of human body; K-means cluster; particle swarm optimization

## 1 引言

近年来,随着汽车使用率的逐年递增,城市道路交通事故的发生频率也不断升高。行人作为城市交通的主要参与者和相对弱势群体,保证其生命财产安全是智能车辆研究领域的重要课题。尽管先进驾驶辅助系统(Advanced Driver Assistance System, ADAS)已经得到了有效应用,但是如何克服人体姿态、光照、遮挡等因素的干扰,实现更高精度的行人检测与定位,进而提升辅助驾驶系统的可靠性及安全系数,为车辆紧急避险及路径规划提供有效的数据输入,是智能车辆自动驾驶亟需解决的问题。

基于运动检测和特征学习的策略是目前解决行人检测问题的主要途径,其中基于特征学习的方法在近年来取得了较大进展。Viola等<sup>[1]</sup>对图像提取 Haar-like 特征,取得了良好的测试效果。Dalal等<sup>[2]</sup>提出的 HOG 特征极大的促进了行人检测算法的发展。针对 HOG 特征所存在的局限性,研究者提出了一些具有代表性的改进策略。Wang等<sup>[3]</sup>将 LBP 特征与 HOG 特征进行融合,弥补了 HOG 纹理信息缺失的不足。Felzenszwalb等<sup>[4]</sup>提出了一种可变形部件模型(Deformable Parts Model, DPM),通过设置隐藏变量,采用隐结构 SVM(Support Vector Machine, SVM)进行训练,一定程度上克服了目标遮挡及形变所造成的干扰。之后的改进算法 LatSVM-V2<sup>[5]</sup>在其基础上融入了多模型和上下文信息,检测的鲁棒性与准确率得到了进一步提升。Pedersoli等<sup>[6]</sup>基于多分辨率分层可变形部

件模型和由粗到精的推理策略,在搜索空间中消除相对次要区域,大幅提高了检测速度,并获得了较高的准确率。此外,基于深度学习模型的检测方法在公开数据集上取得了优良的测试结果,但是鉴于 DPM 模型的经典思路及完整的数学基础,其性能依然存在较大的提升空间。此外,将两种模型进行融合的方法也获得了突破性进展。Ouyang等<sup>[7]</sup>提出一种 UDN 算法,利用 CNN 和 DPM 思想构筑神经网络,有效降低了检测的误检率和漏检率。文献[8]结合卷积神经网络提出一种形变约束池化方法,通过几何约束和惩罚参数实现对目标形变的建模。

由于传统基于人体部件的行人检测方法,通常仅利用子模型与主模型的空间先验知识,以及若干人体部件对星型模型<sup>[5]</sup>进行编码,对于各部件之间的关联属性缺乏必要的描述,进而限制了检测性能的进一步提高。为了克服上述问题,本文基于一种灵活的树图模型<sup>[9]</sup>,对人体各部件之间的空间关系和局部刚度进行描述,提出了一种离线训练、在线检测的交通环境行人检测方法。首先定义人体各部件之间的父子关系,并结合 K-means 算法<sup>[10]</sup>对父子部件所属类型进行聚类,获取训练样本的隐藏变量。之后,为了兼顾类内紧密性和类间分离性,基于均方误差(Mean Square Error, MSE)和戴维森堡丁指数(Davies Bouldin Index, DBI)构建包含两阶段适应度函数的混合粒子群聚类(Hybrid Particle Swarm Clustering, HPSC)算法,解决了 K-means 算法对初始聚类中心敏感,且聚类中心数量凭经验而定的缺陷<sup>[11-12]</sup>,进一步提高检测模型的判别性能。在检测阶段,利用离线训练所得模型,结合动态规划算

法<sup>[13-14]</sup>求解状态转移方程,进而获取测试样本的行人检测结果。最后,在 INRIA<sup>[2]</sup>和 ETH<sup>[15]</sup>数据集上对所提行人检测方法的性能进行了测试。

## 2 基于人体树图模型的行人检测

利用人体树图模型对各个部件之间的空间关系及局部刚度进行描述,并将部件类型定义为隐藏变量,通过坐标下降法<sup>[9]</sup>求解隐结构 SVM<sup>[16]</sup>,获取行人检测模型。之后,基于动态规划思想和非极大值抑制(Non-Maximum Suppression, NMS)算法完成行人检测任务。

### 2.1 人体树图模型

传统的 DPM 采用多模型的策略解决了视角变化对于人体检测的影响,并以不同视角下的人体部件共现约束关系为先验知识,使得行人检测结果更加合理。但是在模型训练过程中,包含丰富人体姿态信息的各部件之间空间关系及局部刚度通常没有得到有效利用,而这些先验知识在实验过程中对于提高检测模型的鲁棒性及准确性具有重要意义。

基于上述分析,本文采用一种树图结构对人体进行建模<sup>[9]</sup>。区别于多模型中,不同视角对应不同部件类型的策略,树图模型首先对各部件之间的父子关系进行定义,并对父子部件的相对位置执行 K-means 聚类,进而获取全部训练样本的部件类型信息。

将某幅图像表示为  $I$ , 设其中的待检测人体包含  $L$  个部件,第  $i$  个部件的像素坐标可描述为  $p_i = (x, y), i \in \{1, \dots, L\}$ 。用  $t_i$  表示部件  $i$  的所属类型,由当前部件与其父部件的相对位置决定,这种位置关系的种类取决于 K-means 算法的聚类中心数量。即设任意父子部件存在  $K$  种位置关系,则  $t_i \in \{1, \dots, K\}$ ,且所执行 K-means 算法的聚类中心数量也为  $K$ 。构建人体树图模型为  $G = (V, E)$ ,其中  $V$  表示人体部件集合, $E$  表示父子部件连接而成的线段集合。令  $t = \{t_1, \dots, t_L\}$ ,定义一个兼顾部件类型得分及父子部件组合得分的求和函数,用于评价各部件类型得分:

$$S(t) = \sum_{i \in V} b_i^{t_i} + \sum_{ij \in E} b_{ij}^{t_i t_j}, \quad (1)$$

其中: $i$ 与 $j$ 互为父子部件, $b_i^{t_i}$ 表示部件 $i$ 属于类型 $t_i$ 的可能性; $b_{ij}^{t_i t_j}$ 表示部件 $i$ 属于类型 $t_i$ ,部件

$j$ 属于类型 $t_j$ ,且两者同时存在的可能性。

依据公式(1)定义图像  $I$  中不同部件类型及像素位置的总体评分函数,如式(2)所示:

$$S(I, p, t) = S(t) + \sum_{i \in V} \omega_i^{t_i} \cdot \varphi(I, p_i) + \sum_{ij \in E} \omega_{ij}^{t_i t_j} \cdot \psi(p_i - p_j), \quad (2)$$

其中: $\omega_i^{t_i}$ 表示部件 $i$ 类型为 $t_i$ 时的滤波器模板, $\varphi(I, p_i)$ 是在像素坐标 $p_i$ 处的 HOG 特征向量。 $\omega_{ij}^{t_i t_j}$ 表示父子部件 $i$ 和 $j$ 分别在类型 $t_i$ 和 $t_j$ 下的滤波器模板,包含了相应的静态位置与共生信息。另外, $\psi(p_i - p_j) = [dx \ dx^2 \ dy \ dy^2]$ 代表父子部件 $i$ 与 $j$ 的位置关系,且 $dx = x_i - x_j, dy = y_i - y_j$ 。

图 1 为人体树图示意图<sup>[9]</sup>(彩图见期刊电子版),其中人体部件数量  $L$  为 26,部件类型数量  $K$  为 4。左图即为人体各部件之间的父子关系示意,各部分由肢体末端向头部依次互为父子关系,头部件在训练时被设置为根节点。由于训练所采用的 Image Parse 数据集<sup>[17]</sup>仅提供了 14 个人体标注点,为了进一步丰富姿态信息,对人体部件进行了扩充,如左图中的红色星型标注点所示,其具体原理参见文献[9]。

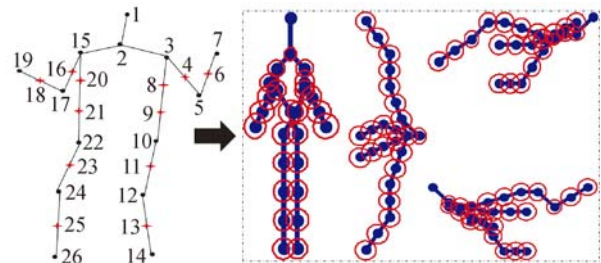


图 1 人体树图模型示意图

Fig. 1 Schematic diagram of tree-structured graphical model of human body

### 2.2 模型训练与行人检测

为了训练行人检测模型,采用坐标下降法<sup>[9]</sup>解决隐结构 SVM<sup>[16]</sup>问题。设训练正负样本图像均为  $I_k$ ,负样本中并未包含任何有效人体。 $p_k$ 代表第  $k$  个正样本中人体部件的像素坐标集合, $t_k$ 代表相应的部件类型集合,则初始优化问题如公式(3)所示:

$$\begin{aligned} & \arg \min_{\alpha, \xi \geq 0} \frac{1}{2} \beta \cdot \beta + C \sum_k \xi_k \\ & \text{s. t. } \forall k \in \text{pos } \beta \cdot f(I_k, z_k) \geq 1 - \xi_k \end{aligned}$$

$$\forall k \in \text{neg}, \forall z\beta \cdot f(I_k, z_k) \leq -1 + \xi_k, \quad (3)$$

其中:  $\beta = (\omega, b)$  为待训练的模型参数集合,  $\omega$  及  $b$  代表滤波器模板集合与类型支持率集合。将公式(2)表述为线性形式  $\beta \cdot f(I, z) = S(I, z)$ , 且  $z_k = (p_k, t_k)$ 。此外,  $\xi_k$  和  $C$  分别表示松弛变量和惩罚因子。

在检测阶段, 结合动态规划思想<sup>[13]</sup>, 利用检测模型对图像进行多尺度下的行人检测。首先基于公式(1)和(2)定义动态规划过程中的状态转移方程, 如公式(4)所示:

$$m_i(t_j, p_j) = \max_{t_i} b_{i_j}^{t_i} + \max_{p_i} \text{score}(t_i, p_i) + \omega_{i_j}^{t_i} \cdot \phi(p_i - p_j). \quad (4)$$

相应的指标函数, 如公式(5)所示:

$$\text{score}_i(t_i, p_i) = b_{i_i}^{t_i} + \omega_{i_i}^{t_i} \cdot \varphi(I, p_i) + \sum_{h \in \text{kids}(i)} m_h(t_i, p_i), \quad (5)$$

其中  $\text{kids}(i)$  代表部件  $i$  的子部件集合。

其次, 对单张图像的多个检测结果进行阈值化处理, 并采用 NMS 算法, 得到全部的最优人体部件位置, 完成行人检测任务。

### 3 基于 HPSC 算法的部件类型聚类

在训练过程中, 部件的类型将被视为隐藏变量, 因此, 各部件类型的数量和父子部件相对位置的聚类结果都将直接影响到行人检测模型的性能。然而, 传统 K-means 算法<sup>[10]</sup>对初始聚类中心存在较强的依赖性, 并且聚类中心的个数及初始化通常具有较强的随机性和盲目性, 导致聚类结果稳定性差且准确率较低。为了解决上述问题, 采用一种包含两阶段适应度函数的混合粒子群聚类算法, 在有效估计各部件最优聚类中心个数的同时, 消除随机初始化对聚类准确率造成的影响。

#### 3.1 部件类型数量优化

K-means 算法作为一种无监督学习方法, 其核心思想是最小化所有样本到所属聚类中心的欧氏距离和, 并采用迭代的方式实现收敛。对于  $N$  维向量  $x$  和  $y$ , 其欧氏距离计算如下:

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2}. \quad (6)$$

根据人体实际构造特点, 各父子部件之间的活动角度存在较大差异, 例如头部和颈部互为父子部件, 其像素坐标连线相对于水平方向的角度

变化范围要明显小于肩部和肘部, 因此, 单纯对各部件对采用统一的部件类型数量<sup>[9]</sup>会导致模型的泛化能力减弱, 影响检测性能。为了更加合理地确定各部件类型的数量, 本文基于 DBI<sup>[18]</sup>对 K-means 算法的聚类中心数量在  $P_{\min} \sim P_{\max}$  范围内进行评估。该指标能够同时兼顾类内紧密性和类间分离性两个方面, 并使数据点得到更加准确地划分。

首先, 引入 MSE 对类内紧密性进行评估, 其计算公式如下:

$$\text{MSE}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} d(x_j, C_i)^2, \quad (7)$$

其中:  $N_i$  表示第  $i$  类的数据点个数,  $x_j$  表示属于聚类中心  $C_i$  的第  $j$  个数据点;

其次, 设任意两类的聚类中心分别为  $C_i$  和  $C_j$ , 则两个类别之间的亲密度  $Q_{i,j}$  可以定义为:

$$Q_{i,j} = \frac{\text{MSE}_i + \text{MSE}_j}{d(C_i, C_j)}, \quad (8)$$

$Q_{i,j}$  的数值越大表示两个类别之间的分离性越差。取第  $i$  类与其余类别的亲密度最大值, 记为  $Q_i = \max(Q_{i,j})$ , 且  $i \neq j$ ;

最后, 通过计算每个类别的最大亲密度均值得到当前的 DBI, 如公式(9)所示:

$$\text{DBI} = \frac{1}{M} \sum_{i=1}^M Q_i, \quad (9)$$

其中  $M$  表示聚类的类别数量。DBI 数值越小表示 K-means 算法的分类效果越理想。

#### 3.2 初始聚类中心优化

由于 K-means 算法<sup>[10]</sup>对初始聚类中心较为敏感, 随机初始化容易使算法陷入局部最优, 影响人体部件分类结果, 进而导致检测模型性能下降。因此, 本文基于 MSE 和 DBI, 构建包含两阶段适应度函数的混合粒子群聚类算法。此外, 为了保证各代粒子的种群多样性, 提高算法全局寻优性能, 引入变异算子, 并采用标准差对种群的离散程度进行描述, 进而动态地调节各代种群的变异概率。

粒子群优化 (Particle Swarm Optimization, PSO) 是受鸟群捕食行为启发而提出的一种仿生智能算法<sup>[19]</sup>。为平衡其全局和局部搜索能力, 对惯性权重  $\omega \in [\omega_{\max}, \omega_{\min}]$  采取自适应调节策略, 如式(10)~(12)所示:

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_{1j}(t)(p_{ij}(t) - x_{ij}(t)) + c_2 r_{2j}(t)(p_{gj}(t) - x_{ij}(t)), \quad (10)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1), \quad (11)$$

$$\omega_t = \begin{cases} \omega_{\min} + \frac{(\omega_{\max} - \omega_{\min}) \times (f - f_{\min})}{f_{\text{avg}} - f_{\min}}, & f \leq f_{\text{avg}}, \\ \omega_{\max}, & f > f_{\text{avg}} \end{cases}, \quad (12)$$

其中:  $c_1, c_2$  为常数,  $r_{1j}$  和  $r_{2j}$  为  $0 \sim 1$  区间的随机数,  $p_{ij}$  与  $p_{sj}$  表示个体及群体最优位置,  $v_{ij}$  和  $x_{ij}$  表示粒子速度和位置。公式(12)用于计算第  $t$  代惯性权重  $\omega_t$ , 其中,  $f$  代表粒子适应度,  $f_{\text{avg}}$  和  $f_{\min}$  代表上一代粒子平均适应度及最小适应度。

受遗传算法启发,对粒子群算法引入变异算子用于应对算法陷入局部最优的情况。为了获取各代种群的变异概率,采用标准差  $\sigma$  计算粒子之间的离散程度,其值越小表示粒子多样性越差,计算公式如下:

$$\sigma = \sqrt{\frac{1}{P} \sum_{i=1}^P (f_i^t - f_{\text{avg}}^t)^2}, \quad (13)$$

其中:  $P$  表示种群规模,  $f_i^t$  和  $f_{\text{avg}}^t$  分别代表第  $t$  代种群的粒子适应度和平均适应度。由此可定义第  $t+1$  代粒子的变异概率  $G_{t+1}$  为:

$$G_{t+1} = \frac{\lambda}{\lambda + P \cdot \sigma}, \quad (14)$$

其中  $\lambda$  表示比例增益,取值范围为  $[2, 4]$ 。对第  $t+1$  代粒子进行逐个判断,若  $\text{rand} \in [0, G_{t+1}]$ ,则对其当前位置执行高斯变异运算,即为  $x_{ij} = p_{sj} \times (1 + 0.5 \times \text{randn})$ ,  $p_{sj}$  为第  $t$  代种群最优位置。

前文中所提 HPSC 算法流程如下所示:

Step1: 定义自适应变异粒子群算法参数,设置  $T$  表示聚类中心数量,种群规模  $P$  为 50,最大迭代次数为 200。

Step2: 定义收缩因子  $c_1 = c_2 = 2$ , 惯性权重上下限  $\omega_{\max}$  和  $\omega_{\min}$  分别为 0.9 和  $0.4^{[20]}$ 。

Step3: 针对不同的父子部件,设置位置和速度区间为  $\text{Var}_{\max}$  和  $\text{Var}_{\min}$ , 其具体取值依据全部训练样本中某一对父子部件相对坐标的最大值和最小值而定。

Step4: 按照初始化种群规模设置如下参数:

(1) 初始化粒子位置,生成一个  $m \times T$  的矩阵,其中各个元素的数值在  $\text{Var}_{\max}$  和  $\text{Var}_{\min}$  之间随机生成,  $m$  代表部件数据维数为 2;

(2) 通过计算各个粒子的 DBI 得到相应的初始适应度值,记为  $\text{Cost1}$ ;

(3) 定义粒子初始速度为  $m \times T$  的矩阵;

(4) 定义  $\text{Cost2}$  为各个粒子的 MSE,其数值等于粒子的  $T$  个  $\text{MSE}_i$  之和。

Step5: 选择 DBI 数值最小的粒子作为初始全局最优粒子。

Step6: 循环执行 Step7 与 Step8,直到满足终止条件或达到最大迭代次数。

Step7: 对第  $i$  个粒子执行以下过程:

(1) 通过公式(10)更新粒子速度;

(2) 利用公式(11)更新粒子位置;

(3) 通过公式(9)和公式(7)分别计算当前粒子的 DBI 及 MSE;

(4) 若粒子  $i$  的  $\text{Cost1}$  小于其最优  $\text{Cost1}$ ,或者粒子  $i$  的  $\text{Cost1}$  等于其最优  $\text{Cost1}$ ,且  $\text{Cost2}$  小于其最优  $\text{Cost2}$ ,则更新粒子  $i$  的最优位置;

(5) 按照(4)的规则更新种群的全局最优位置;

Step8: 若  $i > P$ ,则通过公式(12)更新  $\omega$ 。根据公式(14)计算  $G_{t+1}$ ,并对满足条件的粒子执行变异运算,且  $t = t + 1$ ;否则  $i = i + 1$ ,返回执行 Step7。

Step9: 若满足循环终止条件或达到最大迭代次数,则以当前聚类中心为初始聚类中心,执行 K-means 算法;否则返回 Step6 继续执行算法。

根据所提 HPSC 算法即可获得更加合理的 K-means 算法初始聚类中心,完成对人体部件类型聚类的优化,切实提高检测模型性能。

## 4 实验过程及结果分析

### 4.1 实验环境及测试数据集

本文分别使用 INRIA 数据集<sup>[2]</sup>和 ETH 数据集<sup>[15]</sup>对所提行人检测方法的有效性进行验证。INRIA 数据集图像清晰度较高,其测试集中包括正样本图像 288 张,分辨率为  $447 \times 358$  至  $1201 \times 960$ ,负样本图像 453 张,分辨率为  $320 \times 240$  至  $640 \times 480$ 。ETH 数据集包含 3 个子集,本文采用较为常用的 Setup1。该子集包含 3 个视频序列帧(Seq01~Seq03),分别包含 999、936 和 354 帧图像,其分辨率均为  $640 \times 480$ 。

由于检测模型的训练需要有效的人体部件信息,因此本文采用全部的 Image Parse 数据集<sup>[17]</sup>样本作为训练正样本。该数据集共包含 305 张标注有全身人体关节的图像。此外,将 INRIA 数据

集的训练集负样本作为检测模型的训练负样本,因为其中不包含任何有效的人体结构。

本文的硬件平台为 Intel Core i7-5930K CPU, 3.50 GHz, 64 GB 内存;软件环境采用 Windows 系统下的 Matlab 2014a。

### 4.2 各部件类型数量计算及聚类

基于 26 个关节点的人体树图结构训练行人

检测模型,如图 1 所示。各部分由肢体末端向头部依次互为父子关系,头部件为根节点,例如部件 2 为部件 1 的子部件;部件 15 有 2 个子部件,为部件 20 和部件 16。

依据上述 3.1 和 3.2 部分内容,计算各部件类型数量,并获取 K-means 初始聚类中心,进而完成部件类型聚类,具体步骤如下:

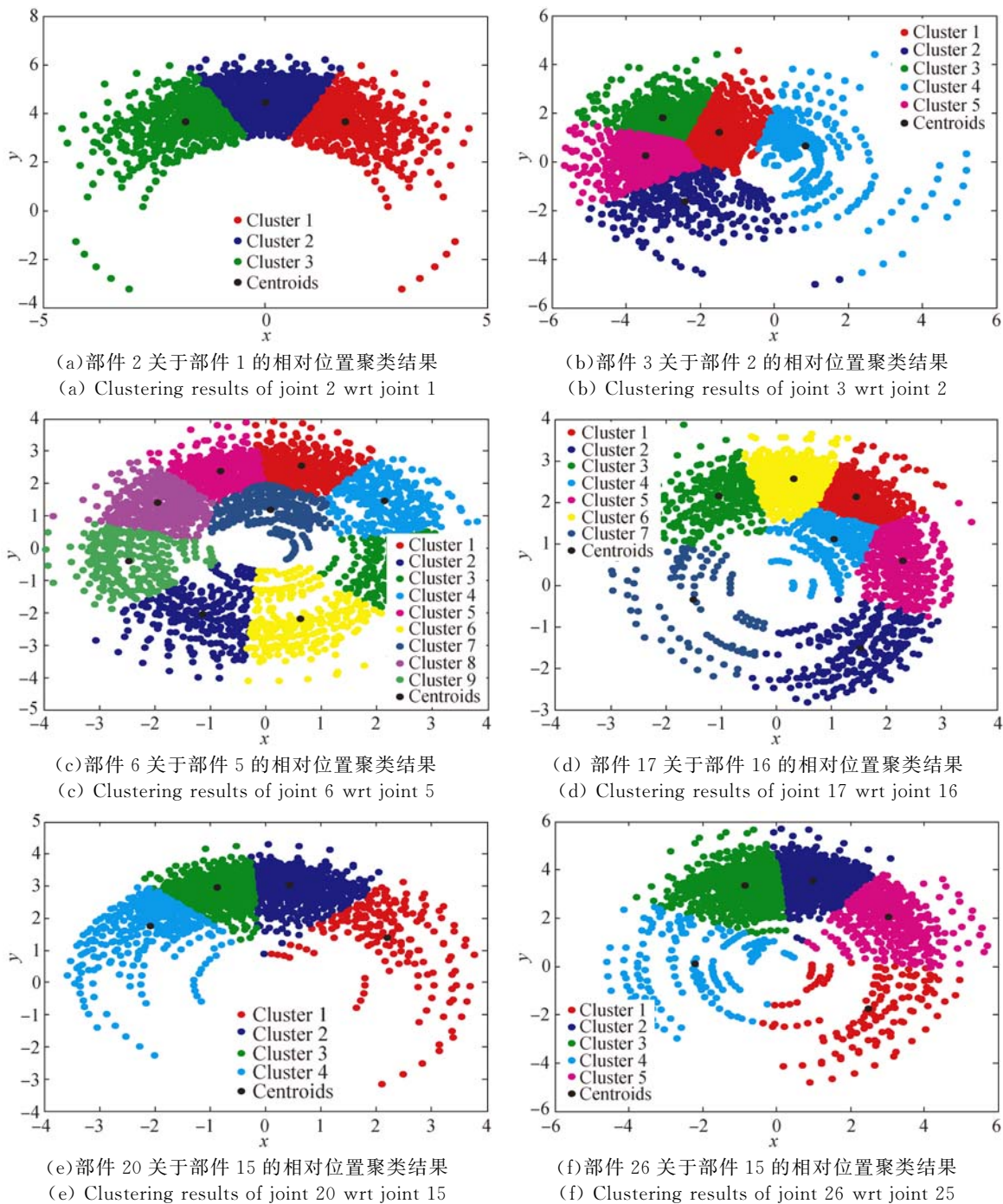


图 2 部分人体部件类型聚类结果

Fig. 2 Clustering results of partial body parts

首先,设置各部件(根节点除外)的类型数量范围,即  $P_{\min}=3, P_{\max}=9$ 。为了进一步增加训练正样本数量,提高模型的泛化能力,将 305 张训练正样本沿水平向左和向右两个方向,分别依次旋转  $15^\circ$  和  $7.5^\circ$ ,并转化对应的部件标注信息,使得训练正样本数量扩充为 1 525 张。之后采用镜像处理,将训练正样本扩充为 3 050 张;

其次,基于图 1 所示部件父子关系和训练正样本中人体的标注信息,通过公式(6)~(9)获取各部件类型数量在区间  $[P_{\min}, P_{\max}]$  内的最优值。计算得到各部件类型数量依次为:3、5、5、7 和 9 (部件 2~6);9、4、4、4 和 6 (部件 7~11);6、6、5、5 和 5 (部件 12~16);7、9、9、4 和 4 (部件 17~21);4、6、6、6 和 5 (部件 22~26);

最后,对各父子部件的相对位置进行聚类。在兼顾类内紧密性和类间分离性的情况下,利用所提 HPSC 算法,获取 K-means 算法的初始化聚类中心。之后通过执行 K-means 算法完成对人体部件类型的聚类优化。部分人体部件类型聚类结果如图 2 所示。

根据实验结果,鉴于训练样本做了旋转和镜像处理,左半人体和右半人体的部件坐标位置呈对称关系,因此对应的部件类型聚类结果基本一致,如部件 3 和 15,26 和 14,以及 19 和 7 等。此外,不同部件的类型聚类结果和聚类中心数量均存在较大差异(如部件 2 与部件 6,部件 3 与部件 20),进而证明了所提部件类型数量优化和 HPSC 算法的必要性和有效性。

### 4.3 检测性能分析

为了检验所提行人检测方法的性能,基于人体树图模型,将所提方法在 INRIA 数据集<sup>[2]</sup>和 ETH 数据集<sup>[15]</sup>上进行了验证。

根据 4.2 部分所得各部件类型数量及聚类结果,对每个训练样本的部件类型进行重新定义,获取隐藏变量。基于公式(1)~(3)的隐结构 SVM 框架对行人检测模型进行训练。之后,结合公式(4)和(5),利用动态规划思想<sup>[14]</sup>,对图像进行多尺度下的行人检测,并结合 NMS 算法,得到全部的最优人体部件位置,完成行人检测任务。在两个标准数据集上的检测示例如图 3 所示(彩图见期刊电子版)。

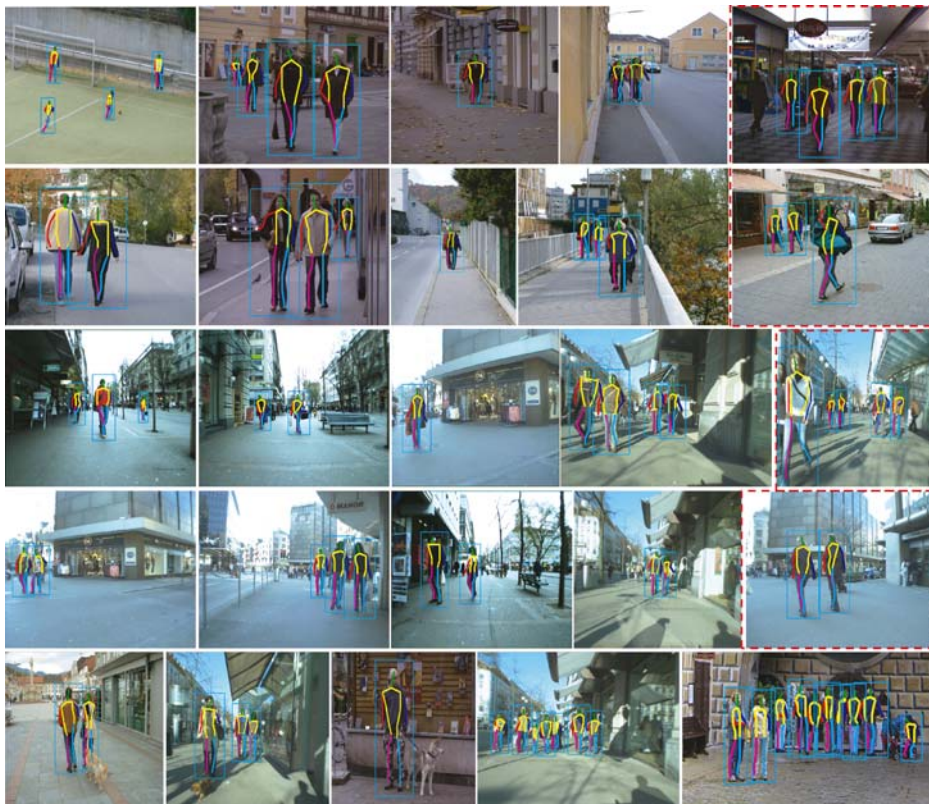


图 3 INRIA 和 ETH 数据集上的行人检测结果示例

Fig. 3 Examples of pedestrian detection results on INRIA and ETH data sets

图 3 中第 1、2 行为数据集 INRIA 上的检测结果示例;第 3、4 行为数据集 ETH 上的检测结果示例;第 5 行为两个数据集中,部分特殊样本的测试结果展示。根据实验结果可以得到以下结论:

首先,ETH 数据集的检测环境更加贴近于实际交通场景,大量行人目标受到光照、尺度及遮挡因素的影响,检测难度较大。INRIA 数据集的目标人体相对分散,且分辨率普遍较高,光照影响弱,检测难度相对较小;

其次,融合部件类型数量优化和 HPSC 算法训练得到的人体树图模型,能够有效地检测出行人目标,具有较强的鲁棒性和准确性。图中的彩色线条为检测出的人体部件连线,其连接次序与图 1 中的人体树图结构一致。对于第 5 行展示的特殊测试样本,所提方法依然能够检测出其中的人体目标。如第 5 行中 3 个牵着狗的测试图像,尽管狗相比于人体同样具有四肢和头部,但其姿态与常规的行人树图结构存在较大差别,因此并没有被错误地检测为行人目标。在第 5 行 2 个行人较为密集测试图像中,尽管存在部分人体被人群遮挡的情况,但是检测模型基于部件间的共生约束关系,同样准确、完整地检测出了不同姿态下的人体目标;

最后,基于人体树图的检测方法能够良好地应对复杂的人体形变,具有较强的容错性。即少数部件位置的估计失败并不影响最终行人检测结果的准确性。如图 3 第 2 行中红色虚线所框出的图像所示,由于背包行人的左侧手臂被包遮挡,所以导致检测模型对于该手臂上的人体部件位置估计失败。但是由于检测模型中融入了父子部件之间的共生约束关系,所以在大部分人体部件位置估计正确的情况下,错误估计的手臂位置不会脱离正常的行人姿态,即手臂依然会连接着人体躯干,且其上所有部件会参照头部部件与颈部部件(参见图 1 中的部件 1 和 2)之间的距离成比例的分布,从而使得行人检测结果依然是准确的。

此外,所提方法对于分辨率低或个体颜色与背景趋于一致的目标,存在部分检测失败的情况。如图 3 第 1、3 和 4 行中,红色虚线所框出的图像所示。具体而言,由于检测过程主要基于图像的 HOG 特征,其本质是梯度的统计信息,而梯度主要存在于图像中的边缘区域。因此,对于人体轮廓像素值与背景像素值较为接近的情况,HOG 特征的判别性能将会显著下降,进而降低了检测模型对该类行人目标的检测成功率。如第 1 行中

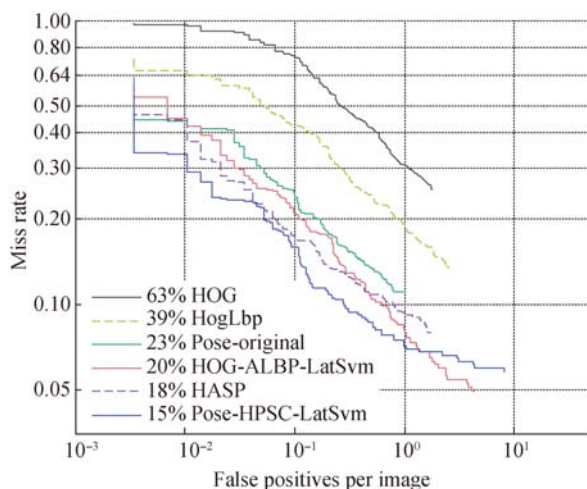
左侧目标的着装偏暗,与昏暗的背景像素值较为接近,并且头部与背景木框的像素值相近,使得该区域检测出的人体树图结构评分过低,导致模型对该目标的检测失败。同理,第 4 行中右侧的两个人体着装分别与各自背景的像素值相近,并且由于图像分辨率较低使得人体轮廓相对模糊,导致该区域并没有检测出有效的人体树图结构。

图 3 第 3 行中右侧人体尽管色彩不同于背景,但是由于图像分辨率较低,用于描述弱小目标边缘的像素点较少,其边缘变得相对模糊,导致该区域同样没能检测出有效的人体树图结构,进而使得对该目标的检测失败。

#### 4.4 检测结果比较

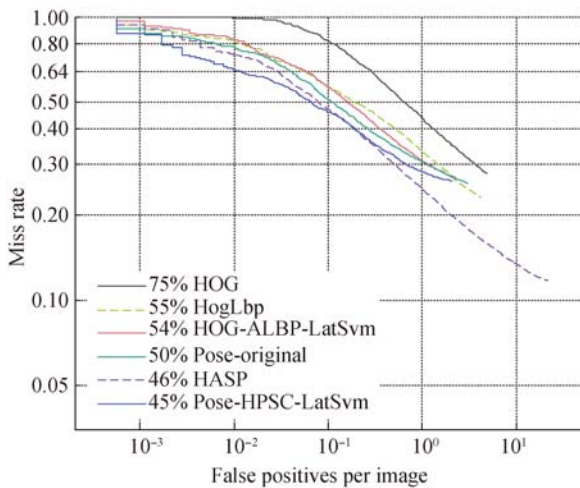
为了进一步验证所提方法的先进性,选取 5 种相关行人检测算法进行对比分析。漏检率 (Miss Rate)-FP 曲线 (False Positives Rate) 是 ROC 曲线的一种表现形式,被普遍应用于行人检测、目标跟踪和显著性检测等算法的性能评估。本文采用 Miss Rate-FPPI (FP Per Image) 曲线对多种行人检测算法进行对比分析。

所选取的对比算法包括:HOG<sup>[2]</sup>,HogLbp<sup>[21]</sup>,HOG-ALBP-LatSvm<sup>[22]</sup>,HASP<sup>[23]</sup>和文献[9]的方法(Pose-original)。上述方法均采用了 HOG 特征,其中 HOG-ALBP-LatSvm、HASP 和 Pose-original 在训练过程中均使用了隐结构 SVM。此外,HOG 和 HogLbp 的实验数据来源于 Caltech Pedestrian Detection Benchmark<sup>[24]</sup>。所提 Pose-HPSC-LatSvm 与其余 5 种方法,在 INRIA 和 ETH 数据集上的 Miss Rate-FPPI 曲线如图 4 所示。



(a) INRIA 数据集上的 Rate-FPPI 曲线

(a) Rate-FPPI curve on INRIA dataset



(b) ETH 数据集上的 Rate-FPPI 曲线

(b) Rate-FPPI curve on ETH dataset

图 4 INRIA 和 ETH 数据集上的 Rate-FPPI 曲线

Fig. 4 Rate-FPPI curve on INRIA and ETH data sets

对于 INRIA 数据集, Pose-HPSC-LatSvm 的丢失率为 15.43%, 远低于传统的 HOG-SVM 方法。相比于同样采用了 HOG 特征和隐结构 SVM 的 HOG-ALBP-LatSvm、HASP 算法, 本文所提方法的丢失率分别降低了 4.76% 和 2.84%。

对于 ETH 数据集, 所得对比结果与 INRIA 数据集类似。Pose-HPSC-LatSvm 相比于 HOG-ALBP-LatSvm 和 HASP 算法, 丢失率分别下降了 9.2% 和 1.29%, 并且明显优于传统基于 HOG 和 HogLbp 特征的行人检测方法。

此外, 所提方法在检测性能上明显优于的原始 Pose-original 方法, 在 INRIA 和 ETH 数据集上的丢失率分别下降了 8.14% 和 5.05%。上述对比结果进一步验证了所提 Pose-HPSC-LatSvm 方法的有效性和准确性。

## 参考文献:

- [1] VIOLA P, JONES M J. Robust real-time face detection [J]. *International Journal of Computer Vision*, 2004, 57(2): 137-154.
- [2] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Diego, USA, CVPR, 2005: 886-893.
- [3] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling [C].

## 5 结 论

为提高交通环境下行人检测的准确率, 本文提出了一种基于人体树图模型的行人检测方法。通过定义部件间的共生关系与局部刚度, 增强模型对于复杂人体形变的应对能力。在部件类型聚类过程中, 为兼顾类内紧密性与类间分离性, 弥补 K-means 算法对初始聚类中心依赖性强且类别数量需人为干预的缺陷, 引入 MSE 和 DBI 构建具有两阶段适应度函数的混合粒子群算法。之后, 利用坐标下降法求解隐结构 SVM 获取行人检测检测模型。在线上检测阶段, 结合动态规划算法, 对图像进行多尺度检测, 通过非极大值抑制得到最终的行人检测结果。实验结果表明, 通过定量和定性分析, 所提方法相比于其他 5 种相关行人检测算法具有明显性能优势。在 INRIA 和 ETH 数据集上的丢失率与原始 Pose-original 方法相比分别下降了 8.14% 和 5.05%, 证明了所提改进策略的有效性, 以及整体行人检测方法的准确性与鲁棒性。

由于所提行人检测方法从根本上依靠图像的 HOG 特征, 而 HOG 特征性能又对图像中的边缘信息较为敏感, 导致在图像分辨率较低或目标像素值与背景趋于一致时存在检测失败的情况。未来将对图像提取更多互补性特征, 以丰富行人目标信息, 进一步提高检测算法的鲁棒性。此外, 将对模型的训练样本数量进行扩充, 并增加被遮挡人体和弱小目标人体的训练样本比例, 提高模型对该类目标的检测成功率。

2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, San Francisco, USA, CVPR, 2010: 32-39.

- [4] FELZENSZWALB P F, MCALLESTER D, RAMANAN D. A discriminatively trained, multi-scale, deformable part model [C]. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, USA, CVPR, 2008: 1-8.
- [5] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, *et al.*. Object detection with discriminatively trained part-based models [J]. *IEEE Trans-*

- actions on *Pattern Analysis & Machine Intelligence*, 2010, 32(9): 1627.
- [6] PEDERSOLI M, VEDALDI A, GONZALEZ J, *et al.*. A coarse-to-fine approach for fast deformable object detection [J]. *Pattern Recognition*, 2015, 48(5): 1844-1853.
- [7] OUYANG W, ZHOU H, LI H, *et al.* Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, PP(99): 1-1.
- [8] OUYANG W, ZENG X, WANG X, *et al.*. DeepID-Net: object detection with deformable part based convolutional neural networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, 39(7): 1320-1334.
- [9] YANG Y, RAMANAN D. Articulated human detection with flexible mixtures of parts [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2013, 35(12): 2878.
- [10] 潘楠, 孙亚军, 刘益. 线缆盗割痕迹的激光检测特征高效匹配[J]. *光学精密工程*, 2017, 25(10s): 183-190. PAN N, SUN Y J, LIU Y. Efficient matching technique for laser detection features of cable cutting traces [J]. *Opt. Precision Eng.*, 2017, 25(10s): 183-190. (in Chinese)
- [11] WU J, LIU H, XIONG H, *et al.*. K-Means based consensus clustering: a unified view [J]. *IEEE Transactions on Knowledge & Data Engineering*, 2015, 27(1): 155-169.
- [12] CELEBI M E, KINGRAVI H A. A comparative study of efficient initialization methods for the k-means clustering algorithm [J]. *Expert Systems with Applications*, 2012, 40(1): 200-210.
- [13] 李庆武, 席淑雅, 王恬, 等. 结合位姿约束与轨迹寻优的人体姿态估计[J]. *光学精密工程*, 2017, 25(4): 1060-1069. LI Q W, XI SH Y, WANG T, *et al.*. Human pose estimation based on configuration constraints and trajectory optimization [J]. *Opt. Precision Eng.*, 2017, 25(4): 1060-1069. (in Chinese)
- [14] 黄建余, 谭继波, 吴文堂, 等. GEO 碎片天基观测图像速度估计与动态规划检测[J]. *光学精密工程*, 2017, 25(2): 537-546. HUANG J Y, TAN J B, WU W T, *et al.*. Velocity estimation of GEO debris space-based observation image and dynamic programming detection [J]. *Opt. Precision Eng.*, 2017, 25(2): 537-546. (in Chinese)
- [15] ESS A, LEIBE B, SCHINDLER K, *et al.*. A mobile vision system for robust multi-person tracking [C]. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Anchorage, USA, CVPR, 2008: 1-8.
- [16] NIE B X, XIONG C, ZHU S C. Joint action recognition and pose estimation from video [C]. 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA, CVPR, 2015: 1293-1301.
- [17] RAMANAN D. Learning to Parse Images of Articulated Bodies [C]. *Proc. Advances in Neural Information Processing System*, Vancouver, Canada, NIPS, 2007: 1-8.
- [18] PIVIDORI M, STEGMAYER G, MILONE D H. Diversity control for improving the analysis of consensus clustering [J]. *Information Sciences*, 2016, s361-362(C): 120-134.
- [19] FONG S, WONG R, VASILAKOS A V. Accelerated PSO swarm search feature selection for data stream mining big data [J]. *IEEE Transactions on Services Computing*, 2016, 9(1): 33-45.
- [20] 段海滨, 张祥银, 徐春芳. 仿生智能计算[M]. 北京: 科学出版社, 2011. DUAN H B, ZHANG X Y, XU CH F. *Bio-inspired Computing* [M]. Beijing: Science Press, 2011. (in Chinese)
- [21] WANG X, HAN T X, YAN S. An HOG-LBP human detector with partial occlusion handling [C]. 2009 IEEE International Conference on Computer Vision, Kyoto, Japan, ICCV, 2009: 32-39.
- [22] LIU Y, ZENG L, HUANG Y. An efficient HOG-LBP feature for pedestrian detection [J]. *Signal Image & Video Processing*, 2014, 8(1): 125-134.
- [23] LIU Y, ZOU L, LI J, *et al.*. Segmentation by weighted aggregation and perceptual hash for pedestrian detection [J]. *Journal of Visual Communication & Im-*

*age Representation*, 2016, 36(C): 80-89.

- [24] ANGELOVA A, KRIZHEVSKY A, VANHOUCKE V. Pedestrian detection with a Large-Field-Of-View

deep network [C]. 2015 *IEEE International Conference on Robotics and Automation*, Washington, USA, ICRA, 2015: 704-711.

作者简介:



孟晓燕(1990—),女,山西晋中人,博士研究生,主要从事图像处理与模式识别,智能车辆等方面的研究。E-mail: 529108419@qq.com

导师简介:



段建民(1959—),男,北京人,博士,教授,主要从事车辆环境识别与自动驾驶技术,网络化测控系统与现场总线技术,嵌入式汽车电子控制技术等方面的研究。E-mail: jmduan@bjut.edu.cn