

基于区域复合概率的行人候选框生成

章 剑, 肖 婷

(重庆大学通信工程学院, 重庆 400044)

摘 要: 行人检测在汽车驾驶辅助系统和智能视频监控等领域有广泛的应用, 而行人候选框的生成是行人识别、定位及跟踪的一项重要前期工作. 本文提出一种基于区域复合概率(Local Mixture Probability, LMP)模型的在线生成行人候选框的方法. 该方法根据区域相似性将监控场景划分为多个子区域, 随之对各区域内行人的位置和尺度分别建立泊松模型和高斯模型. 通过模型的学习与更新可以获取目标出现的概率信息以及目标尺度的分布情况, 从而为候选框的生成提供依据, 避免遍历搜索的盲目性. 实验结果表明, 该算法能够在生成较少数目候选框的情况下获得较高的覆盖率.

关键词: 机器视觉; 行人检测; 区域复合概率模型; 生成候选框

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2018)07-1719-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2018.07.025

Pedestrian Proposal Generation Based on Local Mixture Probability Model

QIN Jian, XIAO Ting

(College of Communication Engineering, Chongqing University, Chongqing 400044, China)

Abstract: Pedestrian detection is widely applied in driver assistance systems and video surveillance fields, while proposal generation is a significant preliminary work for pedestrian recognition and tracking. This paper proposes a method for fast online proposal generation using Local Mixture Probability (LMP) model. Poisson model and Gaussian model are separately established for online learning location and scale of pedestrians after region-dependent segmentation according to local similarity. Based on learning and updating models, both the probability of pedestrians occurrence and the probability distribution of the scale in specific regions can be obtained, which provides bases for pedestrian proposal generation and avoids searching blindness. Experiments on Caltech Pedestrian dataset show that LMP can achieve higher recall by fewer pedestrian detection proposals.

Key words: machine vision; pedestrian detection; LMP model; proposal generation

1 引言

行人检测在智能视频监控、智能汽车及交通中都有着重要的应用. 通过对行人的检测、跟踪、轨迹分析及行为识别, 系统可检测异常事件并报警^[1]. 行人检测与定位的准确性直接影响后续工作乃至整个系统的性能. 一般, 行人检测主要由“生成检测候选区域”和“行人识别”两个部分组成. 即先在图像帧中生成可能的目标框, 然后采用识别算法对该目标框进行检测并识别是否为行人. 其中“生成检测候选区域”是识别行人的一项重要的前期工作, 对最终的行人检测的准确率和

效率均有很大的影响. 如果前期生成的候选框与真实目标的覆盖度较大, 且候选框的数目能尽量少, 就会对整个行人检测系统产生正面的作用, 即提升行人检测的准确率, 并缩短搜索的时间^[2].

针对该问题, 学者提出了多种生成行人候选框的方法. 如 Bing^[3] 和 Edge Boxes^[4] 这类基于单帧图像信息的无类别候选框生成方法, 主要利用单帧图像中的边缘、纹理等特征信息. Bing^[3] 训练了一个简单的线性分类器通过类似滑动窗的方式来过滤候选框, Edge Boxes^[4] 则通过计算窗口内边缘个数进行打分, 最后排序得到候选区域. 该方法处理速度快, 但是不适合用于智

能监控、无人驾驶等需要进行特定目标检测的应用场景中. 另一类算法考虑到行人检测中的目标大多处于运动状态, 因此先检测出运动目标作为候选框, 再进行识别. 如文献[5~7]等利用运动目标的光流随时间变化的特性, 把光流计算得到的运动场作为一个重要的识别特性来检测运动目标. 该方法虽然能够描述运动信息, 但对于场景中一些静止或者运动缓慢的目标, 该方法几乎不能捕捉到信息, 无法有效地生成候选框. 还有一类方法则采用遍历搜索的方式, 如滑动窗搜索方式^[8], 但该方法会产生数以万计的候选框, 计算量太大从而降低了目标检测的效率; 后来发表的 Selective Search^[9]方法综合了蛮力搜索和图像分割, 旨在找出可能的目标位置来进行物体的识别. R-CNN^[10]与 Fast R-CNN^[11]检测器均采用 Selective Search 算法生成目标候选框, 处理速度慢. 此外, 一些算法通过共享候选框生成算法和检测算法的信息从而降低运算量. 如 Faster R-CNN^[12]基于 Fast R-CNN^[11]引入了 Region Proposal Networks (RPN), 与检测网络共享整幅图片的卷积特征, 虽然加快了处理速度, 但是由于候选框生成算法与目标识别算法紧密耦合, 该方法没有普适性. 而 OL-GMPG^[13]方法采用对研究场景中的每个像素点建立高斯模型的方式获取行人的位置和尺度信息进而生成候选框, 但其未考虑目标分布的局部相似性及区域合并, 复杂度高.

在静态监控及车载监控等场景中, 上述方法均存在一定局限性, 本文提出了一种基于区域复合概率 (Local mixture probability, LMP) 模型的新算法. LMP 算法充分利用研究场景的行人分布的信息以及视频图像序列之间的相关性, 将监控场景划分成多个具有局部相似性的子区域, 训练高斯泊松复合模型, 预测出现在每个区域的行人的密度和尺度信息. 在之后的候选框生成过程中, 对各个区域进行不同步长的滑动窗搜索, 其中通过泊松模型提供的位置信息调整步长, 而多尺度搜索则依据高斯模型提供的尺度信息实现. 此外该算法采用了在线学习机制, 模型更新与检测同步进行, 处理速度快. 针对 Caltech Pedestrian 数据库的 set00 数据集, 当在每幅图片上平均设置 5198 个行人候选框时, 获得了 90.9% 以上的平均覆盖率.

2 区域复合概率模型原理

行人候选框的生成问题实际上可以视为求解视频中行人可能出现位置 L 、表观尺度 S 的分布 $F(L, S)$, 如果预测值逼近真实值, 则可以以尽可能少的候选框最大限度地准确覆盖所有真实的目标, 降低识别运算量并提高准确率. 其中, L, S 均为时间 t 的函数, 因此求解问题可写为 $F(L(t), S(t))$. 本文的 LMP 算法通过区域

分块和高斯泊松复合模型来进行求解.

2.1 区域分块

在静态监控及车载监控等场景下, 行人出现的位置和频率分布具有局域一致性, 即在某区域内行人的尺度分布不会有太大的差别, 行人出现的频率应是稳定或缓慢变化的. 因此, 可以将场景划分为多个具有相似性的子区域然后分别建立模型. 此外, 由于摄像头的透视效应, 行人的尺度分布在整个场景中并不一致, 近处行人尺度较大, 远处行人尺度较小. 所以, 在整个场景中采用相同的尺度及步长搜索会造成大量冗余, 而通过区域分块并单独建模则可以避免.

LMP 算法首先将监控场景划分为多个具有局部相似性的子区域. 对于场景 R , 将其划分为 k 个区域, 分别为 $\{R_1, R_2, R_3, \dots, R_k\}$, 则整个场景中行人候选框的生成问题可转化为对各区域 $R_i (i = 1, 2, \dots, k)$ 分别建模分析. 即: $F(L(t), S(t)) \Rightarrow \bigcup_{i=1}^k F_{R_i}(L_i(t), S_i(t))$

实际处理中, 考虑到分块处理的复杂度以及研究场景的特点, 没有必要将研究场景进行过度的细分, 本文直接将场景分成 $k = 12$ 份 (3×4), 如图 1 所示.

1	4	7	10
2	5	8	11
3	6	9	12

图1 区域分块示意图

2.2 高斯泊松复合模型

行人候选框的生成问题实际上就是预测在当前帧中行人的可能位置参数和尺度参数, 可分别建立模型进行求解.

2.2.1 位置模型

位置模型用于拟合行人的位置信息, 其中, 预测当前帧是否有行人出现及出现的位置可以转换为求取各子区域下一时间段行人是否会出现的问题, 即通过对每个子区域建立位置模型求解相应区域下一时间段行人出现的概率. 实际处理中, 我们将该问题转化为求解某区域在一段时间内行人出现的次数.

泊松过程是描述随机事件累计发生次数的基本数学模型之一, 在应用中很多场合都满足这些条件, 可以假设监控场景某区域一段时间内行人出现的次数是一泊松过程.

设 $\{N(t), t \in T = [0, +\infty)\}$ 是一强度为 λ 的泊松过程, 则对任意固定的 $t > 0$, $N(t)$ 服从泊松分布 $\pi(\lambda t)$, 即 $P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k = 0, 1, 2, \dots$ 泊松分布的

参数 λ 是单位时间(或单位面积)内随机事件的平均发生率.

假设监控场景中某区域行人出现的次数服从泊松分布,即建立行人位置模型 $L_i(t) \sim \pi(\lambda t)$,泊松分布的期望和方差均为 λ ,则通过该区域泊松分布的期望可以间接预测出该区域出现的行人次数.本文通过对前 m 帧图像每 t 时间内各子区域出现的行人次数进行泊松估计得到 λ 的初始值 λ_0 ,随之在线学习行人在各区域 t 时间内出现的次数,并实时更新泊松过程的均值与方差.参数的在线更新可采用梯度下降法并利用前 t_p 时间内行人出现的次数进行迭代,如式(1):

$$\lambda_{t+1} = \lambda_t - \beta \left(1 - \frac{\sum_{i=t-t_p}^t x_i}{\lambda t_p}\right) \quad (1)$$

上式中, λ_t 与 λ_{t+1} 分别为 t 时刻与 $t+1$ 时刻泊松分布的均值, β 为模型的学习率.

2.2.2 尺度模型

众所周知,高斯分布(正态分布)是一个在很多领域得到广泛应用的重要分布.现实中许多随机变量均服从或者近似服从高斯分布.

若随机变量的概率密度服从函数:

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), x \in (-\infty, +\infty) \quad (2)$$

其中, μ 和 σ 是常数, $\sigma > 0$,则 X 服从高斯分布,记为 $x \sim N(\mu, \sigma^2)$,称 x 为正态随机变量.当 $\mu=0, \sigma=1$ 时,该分布称为标准正态分布.

针对静态监控或车载监控场景,假设各分块区域内行人的高度近似服从正态分布,即建立行人尺度模型 $S_i(t) \sim N(\mu, \sigma^2)$,则可以用数学期望 μ 和标准差 σ 两个参量来进行描述.在该假设下,各区域高斯分布的期望 μ 即为该处出现概率最大的行人的高度,而标准差 σ 决定了高斯分布的幅度大小.本文通过对前 m 帧图像的行人高度进行正态估计分析得到各区域高斯模型的均值 μ_0 及标准差初始值 σ_0 .针对该块区域,随着每帧图片出现行人高度 S_i 的引入,高斯模型的均值与标准差被不断更新.具体更新过程如下式(3)和(4):

$$\mu_{t+1} = (1-\alpha) \times \mu_t + \alpha \times S_t \quad (3)$$

$$\sigma_{t+1} = \sqrt{(1-\alpha) \times \sigma_t^2 + \alpha \times (S_t - \mu_t)^2} \quad (4)$$

上式中, μ_t 与 μ_{t+1} 分别为 t 时刻与 $t+1$ 时刻高斯模型的均值, σ_t 与 σ_{t+1} 分别为 t 时刻与 $t+1$ 时刻高斯模型的标准差, α 为模型学习率.

3 区域复合概率模型算法

LMP 算法通过对每个分块区域建立高斯泊松复合模型并在线学习更新可以得到各区域行人位置和尺度

信息,以不同步长生成对应不同尺度的行人候选框.根据泊松模型学习得到的强度参数 λ 可以预测区域内出现行人的次数 n 从而设置生成候选框的密度.如表 1 所示,分别对不同密集度的区域设置不同的候选框平移的步长 step.

表 1 不同密集度区域中候选框生成的相应处理

区域处理	稀疏分布区域 $n \leq n_{\min}$	中等分布区域 $n_{\min} < n < n_{\max}$	密集分布区域 $n \geq n_{\max}$
步长设置(step)	l_1	l_2	l_3

其中, $l_1 > l_2 > l_3$. 然后分别对各区域出现的行人的高度建立高斯模型,通过在线学习更新得到出现在各区域的行人尺度信息.考虑到尽量减少候选框获得高覆盖率,我们以相同的步长在行人基准高度正负方向设置 3 个候选框,能够实现良好的效果.具体地,将模型更新得到的均值作为基准尺度,即 $h_s = \mu$,然后按照一定的比例 *ratio* 对候选框进行放大和缩小,实现多尺度搜索,候选框尺度的上界和下界值计算如下:

$$h_{\max} = h_s (1 + \text{ratio}) \quad (5)$$

$$h_{\min} = h_s (1 - \text{ratio}) \quad (6)$$

本文中,候选框的宽高比锁定为 k .

LMP 算法生成候选框的流程如图 2 所示.

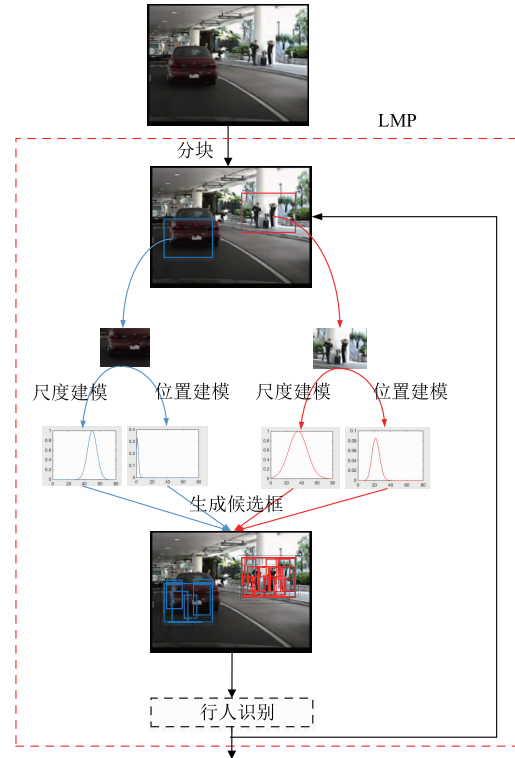


图2 LMP算法生成候选框示意图

(1) 区域分块. 如图 1 所示,根据场景复杂度将检测视频分成 N 个子区域.

(2) 建立初始模型. 对目标位置建立泊松模型, 通过对前 m 帧图像中每 t 时间内出现的行人个数进行统计学习并进行参数估计, 得到泊松模型参数初始值. 对目标尺度建立高斯模型, 通过对前 m 帧图像出现的行人尺度大小进行统计分析并进行参数估计, 得到高斯模型参数初始值.

另外, 考虑到实际情况, 开始模型学习时, 由于输入的样本数目不足, LMP 模型暂未收敛. 因此开始时采用传统滑动窗搜索方式生成候选框, 其中, 传统滑动窗搜索方法的搜索窗基准尺寸初始值即为高斯模型的均值初始值. 滑动窗搜索生成候选框的同时更新 LMP 模型参数, 直至前一段时间每帧图片的平均覆盖率超过设定的阈值时将转换为 LMP 模型.

(3) 参数在线更新. 根据式(1)更新各区域位置泊松模型强度 λ , 从而预测各检测区域 t 时间内行人出现的次数 n . 根据式(3)和式(4)更新各区域尺度高斯模型均值 μ 及标准差 σ , 基于均值 μ 设置各分块检测区域中行人尺度的基准, 根据标准差 σ 设置各区域尺度变化的步长, 综合基准及变化步长设置候选框的尺度.

(4) 生成目标候选框. 根据步骤(3)获取的各检测区域 t 时间内行人出现的次数 n , 按照表 1 设置候选框移动步长以控制生成密度, 其中候选框的尺度大小设置是以高斯模型更新的该区域的均值为基准实现的, 根据高斯模型更新的标准差调整各区域设置的 $ratio$ 值, 可以提高行人候选框与标注框的覆盖率. 最后, 应用多尺度滑动窗搜索方式在区域内生成检测框.

(5) 行人识别分类. 将步骤(4)生成的行人候选框输入后续的行人分类器进行识别, 得到识别结果. 再将识别后的目标信息作为新样本, 输入到步骤(3)进行模型参数更新.

4 实验及分析

实验环境: 处理器 Intel Pentium, 内存 4GB, 操作系统 Windows7, 实验平台 MATLAB R2014b. 实验选用 Caltech Pedestrian Dataset^[14] (CPD) 中的 set00 数据集, 总共有 17955 张图片. CPD 是目前规模较大的行人数据库, 共 10 小时, 分辨率为 640×480 , 30 帧/秒, 采用车载摄像头拍摄, 并且对其行人目标的位置大小遮挡等情况进行了详细标注. 本文使用平均每帧图片设置的候选框数目以及平均召回率评估候选框生成的数量及准确度. 当候选框与目标的重合度 (IOU) 大于设定的阈值时, 检测到判定目标. IOU 值越高, 说明候选框与真实目标的重合度越大, 则后续识别算法正确识别目标的概率越大.

4.1 位置模型测试

本文按照图 1 对场景进行分块, 然后对前 2400 帧

图像中每个分块区域每秒(30 帧)出现的行人的次数进行统计然后进行泊松参数估计建立初始模型. 其中, 泊松模型的学习率 $\beta = 0.8$, $t_p = 20$. 本文设置 $n_{\min} = 10$, $n_{\max} = 30$. 因此, 根据表 1, 步长 $l_1 = 10$, $l_2 = 8$, $l_3 = 6$. 输入各分块区域当前的 λ 值以及前 20 秒内行人出现的次数, 根据式(1)分别对各分块区域位置模型进行实时在线更新, 从而估计各分块区域下一时间段(1 秒)内行人出现的次数. 为了测试位置模型的在线学习能力, 我们将每个分块区域学习预测得到的行人出现次数与行人位置分布的真实情况进行了对比.

图 3(a) 是第 10200 帧(第 340 秒)的实际场景, 图 3(b) 是位置模型在对应时刻预测的出现行人次数. 由(b)可知, 行人主要分布在区域 2、8、11, 而其他区域则相对很少. 对照其实际场景(a)可以看到, 区域 2 对应左侧人行道, 区域 8、11 对应右侧人行道, 均为实际场景中行人密集区域; 区域 1、4、7、10 位于视线上方空中, 因此行人很少; 区域 3、6、9、12 位于车前较近位置, 在车辆行驶过程中也很少会出现行人, 因此行人也较少; 而区域 5 比较特殊, 虽然与密集区域 2、8、11 均处于正前方临近区域, 但在第 340 秒左右时, 该区域对应的是前进方向的车道, 因此行人很少出现. 可见, 位置模型较好的拟合了实际情况.

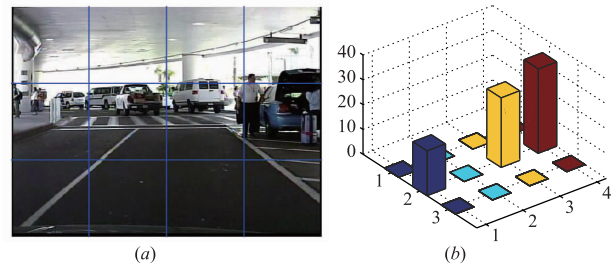


图3 (a)第10200帧(第340秒)的实际场景
(b)预测第340秒行人出现次数

4.2 尺度模型测试

本文假设行人的高度服从正态分布, 同理, 按照图 1 的分块方式对前 2400 帧图像每个分块区域中出现行人的身高进行统计随之进行正态参数估计, 可以得到各区域出现行人高度的初始值.

如图 4(a)(b) 所示分别为第 10200 帧及 16410 帧图片中 12 块区域尺度模型学习得到的均值及标准差. 从中可知, 不同区域尺度的均值不同, 标准差也不相同. 其中, 出现在区域 3、6、9、12 处的行人高度值最大, 区域 2、5、8、11 处次之, 区域 1、4、7、10 处则最小. 根据实际经验, 距离摄像头越远, 行人的尺度越小, 反之越大. 由检测区域的分块处理可知, 区域 3、6、9、12 离摄像头最近, 区域 2、5、8、11 次之, 区域 1、4、7、10 则离摄像头最远, 同样与图 4(a)(b) 中数据吻合. 再者, 通过对比图 4(a)

(b)中不同场景对应的尺度模型学习到的参数值,可以看出尺度模型能够在线学习更新各分块区域的尺度信息.因此,通过尺度模型的学习,我们也可以直观的得知行人的远近信息.

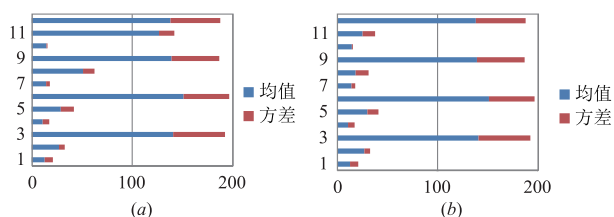


图4 (a) 10200帧各分块区域尺度模型更新得到的均值及标准差
(b) 16410帧各分块区域尺度模型更新得到的均值及标准差

此外,本文对 *ratio* 值进行了实验比较分析,当 *ratio* = 0.3 时即尺度变化步长为 0.3 倍基准尺度时效果最佳.另外,根据文献[14],行人候选框的宽高比 *k* 设置为 0.5.

为了更好地比较分析高斯尺度模型的效果,我们将每个分块区域学习得到的高度大小与真实标注的行人高度进行了对比.如图 5(a)(b)所示分别为第 11 块区域中 6900 帧至 7100 帧图片以及第 8 块区域中 9950 帧至 10200 帧图片的拟合情况.由图可知,行人的真实高度均在基准高度 h_s 附近,且位于上界 h_{max} 和下界 h_{min} 之内,说明 LMP 的尺度模型可以准确地拟合实际目标尺度.此外,这种在行人基准高度上下界内设置候选框的做法保证了出现多个不同尺度行人的情况下的覆盖率,避免盲目增加尺度数量来提高覆盖率,有利于提高行人的检测率.

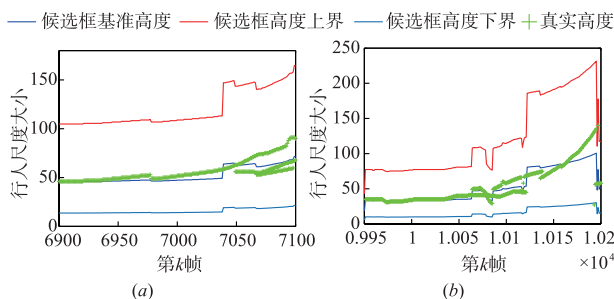


图5 (a)第11块区域6900帧至7100帧行人尺度学习情况
(b)第8块区域9950帧至10200帧行人尺度学习情况

4.3 候选框生成效果测试

4.3.1 LMP 算法候选框生成效果测试

候选框生成算法的目标是生成较少的候选框实现较高的覆盖率,从而保证候选框生成的实时性和准确性.本文提出的 LMP 算法可通过调整步长来改变每帧图片生成候选框的数目.如表 2 所示为设置不同的步长情况下相应的平均每帧图片生成候选框的数目.

如图 6 所示为 LMP 算法在不同数目的候选框下

置不同 IOU 阈值(0.5 ~ 0.7)的实验结果.当设置阈值为 0.5 时,该算法每帧图片生成 5198 个候选框的覆盖率能够达到 90.9%.当设置的 IOU 阈值增大至 0.6 时,每帧图片的平均覆盖率有所下降,为 84.90%.此外,固定 IOU 阈值,增加或是减少生成候选框的个数,图像的平均覆盖率也会有相应的变化.当在每幅图像上平均设置 2607 个框时,可获得 86% 以上的平均覆盖率;而当每幅图像设置候选框数超过 10067 时,平均覆盖率可提升至 92.8%.

表 2 设置不同的步长相应的每帧图片生成候选框的数目

步长 l_1 (像素点)	步长 l_2 (像素点)	步长 l_3 (像素点)	生成候选框 数目/帧(个)
16	12	8	2607
10	8	6	5198
10	6	4	10067

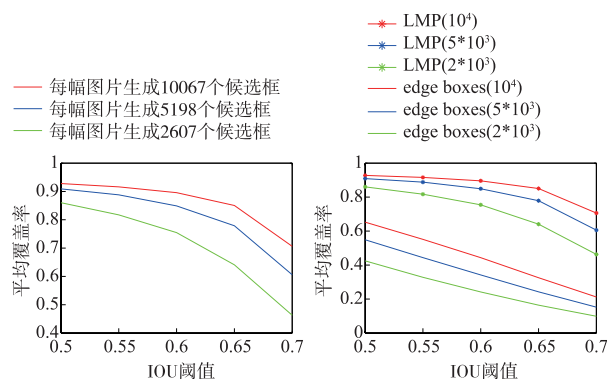


图6 LMP算法生成不同数目的候选框下的测试结果
图7 LMP算法与Edge Boxes算法生成不同数目的候选框下的测试结果

4.3.2 与先进算法的比较

Edge Boxes^[4]算法在 Pascal VOC 2007 数据集上获得了较好的测试效果,为了测试其在行人候选框生成中的效果,本文基于 Caltech set00 行人数据集对 LMP 和 Edge Boxes 算法生成候选框的效果进行了比较.如图 7 所示,当设置 10^4 数量级的行人候选框时,LMP 算法每帧图片的平均覆盖率达 92.78% (IOU 阈值设置为 0.5),Edge Boxes 算法实现的平均覆盖率为 65.26%.在生成其它同等数量级行人候选框时,LMP 在不同 IOU 阈值的情况下所获得的平均覆盖率均优于 Edge Boxes 算法.此外,当每帧图片平均设置 10^4 数量级的行人候选框时,随着 IOU 阈值的增加,LMP 与 Edge Boxes 算法生成候选框的平均覆盖率的差距越来越大,间接说明了 LMP 生成候选框的质量很高.

为了方便比较,滑动搜索方法^[8]具体的参数设置和 LMP 算法一致.如图 8 为三种方法在不同数目的候选框下的测试结果,同样地采用每帧图片的平均覆盖

率来度量算法的效果,其中 IOU 阈值设置为 0.5. 由图可知,相较于滑动搜索窗和 Edge Boxes 两种方法,LMP 算法表现出了极大的优势. 具体地,LMP 在每幅图片平均设置 3000 个左右的候选框就能获得 86% 以上的覆盖率,而 Sliding Window 需要三倍数量的候选框才能达到相同高的覆盖率,对于 Edge Boxes 算法,候选框的数量要求更大.

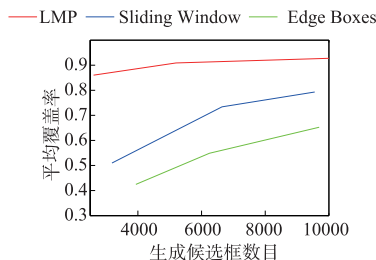


图8 三种方法生成不同数目的候选框下的测试结果

综合以上分析,通过建立区域复合概率模型并进行在线学习,LMP 算法在行人候选框的生成上效果很好,取得了很大的提升.

4.4 LMP 算法生成候选框对行人跟踪辅助效果分析

基于计算机视觉的行人检测及跟踪在车辆辅助驾驶系统中有着重要的应用,是当前智能车辆领域最为活跃的研究课题之一. 其核心是利用安装在车辆上的摄像机检测行人,从而估计出潜在的危险以便采取策略保护行人. 本文的主要工作生成检测候选框在其中具有重要的辅助意义. 由于后续识别模块是对输入的各个候选框(目标)进行分析. 因此,要求候选框生成算法不仅能够准确,而且能够及时地覆盖真实目标.

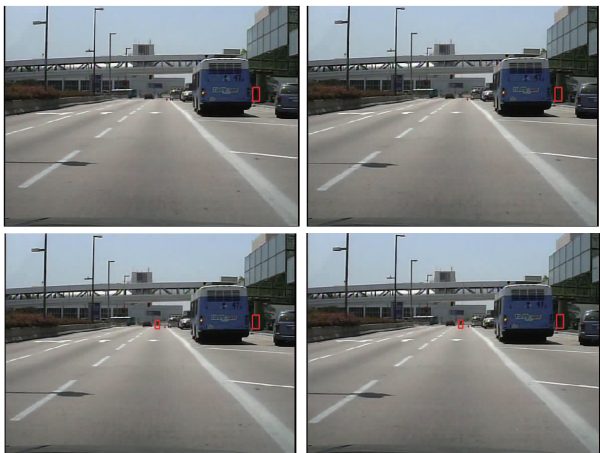


图9 第2956帧出现新的行人

本文在 Caltech 行人数据库中对 LMP 算法的“及时性”也进行了测试. 如图 9 所示为 2956 帧处出现新的行人时的情况,通过高斯泊松复合模型的在线学习生成

检测框,新出现的行人均能够及时被生成的候选框覆盖到. 图 10 展示了行人目标出现后第 k 帧被 LMP 算法生成的候选框覆盖的概率情况,可见,在 5 帧之内覆盖到新出现的行人的概率就能够达到 0.887,随时间增加,覆盖效果将越来越好,在 10 帧内可以达到 0.915,这对于之后的行人跟踪提供了巨大的辅助作用.

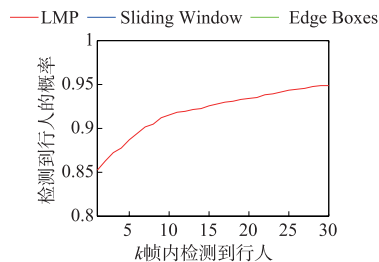


图10 行人目标出现后第 k 帧被 LMP 候选框覆盖的概率

5 结束语

本文提出了一种新的基于区域复合概率模型在线生成行人候选框的方法. 该方法主要是通过对场景区域内行人分布的位置和尺度建立高斯泊松复合模型,从而获取行人密度和尺度信息,为候选框的生成提供依据. 实验结果表明,LMP 算法相较于滑动窗搜索和 Edge Boxes 算法,在每帧图像可设置较少数量的候选框以获得较高的覆盖率,而且在识别实时性上有大幅度改进,为后续的行人识别及跟踪过程提供了较为准确的候选框. 目前,该方法主要针对静态或逐渐变化的场景,而且该方法与后续识别、跟踪算法是基本相独立的,仅要求识别算法反馈识别结果即可. 将来的工作会考虑加入具体的识别与跟踪算法,以便看到 LMP 算法在综合系统应用中的效果.

参考文献

- [1] 杨涛,李静,等. 基于场景模型与统计学习的鲁棒行人检测算法[J]. 自动化学报,2010,36(4):499-508.
YANG Tao, LI Jing, et al. Scene modeling and statistical learning based robust pedestrian detection algorithm [J]. Acta Automatica Sinica, 2010, 36(4): 499-508. (in Chinese)
- [2] HOSANG J, BENENSON R, et al. What makes for effective detection proposals? [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 38(4): 814-830.
- [3] CHENG M M, ZHANG Z, et al. BING: Binarized normed gradients for objectness estimation at 300fps [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2014. 3286-3293.

- [4] ZITNICK C L, DOLLAR P. Edge boxes: Locating object proposals from edges [A]. 13th European Conference on Computer Vision [C]. Zurich, Switzerland, 2014. 391 – 405.
- [5] HORN B K P, SCHUNCK B G. Determining optical flow [J]. Artificial Intelligence, 1981, 17(1 – 3): 185 – 203.
- [6] TERZOPOULOS D. Regularization of inverse visual problems involving discontinuities [J]. IEEE Trans PAMI, 1986, 8(4): 413 – 424.
- [7] LUCAS B, KANADE T. An iterative image registration technique with an application to stereo vision [A]. International Joint Conference on Artificial Intelligence [C]. Canada, 1981. 674 – 679.
- [8] GERONIMO D, LOPEZ A M. Vision-Based Pedestrian Protection Systems for Intelligent Vehicles [M]. New-York: Springer Publishing Company, 2013.
- [9] UJILING S R, SANDE K E, et al. selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104 (2): 154 – 171.
- [10] GIRSHICK R, DONAHUE J, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [A]. 27th IEEE Conference on Computer Vision and Pattern Recognition [C]. US: IEEE, 2014. 580 – 587.
- [11] GIRSHICK R. Fast R-CNN [A]. 15th IEEE International Conference on Computer Vision [C], Santiago, Chile, 2016. 1440 – 1448.
- [12] REN S, HE K, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [A]. 29th Annual Conference on Neural Information Processing Systems [C]. Montreal, QC, Canada, 2015. 91 – 99.
- [13] 覃剑, 王美华. 采用在线高斯模型的行人检测候选框快速生成方法 [J]. 光学学报, 2016, 36(11): 162 – 172.
QIN Jian, WANG Mei-Hua. Fast pedestrian proposal generation algorithm using online Gaussian model [J]. Acta Optica Sinica, 2016, 36(11): 162 – 172. (in Chinese)
- [14] DOLLAR P, WOJEK C, et al. Pedestrian detection: an evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34(4): 743 – 761.

作者简介



覃 剑 男, 博士, 1977 年 5 月生于陕西宝鸡. 重庆大学通信工程学院副教授, 研究方向为视频分析及传输.
E-mail: qinjian@cqu.edu.cn



肖 婷 女, 硕士研究生, 1993 年 12 月生于江西吉安. 研究方向为计算机视觉、目标检测与识别分类.
E-mail: 20151202011@cqu.edu.cn