

改进的基于人头检测的行人跟踪算法

马金鹏

(上海大学通信与信息工程学院 上海 200444)

摘要: 针对已有的基于人头检测的行人跟踪算法在复杂背景下易发生目标跟踪错误的问题,提出一种改进的基于人头检测的行人跟踪算法,以适应多种复杂场景下的行人跟踪。该方法首先采用前景分割技术提取目标行人,在 Adaboost 分类器中加入人体躯干负样本,再结合 Haar-like 特征在前景的基础上检测人头,通过检测到的人头建立目标跟踪链实现对场景中行人的跟踪。实验结果表明,所提算法降低了人头的误检率和漏检率,提高了在多种复杂场景中对行人跟踪的鲁棒性。

关键词: Adaboost 分类器;人头检测;目标跟踪链;行人跟踪

中图分类号: TP751.1 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Improved pedestrian-tracking method based on head detection

Ma Jinpeng

(School of Communication and Information Engineering of Shanghai University, Shanghai 200444, China)

Abstract: Considering the problem of existed pedestrians tracking method based on head detection often made mistakes during tracking in complex backgrounds, this paper proposes an improved pedestrian-tracking method based on head detection to adapt to a variety of complex scenarios. This method firstly uses the foreground segmentation technology to extract the target people, and add human body negative samples in adaboost classifier which combined with the feature of haar-like to detect head, then establish head-tracking chains to fulfill the tracking of pedestrians by using the gotten head. Experiments show that the improved method reduces the error rate of head detection and improve the correction rate of human-tracking in complex backgrounds.

Keywords: Adaboost classifier; head detection; head-tracking chains; tracking of pedestrians

0 引言

对复杂场景中行人的跟踪是智能视频监控系统的重要环节,利用监控设备得到的监控视频对场景中的行人信息进行分析,在智慧交通、安全监控和刑侦破案等方面有着重要的应用价值。目前对行人检测和跟踪算法主要可分为两类,第一类算法采集整个行人作为样本,学习整体特征并对场景中的完整人体进行检测和跟踪^[1-2],这类算法可以提供关于人体特征的丰富信息,通过充分的特征学习能获得较好的检测结果和跟踪结果,但当在密集场景中发生遮挡严重时,破坏了整体特征而造成漏检,此外较大的模板意味着较高的计算复杂度,降低了跟踪算法的实时性;第二类算法则是基于人头检测实现行人跟踪^[3-4],这类算法将非刚性问题转化为刚性问题解决,人头是刚体,且形状不易随人体运动或摄像机视角变化而改变,此外在拥挤场景中,相比人体其他部位,人头被遮挡的概率较小,因此在人群拥挤的场

景内能获得较好的检测结果,但由于人头提供的特征较少,容易产生严重的误检。

通过对上述不同检测与跟踪算法的分析,本文结合以上两种算法的优点,提出一种改进的基于人头检测的行人跟踪算法,首先采用前景分割方法提取目标前景,然后在前景的基础上,采用 Haar-like 特征及 Adaboost 分类器提取目标人头,为提高人头正检率,在分类器训练过程中引入适量人体躯干负样本,然后通过检测到有效人头构建目标外貌模型,创建跟踪链,同时将目标检测结果与跟踪链进行关联,用检测结果修正跟踪链的参数,并不断更新跟踪链,在保证正检率的同时,提高了行人跟踪的准确性。

1 目标前景分割

运动目标分割算法主要包括背景建模、帧间差分、光流法 3 类^[5-6]。针对监控相机通常视角固定的特点,本文采用文献[7]提出的一种基于结构相似性的背景建模方法,具体

流程如下。

首先取视频前 n 帧图像的平均值为初始背景图像。假设背景图像为 Y_1 , 当前图像为 Y_2 , 则计算亮度失真 $S_m(Y_1, Y_2)$ 和对比度失真如下:

$$S_m(Y_1, Y_2) = \frac{2u_1u_2 + c_1}{u_1^2 + u_2^2 + c_1}, S_v(Y_1, Y_2) = \frac{2\sigma_{1,2} + c_2}{\sigma_1^2 + \sigma_2^2 + c_2} \quad (1)$$

式中: u_1 和 u_2 分别是 Y_1 和 Y_2 的区域均值, σ_1 和 σ_2 分别是 Y_1 和 Y_2 的区域方差, $\sigma_{1,2}$ 是 Y_1 和 Y_2 的区域协方差, c_1 和 c_2 为常数。其次计算出当前图像与背景图像间的结构相似图 $SSIM(Y_1, Y_2)$ 和结构相似系数 ssm 如下:

$$SSIM(Y_1, Y_2) = S_m(Y_1, Y_2) \cdot S_v(Y_1, Y_2) \quad (2)$$

$$ssm = (SSIM(Y_1, Y_2))^\gamma \quad (3)$$

式中: γ 为常数。以结构相似系数为基础, 建立并更新背景模型, 计算 t 时刻运动区域反馈系数 d_t 和模糊反馈系数 β_t , 如下:

$$d_t = (1 - \alpha)d_{t-1} + \alpha(1 - ssm_t) \quad (4)$$

$$\beta_t = \frac{ssm_t + d_t}{2} \quad (5)$$

式中: d_{t-1} 为上一时刻运动区域反馈系数, α 表示学习率。基于结构相似性获得的背景图像 B_t 如式(6)所示。

$$B_t = (1 - \beta_t \cdot \alpha)B_{t-1} + \beta_t \cdot \alpha \cdot I_t \quad (6)$$

式中: I_t 为当前帧, B_t 为当前背景图像, B_{t-1} 为上一时刻背景图像。为克服 RGB 色彩空间对光照变化的敏感性, 将当前图像 I_t 和背景图像 B_t 均转换到 YUV 颜色空间(记为 YI_t, UI_t, VI_t 及 YB_t, UB_t, VB_t)进行运动前景区域分割, 分别计算当前帧与背景的亮度、色度及结构差异如下:

$$Y_{diff} = |YI_t - YB_t| \quad (7)$$

$$UV_{diff} = |UI_t - UB_t| + |VI_t - VB_t| \quad (8)$$

$$S_{diff} = ssm_t \quad (9)$$

相应地, 基于亮度的前景分割方法为:

$$Y_{fgt} = \begin{cases} 1, & Y_{diff} > Y_{tht} + s_{th} \\ 0, & \text{其他} \end{cases} \quad (10)$$

其中 Y_{fgt} 为分割结果, Y_{tht} 为亮度分量阈值, s_{th} 为阈值调整参数, 阈值估计方法如下:

$$Y_{tht} = \begin{cases} Y_{tht-1} + \alpha_T(1 + d_t), & Y_{tht} \leq Y_{diff} \\ Y_{tht-1} - \alpha_T(1 - d_t), & \text{其他} \end{cases} \quad (11)$$

其中 α_T 为阈值的更新率, 基于类似的方法获得基于色度的前景掩码 UV_{fgt} , 基于结构变化的分割方法如下:

$$S_{fgt} = \begin{cases} 1, & S_{diff} \leq 0.5S_u \\ 0, & \text{其他} \end{cases} \quad (12)$$

式中: S_u 为结构差异 S_{diff} 的平均值, 对上述 3 种方法获得的分割结果进行融合并生成最终的前景掩码如式(13)所示。

$$F_{gmskt} = S_{fgt} \cap (Y_{fgt} \cup UV_{fgt}) \quad (13)$$

式中: F_{gmskt} 为所获得的前景分割结果, 对 F_{gmskt} 进一步通过中值滤波、边缘检测及孔洞填充等处理以提高准确性。

2 人头检测

在运动目标分割的基础上, 采用 Haar-like 特征及 Adaboost 算法实现人头检测^[8-11]。Haar-like 特征在 Viola and Jones^[13]所提出的 5 种基本特征的基础上进行扩展, 添加了一些中心环绕特征及旋转 45° 的矩形特征, 并归类为边缘特征、中心特征和对角线特征, 如图 1 所示。

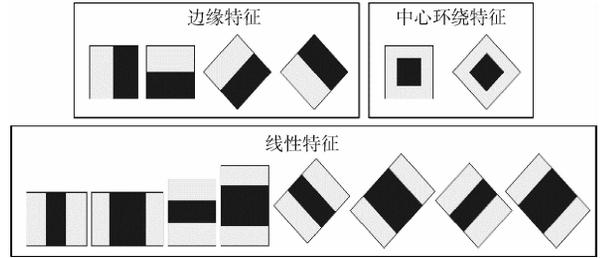


图 1 Haar-like 特征

Haar-like 特征仅规定形状, 其尺寸及在图像中的位置都可以任意改变。计算 Haar-like 特征时, 将每种 Haar-like 特征看作一种模板, 叠加到图像相应位置后计算白色区域像素值减去黑色区域像素值作为特征值。

为合理地选取有效特征, 采集一定数量的正负样本, 采用 Adaboost 算法记录每次迭代过程中分类效果最好的特征作为弱分类器, 通过级联多个弱分类器实现具有较好检测效果的强分类器。为剔除背景的干扰, 首先结合运动目标分割算法从一些视频文件中截取正样本并将背景区域的像素设置为固定值 255, 如图 2 中前 4 行所示, 作为对比, 部分未剔除背景的正样本图像如图 2 第 5 行所示。



图 2 人头正样本示例对比

由图 2 可知, 通过剔除背景区域, 使正样本仅仅包含人头区域的特征, 使得检测器在不同场景中所需检测的目标与背景无关, 避免了相同的人头图像由于更换了背景而造成部分特征失效使检测率降低的问题。

对于人体躯干负样本同样采用运动区域分割方式剔除背景。由于实际应用中只对运动前景区域进行检测, 而背景区域所产生的误检测都将被忽略。因此无需像传统方法

那样采集大量不同场景下数量庞大的负样本图像集,而只需要关注同样属于运动区域的人体衣服、躯干、鞋子等部位可能产生的误检即可,部分采集的人体躯干负样本示例如图3所示。



图3 人体躯干负样本示例

在训练器中通过引入去除背景的躯干负样本以降低对人头的误检率,并将所有人头正样本图像归一化为 16×16 像素,基于Adaboost算法进行监督式学习得到级联检测器。在线检测过程中,首先通过运动目标分割算法将背景区域设置为同一固定值,然后按照分割掩码计算每个前景区域的外接矩形,为确保运动目标的完整性,对外接矩形区域进行适当扩充,并对重叠部分进行合并,在处理后的图像与合并区域进行人头检测。

3 多目标跟踪

3.1 目标外貌模型构建

为获得准确的跟踪结果,需建立目标外貌模型。本文采用文献[5]提出的一种多示例学习方法对外貌进行描述,过程如下:

1) 设定距离 $\alpha, \beta, \gamma (\alpha < \beta < \gamma)$,在当前人头位置周围随机采集若干样本。其中,与当前人头距离小于 α 的样本标记为正样本,距离处于 β 与 γ 之间的为负样本,样本尺寸与当前人头尺寸相同。

2) 生成haar特征池,此处haar特征为图1特征的一种扩展,即按随机的位置、权重、尺度生成 $2 \sim 3$ 个矩形区域作为模板,假设当前人头矩形尺寸为 (row, col) 则每种特征值的计算方法如下:

$$feature(I) = \sum_{j=0}^k w_j \sum_{r=x_1}^{x_2} \sum_{c=y_1}^{y_2} I_r \quad (14)$$

其中矩形数量 $k \in [2, 3]$,权重 $w = \text{rand}(-1, 1)$,坐标 $x_1 = \text{rand}(0, row)$, $x_2 = \text{rand}(x_1, row - x_1)$, y_1, y_2 的定义与 x_1, x_2 相似, $\text{rand}(a, b)$ 为生成 a 与 b 之间随机数字的函数, I 为样本图像。

3) 假设所采集的样本集合为 $X = \{(I_1, C_1), (I_2, C_2), \dots, (I_t, C_t)\}$, t 为样本数量, I 为样本图像, $C \in [0, 1]$ 为样本类别。假设每个特征在正负样本集中均满足高斯分布,则基于式(16)初始化对应弱分类 h_i 的参数 $\{u_{pi}, \delta_{pi}, u_m, \delta_m\}$,其中 $\{u_{pi}, u_m\}$ 分别是正负样本特征值的均值, δ_{pi}, δ_m 则是正负样

本的方差, $i=0, 1, \dots, k$ 为特征总数量。在后续处理过程中,需不断对弱分类器参数进行更新,如下:

$$u_{pi} = (1 - \gamma)u_{pi} + \gamma \frac{1}{n} \sum_{C=1} feature_i(I) \quad (15)$$

$$\delta_{pi} = (1 - \gamma)\delta_{pi} + \gamma \frac{1}{n} \sqrt{\frac{1}{n} \sum_{C=1} (feature_i(I) - u_{pi})^2} \quad (16)$$

式中: γ 为学习率,对 u_m, δ_m 按类似的方式更新。

4) 假设正负样本先验概率相同,计算每种弱分类器对正负样本的后验概率:

$$p_i(C = 1/X) = \sum \left(\log \left(\frac{1}{\delta_{pi}} \exp \left(-\frac{(f_i - u_{pi})^2}{2\delta_{pi}^2} \right) \right) \right) - \log \left(\frac{1}{\delta_m} \exp \left(-\frac{(f_i - u_m)^2}{2\delta_m^2} \right) \right) \quad (17)$$

式中: f_i 为标签 $C=1$ 的样本图像的第 i 种特征值。按类似的方法计算出第 i 个弱分类器对负样本的后验概率 $p_i(C=0/X)$ 。

5) 以似然函数作为度量,采用贪心策略选取前 M 个分类效果最好的弱分类器,级联形成强分类器 H ,如下:

$$h_m = \text{argmax}_l (H_{m-1} + h) \quad (18)$$

$$l = \sum_{C=0,1} y_i \log(p_C^m) + (1 - y_i) \log(1 - p_C^m), y_i \in [0, 1] \quad (19)$$

$$p_C^m = \sigma(H_{m-1} + h_m) \quad (20)$$

6) 采用最终获得的强分类器包含最有效的 M 个弱分类器编号,均值及方差。将这些数据存储于跟踪链中,用于评价待测图像与当前跟踪链的相似度。

3.2 行人跟踪

在前面所获的检测结果的基础上,首先定义第 k 帧的检测结果 D_k 及跟踪链集合 T_k 如下:

$$D_k = \{(x_{k1}, y_{k1}), (x_{k2}, y_{k2}), \dots, (x_{kn}, y_{kn})\} \quad (21)$$

$$T_k = \{t_1, t_2, \dots, t_m\}, t = \{mod_i, f_i, c_i\}, i = 1, 2, \dots, m \quad (22)$$

式中: n 为当前检测结果数量, m 为现存有效跟踪链接数量, mod_i 为外貌分类器, f_i 为第 i 个跟踪链的可信度, c_i 为对应跟踪链的轨迹,则 t 时刻的更新过程如下。首先对所有检测结果进行遍历,在每个检测结果周围一定范围内判断是否存在有效跟踪链,则可能产生以下3种情况。

1) 不存在跟踪链,此时通常为新目标在场景中出现,因此创建一个跟踪链接,初始化其外貌分类器并设置可信度 $f_i=1$;

2) 存在一条跟踪链,此时用当前检测结果修正跟踪链参数,由于检测器是采用大量正负样本进行离线训练的,相比完全依赖在线样本建立的外貌分类器具有更高的可信度,因此将该检测结果设置为跟踪链当前位置,并以较高权重 5γ 更新外貌模型;

3) 存在多条跟踪链,此时用每个跟踪链的外貌分类器对当前检测结果进行测试,将检测结果分配给相似度较高

的跟踪链进行参数更新。

经过上述步骤,一些跟踪链获得了检测结果的支持并进行更新,这些跟踪链的可信度 f_i+1 。另一些跟踪链则没有与之对应的检测结果,这些跟踪链的可信度减1,若可信度仍 >0 ,则采用预测方式进行更新,即在上一时刻人头位置周围随机采样若干候选区域,通过外貌分类器筛选出当前概率最大的候选区域,进行位置及外貌特征的更新。

上述步骤实现了跟踪链的更新过程,若某条跟踪链可信度为0,则删除该条跟踪链。

当出现孤立的检测结果而无跟踪链与之对应时有可能并非新的运动目标出现,而是由于误检造成的。因此上述情况2)中虽然创建了跟踪链,但并非有效的跟踪链。为克服这个问题,当可信度 $f_i > 2$ 时将跟踪链设置为有效跟踪链。有效跟踪链在极少情况下是由误检引起的,具有较高的可信度,因此将这些跟踪链的可信度提高到10。

当运动对象间出现严重互相遮挡的情况时,情况3)可能无法有效应对,即被遮挡的跟踪链错误地与未被遮挡的检测结果关联而错误地更新外貌特征。为克服这个问题,对每次跟踪链的位置更新结果进行遍历,若某条跟踪链的位置与其他跟踪链重叠超过50%时,该跟踪链的人头特征不进行更新。

4 实验结果

为验证本文所提改进的基于人头检测的行人跟踪算法能够适应多种复杂场景,利用单反拍摄了室内、走廊和室外等不同复杂场景的视频进行测试,本文实验分为两部分,一部分是检测实验结果,另一部分为跟踪实验结果。在Windows 7操作系统中,采用Visual Studio 2010结合OpenCV2.4作为研究平台,计算机配置为intel core i5 CPU 2.4GHz,内存4G。

为验证改进后的人头检测算法在人头检测方面的优势,采用基于正负样本的Adaboost训练器结合Haar-like特征实现对人头的检测,检测的实验结果如图4所示。

图4(a)中的视频拍摄于室内,场景中多个运动目标随意走动,图4(b)中的视频拍摄于走廊,多个运动目标由远及近通过走廊,图4(c)中的视频拍摄于室外,场景中穿着各异的行人通过监控区域。图中矩形框内为场景中人头的检测结果。由图中检测结果可以发现,所提检测算法能够准确检测场景中的多个目标人头,对于多个不同复杂场景具有很好的通用性。



(a) 室内人头检测结果



(b) 走廊人头检测结果



(c) 室外人头检测结果

图4 多种复杂场景下的人头检测结果

为验证本文剔除背景后的改进的人头检测算法的有效性,将其与文献[12]进行对比,两者采用的正负样本相同,但文献[12]未剔除样本中的背景,以检测准确率作为衡量标准,统计的对比实验结果如表1所示。

表1 检测准确率对比

视频	检测算法	检测准确率/%
室内	文献[12]	80.31
	本文	85.33
走廊	文献[12]	85.56
	本文	91.61
室外	文献[12]	73.41
	本文	81.92

通过表1的实验对比结果可以发现,本文所提检测算法在各组测试视频当中的检测准确率均高于背景相关的检测算法,这是因为本文在检测过程中将背景设置为白色,降低了背景区域的误检率,此外在训练过程中通过加入适量人体躯干部分的负样本,使检测器降低了对躯干部分的误检率。

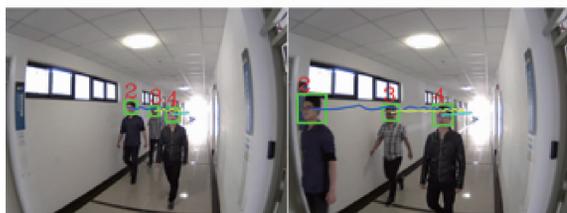
在检测结果的基础上,为进一步验证对行人跟踪的有效性,通过建立跟踪链实现对行人的跟踪。同样在3个相同的视频场景下进行跟踪实验,跟踪的结果如图5所示。

通过图5的跟踪结果可以看出,改进的基于人头检测的行人跟踪算法在不同光照环境、不同摄像机参数及不同背景的监控场合下都能有效地对场景中的行人进行跟踪,且能适应人头尺寸有较大变化的情况,在行人较为密集或存在部分遮挡的情况下能有效对检测结果进行关联而不易产生身份交叉错误。

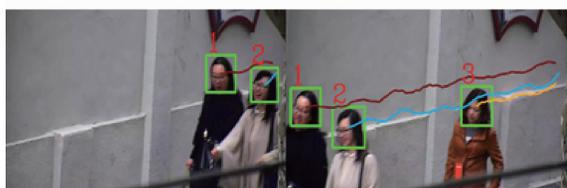
为了验证本文提出算法在不同复杂场景中具有明显的跟踪优势,与文献[6]所提的融合颜色梯度直方图的粒子滤



(a) 室内行人跟踪结果



(b) 走廊行人跟踪结果



(c) 室外行人跟踪结果

图5 多种复杂场景下的行人跟踪结果

波算法进行对比,并对二者跟踪准确率数据进行统计,对比统计结果如表2所示。

表2 跟踪准确率对比

视频	跟踪算法	跟踪准确率/%
室内	文献[6]	84.62
	本文	89.45
走廊	文献[6]	90.51
	本文	93.27
室外	文献[6]	80.32
	本文	85.28

由表2可知,通过剔除背景的影响及通过可信度确认有效跟踪链的方法,大大降低了跟踪错误,使所提算法的正确跟踪结果高于文献[6]对应算法。所采用的基于正负样本后验概率的外貌模型相比文献[6]仅采用正样本颜色和梯度直方图的外貌模型特征信息更丰富,因此身份交换的次数较少,能稳定实现多个复杂场景中行人的准确跟踪。

5 结论

本文在目标前景分割的基础上,利用人头分割技术将非刚性问题转化为刚性问题解决,并在训练器中引入一定量的人体躯干负样本,将每个检测到的人头赋予标签,通过创建跟踪链将同一个人的检测结果串联起来,同时在行人

跟踪的过程中将检测结果与跟踪结果关联起来,用检测结果不断修正跟踪结果,进而实现复杂场景中行人的跟踪。实验结果表明,改进后的算法对场景中行人的离开与进入,行人间的相互遮挡具有较强的鲁棒性。

参考文献

- [1] ZHAO M, SUN D H, TANG Y, et al. Head detection based on 21HT and circle existence model [C]. Proceeding of the 10th World Congress on Intelligent Control and Automation, 2012: 4875-4880.
- [2] 黄凯奇,陈晓棠.智能视频监控技术综述[J].计算机学报,2015(6):1093-1118.
- [3] XIE D, DANG L, TONG R. Video based head detection and tracking surveillance system [C]. Proceedings of IEEE International Conference on Fuzzy System and Knowledge Discovery, 2012: 2832-2836.
- [4] XU R Y, GUAN Y P, HUANG Y Z. Multiple human detection and tracking based on head detection for real-time video surveillance[J]. Multimedia Tools and Applications, 2015,74(3):729-742.
- [5] 胡威.一种新的 Meanshift 快速运动目标鲁棒跟踪方法[J].计算机应用与软件,2014,31(1):140-143.
- [6] 纪青华,禹素萍.基于 Surendra 背景减除法和四帧差分法的目标检测算法[J].计算机应用与软件,2014,31(12):242-244.
- [7] LUO Y, GUAN Y P. Motion objects segmentation based on structural similarity background modeling[J]. IET Computer Vision, 2015, 9(4): 476-488.
- [8] 梁硕,陈金勇,吴金亮.基于分类的视频目标跟踪算法研究[J].无线电工程,2017,47(1):19-22,66.
- [9] HENRIQUES J F, CASEIRO R P, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015,37(3):583-596.
- [10] 刘勇,王晓军,雷玉香.基于视频的施工升降机乘客头部识别方法研究[J].电子测试,2016(11):105-106.
- [11] 张姗姗,景文博,刘学,等.一种基于深度信息的人头检测方法[J].长春理工大学学报(自然科学版),2016(2):107-111,115.
- [12] ALI I, DAILEY M N. Multiple human tracking in high-density crowds [J]. Image and Vision Computing, 2012, 30(12):966-977.
- [13] 曾巧玲,文贡坚.运动目标跟踪综述[J].重庆理工大学学报(自然科学版),2016(7):103-111.

作者简介

马金鹏,1992年出生,硕士研究生,主要研究方向数字图像处理。

E-mail:496096158@qq.com