

# 基于热点区域定义的人数统计方法研究

高 飞 丰敏强 汪敏倩 卢书芳 肖 刚

(浙江工业大学计算机科学与技术学院 杭州 310023)

**摘 要** 行人统计在智能监控领域具有重要意义,但复杂背景环境以及行人运动过程中出现的遮挡现象导致当前方法的准确率并不高。此外,传统过线统计人数的方式的实际适用范围有限。考虑到现有方法的不足,提出了一种基于热点区域定义的人数统计方法。首先,利用自适应学习率背景建模提取运动目标前景,得到前景区域的位置和大小,扫描计算运动目标前景范围内的 HOG 特征,并判别是否存在头肩目标;然后,利用基于 KCF 的目标匹配算法跟踪头肩目标;最后,结合目标运动轨迹与提出的区域人数统计算法进行行人人数统计。采用 24fps 的手机拍摄的长度为 10min、分辨率为  $960 \times 720$  像素的视频做人统计实验。实验结果表明,所提算法在统计人数时正确率可达 93.1%,能满足实时性要求。该方法结合了检测效率和准确率,在背景环境复杂的场景下具有良好的效果,能适应各类人数统计的实际应用场景。

**关键词** HOG 特征,自适应学习率背景建模,目标转移矩阵,热点区域,KCF,人数统计  
中图分类号 TP391.41 文献标识码 A

## Research on People Counting Based on Hot Area

GAO Fei FENG Min-qiang WANG Min-qian LU Shu-fang XIAO Gang

(College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China)

**Abstract** People counting in the field of intelligent monitoring is important, but because of complex background environment and pedestrian movement occlusion phenomenon resulting in the current method accuracy is low, in addition to the traditional line statistics on the number of practical limited scope, taking into account the present that lacking effective methods, we proposed a method of people counting based on hot area. Firstly, the adaptive learning-rate background model is used to extract the foreground of the moving target, and the position and size of the foreground region are obtained. The HOG feature in the foreground region of the moving target is scanned, and the head and shoulder target is determined. Then the target matching Matrix algorithm based on KCF is used to track the head-shoulder target. Finally, the number of pedestrians is calculated by combining the target trajectory and people counting method based on hot area. The number of the video is  $960 \times 720$  pixels with a resolution of  $960 \times 720$  pixels. The correctness of the algorithm reached 93.1%, and it can meet the real-time requirement. The method we proposed combines the detection efficiency and accuracy, and has good effect in scenes with complex background environment, which can meet various practical application scenarios of people counting.

**Keywords** HOG, Adaptive learning-rate background model, Target transformation matrix, Hot area, KCF, People counting

在当今社会,智能视频监控在各个领域扮演着越来越重要的角色,其中的热点之一是基于图像识别的行人统计,而行人运动中的遮挡、轮廓非刚性、光照以及复杂的背景环境等因素给正确的行人统计带来了一定的困难。

通常而言,行人统计算法融合了运动物体检测、行人检测与分割、形状分析、特征提取和目标跟踪等多项技术。根据手段的不同,可将其分为两大类:1)间接法(或称基于映射或者度量的统计方法),其核心思想是通过建立场景特征与行人数量的函数关系来测算人数。Hashemzadeh M 等<sup>[1]</sup>提出基于关键点与分割所得前景的大小、密度以及人群的遮挡程度所提取的特征估计人数;常庆龙等<sup>[2]</sup>提出通过归一化前景和角

点信息计算遮挡因子,最后通过后向传播网络完成人数统计;比较经典的还有 Cong Y 等提出的 Flow Mosaicking<sup>[3]</sup>以及 Chen C L 等<sup>[4]</sup>提出的 Crowd counting and profiling 等。上述方法主要应用于行人高度密集的场景中,通过提取图像中的特征并根据估计函数统计行人个数。2)直接法(或称基于检测的方法),核心思想是首先根据特征检测出目标,然后对目标进行跟踪、计数。该方法具有精度高、灵活计数、适应场景广泛等特点,在实际应用中往往被采用。

直接法一般存在两个步骤:行人的检测和目标的跟踪。比较经典的行人检测方法有 Dalal 等<sup>[5]</sup>提出的基于行人身体轮廓的方向梯度直方图(HOG)特征的行人检测方法,目前比

本文受国家自然科学基金资助项目(C12412135,61402410),浙江省自然科学基金资助项目(LY13F020029,LQ14F020004)资助。

高 飞(1974—),男,教授,博士生导师,主要研究方向为计算机视觉与图像处理、数字化设计与制造等;丰敏强(1991—),男,硕士生,主要研究方向为计算机视觉与图像处理、模式识别;汪敏倩(1993—),女,硕士生,主要研究方向为计算机视觉与图像处理;卢书芳(1984—),女,博士,讲师,主要研究方向为三维造型、计算机艺术仿真等;肖 刚(1965—),男,教授,博士生导师,主要研究方向为数字化设计与制造等。

较主流的行人检测都由 HOG 特征演变而来; Felzenszwalb<sup>[6]</sup>等提出的形变部件模型(DPM), 基于人体自身的铰链式结构, “DPM”改进了行人的 HOG 特征, 提高了检测的速度和效果; 基于卷积神经网络(CNN)的检测算法是近年来的热点, 何凯明等<sup>[7]</sup>改进了 Faster R-CNN 的目标检测框架, 使用 ResNet 替换了 VGG16 网络, 并取得了更好的检测效果; 比较典型的还有 Joseph Redmon 等<sup>[8]</sup>提出的一种全新的 CNN 检测框架, 该框架大大提高了检测速度, 可以做到每秒检测 45 帧图像, 但是该框架需要依赖强大的 GPU 等硬件条件。目前的检测方法虽然有着不错的检测效果, 但是并不能很好地适用于实际应用场景。目标的检测时间效率以及检测算法依赖于强大的硬件是未来要解决的问题。

传统的采用直接法统计行人的方法一般采用过线方式统计。唐春晖<sup>[9]</sup>利用垂直单目摄像头拍摄行人头部信息, 并将其输入到学习好的 AdaBoost 人头分类器中来检测目标, 最后采用过线跟踪的方式完成计数。在室内光线充足的环境下, 该算法能完成对统计线附近成功检测到的人头的计数, 而对未能检测到的人头却无法完成计数。文嘉俊等<sup>[10]</sup>、徐超等<sup>[11]</sup>提出了一种基于行人头部的 HOG+SVM 的行人目标检测方法, 但其采用人头的 HOG 特征作为检测目标特征, 容易产生误检, 尤其是行人携带类似圆形物品或者背景出现圆形、球形物体时。其中徐超<sup>[11]</sup>等使用 Camshift 对目标进行跟踪, 该跟踪算法往往用于局部最优跟踪, 对于速度快的物体容易跟丢。周治平等<sup>[12]</sup>采用特征回归和检测结合的方式, 对视频中远景和近景采用不同的人头统计方法, 提高了对行人目标检测的正确率, 但是该文利用检测作为跟踪, 由于行人遮挡等因素的干扰, 并不能得到较为精确的统计结果。Liu J 等<sup>[13]</sup>提出了一种在行人检测的算法上增加基于 RGB 通道的粒子滤波跟踪算法, 该方法在前景与背景颜色区分明显的场景中有着不错的统计效果, 但是一旦行人衣服颜色与背景相近或者目标增多时, 便无法满足多目标跟踪效果的要求。郑翔翔等<sup>[14]</sup>把行人头部分为头顶、侧面、正面和后脑勺 4 个部分, 并训练对应的子分类器, 以提高检测的精确度, 然后再根据图像的直方图信息对目标进行跟踪, 统计行人目标的个数。但是由于存在多个分类器, 需要多个子检测过程, 其在时间效率上难以满足实时性; 此外, 行人头部特征容易受到周围环境的干扰, 如行人穿着深色服装容易引起误检, 或者发色较浅的行人容易漏检等。顾德军等<sup>[15]</sup>根据人头的轮廓特征和颜色特性进行检测, 并利用运动目标的连续性对其进行跟踪, 从而实现人数的统计。该方法实现简单, 对单个/多个不存在遮挡的目标有着不错的效果, 但一旦行人间发生遮挡, 就会出现目标漏检和跟踪丢失情况, 从而影响人数统计的精确度。夏菁菁等<sup>[16]</sup>提出了一种基于人体骨架特征的人数统计算法, 即利用形态学提取人体骨架图, 并得到骨架特征, 通过分析骨架的人头区域特征计算行人人头个数, 实现人数统计。该方法对行人运动过程中出现遮挡的情况有着不错的检测效果, 但是在背景颜色与行人衣着颜色相似的情况下, 行人的骨架特征并不能很好地检测出来; 此外, 该方法对处于静止状态的行人并不适用, 将其应用在实际场景中存在一定的局限性。

针对现有人数统计方法存在算法时间效率低、行人目标检测算法准确率低以及应用场景比较单一等问题, 本文从实际应用出发, 首先检测运动目标前景的范围, 并在该范围内检测头肩目标, 缩短了检测的时间并降低了误检率; 然后根据运

动行人存在连续性的特点进行跟踪, 如果目标未能成功检测到, 则使用 Henriques J F 等<sup>[17]</sup>提出的 KCF 目标跟踪算法对目标位置进行更新, 并得到目标的运动轨迹; 最后利用本文提出的针对热点区域内的人数统计算法完成对行人人数的统计。人数统计系统流程示意图如图 1 所示。

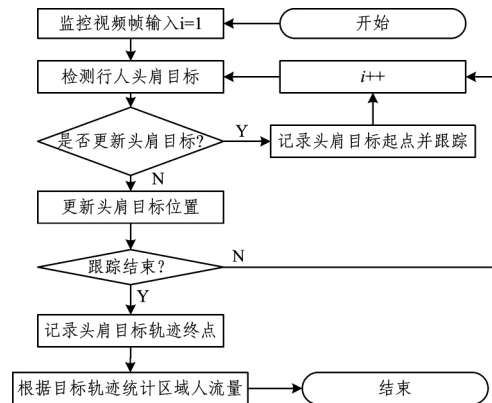


图 1 人数统计总流程示意图

## 1 头肩目标的检测

由于在检测行人的运动过程中可能受到光照变化、姿态改变以及遮挡等影响, 传统的基于全局的行人检测方法并不能满足实际应用的要求; 而目前基于部件/图像块的行人检测方法, 检测模型较为复杂, 虽然在复杂场景下有较好的检测效果, 但是检测速度难以提高。

针对现有目标检测方法的不足, 本文提出了行人头肩目标检测算法融合拍摄场景的空间信息和行人头肩的轮廓信息与颜色信息。首先利用空间信息对行人的运动前景进行分析, 计算行人肩部部位的粗略位置坐标, 然后利用行人头肩轮廓信息与头部的颜色信息完成行人头肩目标的精确定位。

### 1.1 头肩目标前景的提取

常用的背景提取方法有帧差法、光流法、混合高斯模型背景建模法等, 这些方法都有被成功应用于常用的背景建模方法中, 且具有一定的适应性。由于室内环境光线对灯光敏感, 当电压不稳定产生了光线强度变化时, 视频画面容易产生大量离散的噪点, 而行人运动产生的画面像素值变化往往集中在行人运动的区域附近。本文针对室内环境, 提出了一种根据前景像素空间信息改进的自适应  $\beta$  值背景建模方法, 以完成对行人头肩目标前景的提取。

为了得到更好的背景模型, 对整个视频序列  $F$  中的每一帧做中值滤波处理, 用  $F_i$  表示视频序列中的第  $i$  帧图像,  $i=1, 2, \dots, n, n$  为总帧数。此外, 建立  $height \times width$  的前景标记矩阵  $S_i$ , 其中  $height, width$  分别表示单帧图像的高度和宽度(单位: 像素), 并对视频帧做迭代计算, 步骤如下:

(1) 采用加权平均法对彩色帧  $F_i$  灰度化, 得到灰度图  $G_i$ 。

(2) 当  $i=1$  时, 利用灰度图  $G_i$  初始化背景帧, 使得  $B_i(x, y) = G_i(x, y)$ , 并初始化前景标记矩阵  $S_1$ , 即

$$S_1 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}_{height \times width}$$

其中,  $B_i$  表示到第  $i$  帧为止且包括第  $i$  帧所得到的背景矩阵,  $S_i$  表示前景标记矩阵(即各像素为前景的位置标记, 如果该

像素为前景,则标记为 1,否则记为 0),初始第一帧默认为背景,不存在前景。 $B_i(x, y)$ 表示到第  $i$  帧为止且包括第  $i$  帧的背景在  $(x, y)$  处的像素灰度值,  $G_i(x, y)$  表示对应像素坐标  $(x, y)$  处的像素灰度值,  $S_i(x, y)$  表示对应坐标  $(x, y)$  处的前景标记值。

(3) 当  $i > 1$  时,按式(1)~式(6)依次更新视频背景  $B_i$  和前景标记矩阵  $S_i$ 。

$$B_i(x, y) = \beta \times B_{i-1}(x, y) + (1 - \beta) \times G_i(x, y) \quad (1)$$

$$\Delta = |G_i(x, y) - B_i(x, y)| \quad (2)$$

$$S_i(x, y) = \begin{cases} 0, & \Delta < diff \\ 1, & \Delta \geq diff \end{cases} \quad (3)$$

$$hist_i^x = h(\sum_1^y S_i(x, y), Bin) \quad (4)$$

$$hist_i^y = h(\sum_1^x S_i(x, y), Bin) \quad (5)$$

$$\beta = \begin{cases} \beta_1, & \ell(hist_i^x) < u_x \ \& \ \ell(hist_i^y) < u_y \\ \beta_2, & \text{else} \end{cases} \quad (6)$$

其中,  $\beta$  为更新背景的权值系数,  $\Delta$  为中间变量,  $diff$  为像素间的差距控制系数,  $hist_i^x$  与  $hist_i^y$  分别表示第  $i$  帧前景标记矩阵  $S_i(x, y)$  像素点  $x$  列的所有行和像素点  $y$  行所有列的累加值,  $\ell(hist_i^x)$  和  $\ell(hist_i^y)$  分别表示直方图  $hist_i^x$  和  $hist_i^y$  非零值的平均值,  $u_x$  和  $u_y$  表示平均值阈值。式(6)中背景更新  $\beta$  选取不同的值,其中  $\beta_1 > \beta_2$ , 当  $hist_i^x > u_x \ \& \ hist_i^y > u_y$  时,说明前景区域的空间分布比较集中,前景区域为行人的可能性较大,这时选取较大的  $\beta_1$  作为更新背景的权值,以保证行人目标前景的完整和准确;否则说明前景区域的空间分布比较分散,可能是外界干扰(光线变化、摄像头抖动等)引起的,因此选取较小的  $\beta_2$  作为更新背景权值,以加快背景更新的速度。

(4) 根据式(1)~式(3)计算得到的  $S_i$ , 通过以下公式对前景标记矩阵进行聚类并得到  $m$  个子集  $H_i^t$ :

$$\bigcup_{t=1}^m H_i^t = S_i \ \& \ H_i^t \subseteq S_i \quad (7)$$

$$\sum_{x, y} H_i^t(x, y) \geq P \quad (8)$$

$$H_i^{t1} \cap H_i^{t2} = \emptyset; t1, t2 = 1, 2, \dots, m; t1 \neq t2 \quad (9)$$

其中,  $P$  表示前景  $H_i^t$  的面积阈值,当满足式(8)、式(9)条件时,  $H_i^t$  为行人目标前景。由于摄像头在行人目标的上方,根据先验知识得目标前景的上部分为行人的头肩部分。计算目标头肩前景的 4 条边界:

$$left = \min(f_x(H_i^t)) \quad (10)$$

$$right = \max(f_x(H_i^t)) \quad (11)$$

$$top = \min(f_y(H_i^t)) \quad (12)$$

$$bott = top + |right - left| + w \quad (13)$$

其中,  $H_i^t$  表示第  $i$  帧第  $t$  个目标前景;  $left, right, top, bott$  分别表示  $H_i^t$  的左、右、上、下边界,将该区域记为  $D_i^t$ ,  $i$  表示第  $i$  帧,  $t$  表示第  $t$  个目标。在实际应用中,为了更好地检测到目标,  $D_i^t$  的参考范围取自自身的 1.2~1.5 倍。

## 1.2 目标精确定位

利用 1.1 节得到的头肩目标前景,计算行人头肩目标的粗略位置和大小,得到一个目标预测框  $D_i^t$ 。在目标预测框中采用滑动窗口的方式提取每一个滑动窗口内的 HOG 特征,然后根据事先训练好的 SVM 线性核函数精确定位头肩目标位置。具体步骤如下:

(1) 在目标预测框  $D_i^t$  内的原视频帧上使用多尺度的滑动扫描窗口(尺度大小参考范围为  $D_i^t$  的 0.15~1.75 倍,以

0.1 为单位刻度,能保证头肩目标存在于滑动窗口中),滑动窗口以步长为  $w$ (单位:像素)从左往右、从上至下滑动;依次计算滑动窗口中的 HOG 向量,并利用训练好的 SVM 线性分类器判断,如果存在目标,则记录该滑动检测框的大小和位置,继续下一个扫描窗口,否则继续扫描下一个窗口,直到目标预测框  $D_i^t$  的右下角结束。示意图如图 2 所示。

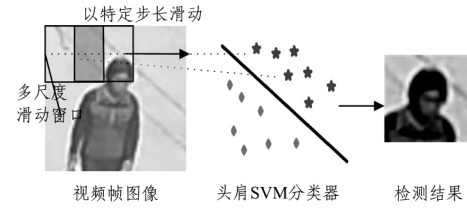


图 2 滑动检测示意图

(2) 如果对同一帧图像内采用滑动窗口式的搜索检测,相同的目标可能会返回多个目标框。对于检测同一个目标返回的多个目标框,利用由步骤(1)得到的所有检测目标结果的中心平均值作为头肩目标的中心位置。记所有的检测结果为  $Tar_k^l$ ,  $k$  表示目标的  $id$  序号,  $id = 1, 2, \dots, n$ ;  $l$  表示同一个目标返回的检测结果数,  $l = 1, 2, \dots, m$ 。目标融合满足如下条件:

$$\frac{Tar_k^{l1} \cap Tar_k^{l2}}{\min(Tar_k^{l1}, Tar_k^{l2})} \geq \theta \quad (14)$$

其中,  $\cap$  表示目标结果  $Tar_k^{l1}$  和目标结果  $Tar_k^{l2}$  的空间面积的交集(单位:像素),  $\min$  表示目标结果  $Tar_k^{l1}$  和目标结果  $Tar_k^{l2}$  中较小的空间面积(单位:像素),  $\theta$  用于评价交集的重合程度,  $\theta$  的参考取值范围为 0.8~0.9。将满足式(14)的目标结果记为  $\hat{T}_k^l$ , 融合目标结果得到的精确的头肩目标位置和大小为:

$$T_k = \frac{1}{m} \sum_{l=1}^m \Gamma(\hat{T}_k^l) \quad (15)$$

其中,  $T_k$  表示检测目标返回结果(待融合目标)融合后精确的头肩目标大小和位置,  $\Gamma$  表示求解各个待融合目标的中心点位置和大小。每帧的目标预测框中的返回检测结果经过目标融合算法的处理后,精确定位头肩目标的大小和位置,用以对行人的精确统计。图 3(a)为未经过目标融合算法处理的某一图像帧,滑动窗口检测返回多个目标框;图 3(b)为经过该算法处理后的图像,融合同一个目标检测得到的多个目标框,进而得到目标的最佳位置和大小。



(a) 目标融合前

(b) 目标融合后

图 3 目标融合效果图

## 2 人数统计算法

### 2.1 KCF 跟踪算法

目前在实时跟踪领域,比较流行的有 CT<sup>[18]</sup>, TLD<sup>[19]</sup>, Struck<sup>[20]</sup>等,但是能兼顾速度和效率的并不多,多数算法能刚好达到实时,而且还是在分辨率不高的视频情况下。KCF 本

质是一种利用循环偏移矩阵构建分类器训练样本的检测跟踪方法,利用循环偏移在目标周围构建分类器的样本,使得图像变成一个循环矩阵。鉴于循环矩阵在傅里叶空间存在可对角化的性质,把矩阵的运算转化成了向量的点乘,从而避免了矩阵求逆的过程,使得算法的复杂度降低了好几个数量级。具体有如下几个步骤:

(1)假设训练的样本集为  $(x_i, y_i)$ , 其线性分类器设为  $f(x) = \langle w, x \rangle + b$ , 则可以通过最小二乘法求解:

$$\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2 \quad (16)$$

其中,  $\lambda$  控制系统的复杂度, 目标是在整个损失函数最小情况下求解  $w$  的值, 令导数为 0, 可以得到  $w = (X^T X + \lambda I)^{-1} X^T y$ 。由于后面是在傅里叶域内计算, 因此把结果统一写成复数域中的形式:

$$w = (X^H X + \lambda)^{-1} X^H y \quad (17)$$

其中,  $X^H$  表示复共轭转置矩阵。对于循环矩阵, 有如下性质:

$$X = F^H \text{diag}(\hat{x}) F \quad (18)$$

其中,  $\hat{x}$  表示  $x$  的傅里叶变换,  $F$  是离散傅里叶变换矩阵, 即满足  $\hat{x} = Fx$ 。把式(18)代入到式(17)中得:

$$\hat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x}^* \odot \hat{x} + \lambda} \quad (19)$$

其中,  $\hat{w}$  为  $w$  的傅里叶变换域,  $\hat{x}^*$  是  $\hat{x}$  的共轭,  $\hat{y}$  为  $y$  的傅里叶变换域,  $\odot$  代表向量对应元素相乘。当样本特征非线性可分时, 需要引入核函数, 把低维空间的计算映射到高维的核空间, 这样在低维空间不可分的情况到高维空间之后就变得可分了。假设  $\phi(x)$  为映射函数, 令  $w = \sum_i \alpha_i \phi(x_i)$ , 把核函数  $K_{ij} = k(x_i, x_j)$  代入后, 最终的回归函数变成:

$$f(z) = \sum_i \alpha_i k(x_i, z) \quad (20)$$

其中,  $\phi(X) = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)]^T$ ;  $f(z)$  是一个向量, 维度和训练的  $y$  一样。

(2)令待检测样本集  $z_j = P^j z$ ,  $P$  为构建循环样本初等矩阵, 那么就可以根据如下公式判断目标移动的位置:

$$f(z_j) = a^T \phi(X) \phi(z_j) \quad (21)$$

本文采用的是高斯核函数, 记为  $K$ 。将高斯核函数代入式(21), 并变换到傅里叶域后得:

$$\hat{f} = (\hat{K}^{xz})^* \hat{a} \quad (22)$$

其中,  $\hat{f}$  为  $z$  样本的响应。

(3)在快速检测得到新目标区域  $t'$  后, 对上一帧的目标模型进行更新, 用以构造  $t'$  的循环位移训练样本:

$$x = (1 - \eta)x_{pre} + \eta x_t \quad (23)$$

其中,  $\eta$  表示模型的学习速率,  $x_{pre}$  由前一帧目标图像  $t$  训练分类器得到,  $x_t$  由当前目标区域图像  $t'$  训练分类器得到。

### 2.2 基于 KCF 的目标路径算法

根据行人移动速度比较缓慢的特点, 同一个目标在相邻两帧图像内有很大部分的重合区域, 相邻两帧中同一目标的像素面积也是渐变的。当目标在相邻两帧检测匹配时, 则为同一个目标, 其匹配条件为:

$$\frac{f(R_{t-1} \ R_t)}{R_t} \geq T \quad (24)$$

其中,  $R_t$  表示  $t$  时刻的某一目标矩形框, 操作符  $\cdot$  表示矩形求

交集运算,  $T$  为控制系数, 本文取 0.8。

为此, 建立跟踪目标匹配矩阵  $M$ , 矩阵行号  $i$  表示当前帧内移动目标的编号,  $i=1, 2, 3, \dots, m$ ; 矩阵列号  $j$  表示上一帧内移动目标的编号,  $j=1, 2, 3, \dots, n$ 。每一次终止搜索后, 没有被匹配的行是当前帧新出现的行人目标; 没有被匹配的列是在当前图像中消失的行人目标。将连续 3 帧被检测匹配的目标确定为行人目标, 并赋予其行人  $id$  号, 同时记录此时目标的中心位置, 为  $a_i$ ,  $i$  为目标的  $id$  号; 当目标走出感兴趣区域时, 对目标停止跟踪, 记录最后的位置  $b_i$ ,  $i$  为目标的  $id$  号, 并从跟踪队列中删除, 行人计数器  $Count+1$ 。算法流程示意图如图 4 所示。根据图 4, 在接下来每一帧中记录目标的位置并加入目标路径队列中, 直到结束跟踪。

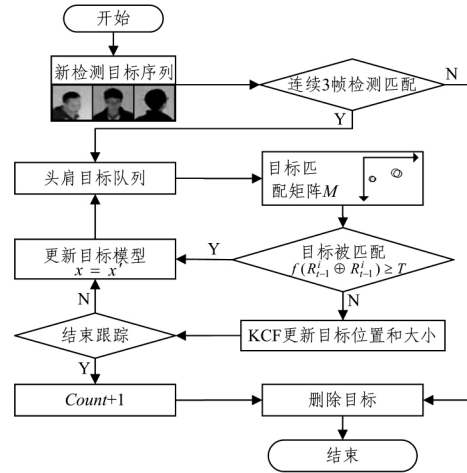


图 4 目标路径算法流程图

多目标行人跟踪普遍存在遮挡问题。当发生遮挡时, 单靠头肩检测算法不能满足每帧都能检测到头肩目标, 因此本文在目标匹配的基础上加入了 KCF 跟踪算法, 提高了目标路径算法的准确性。虽然 KCF 有计算速度快、精确度高的优点, 但是当目标模型外观发生变化(侧身, 低头)时, 由于 KCF 依赖循环矩阵, 初始化矩阵不能适应改变, 所检测到的下一帧预选定框误差可能会与真实目标位置误差很大, 而加入目标检测匹配算法能克服 KCF 的缺点。

### 2.3 热点区域统计算法

针对室内环境中行人可能从 4 个不同方向进出统计的热点区域(本文采用矩形作为热点区域, 行人会从上、下、左、右 4 个方向进入热点区)以及行人在热点区域内停留的时间不确定等问题, 本文在热点区域 4 条边界点划线线段  $a, b, c, d$  周围设定了检测带, 如图 5 所示, 中间虚线内为非检测区域, 对进入该区域的目标不做检测和跟踪, 外黑色实线与内黑色实线间为检测区。以目标从进入检测区域到离开检测区域路径的起点  $a_i$  与终点  $b_i$  构成向量  $P_i$ ,  $i$  为检测到目标的  $id$  号, 根据向量  $P_i$  与边界  $a, b, c, d$  的法向量  $\vec{a}, \vec{b}, \vec{c}, \vec{d}$  (正方向朝外) 的夹角余弦判断:

$$Count = \begin{cases} Count+1, & \cos(P_i, \vec{\mu}) > 0 \\ Count-1, & \cos(P_i, \vec{\mu}) < 0 \end{cases} \quad (25)$$

其中,  $Count$  为行人目标的统计计数器,  $\vec{\mu}$  表示法向量  $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ 。图 5 中圆形目标由统计区域外界穿过边界  $a$ , 当圆形目标接触到非检测区域边界虚线时, 计算该目标的路径和向量  $P_i$ , 代入式(25)并得出行人计数器  $Count+1$ ; 同理, 三角形和正方形的目标也经同样步骤计算, 并更新  $Count$  的值。由于

统计区域的范围和位置是自定义的,检测区域的大小宽度参考范围为统计区域边界内、外两侧大约能容纳 2 个目标的宽度。

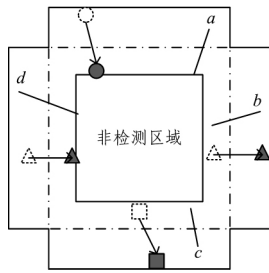


图 5 区域行人统计示意图

### 3 实验结果与分析

本实验的运行环境: Visual Studio 2015, C# + emguCV。实验选取的拍摄器材为 iphone5s, 视频帧分辨率为  $960 \times 720$  像素, 每秒帧数 24fps。

#### 3.1 动态阈值更新背景

图 6(a) 为背景为稳定前的二值图, 图 6(b) 为背景稳定后的二值图。为了更加贴近实际应用场景, 图 6(b) 选取了存在行人的图像。

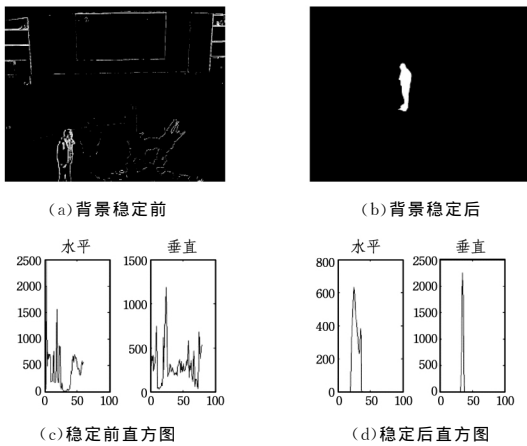


图 6 背景稳定前、后的 2 帧图像

图 6(c)、图 6(d) 为背景稳定前和背景稳定后行、列的前景像素投影直方图(横向、竖向相加求每行、每列的值)。为了更加直观地反映两者的区别, 本文以 12 个像素单位对直方图分区量化。由图 6(c) 可以明显看出, 背景稳定前直方图区间分布广, 而且每行、每列大致像素总和较为接近; 从图 6(d) 可以明显看出直方图较为集中, 与周围其他区间之间的差值很大。根据此特性, 算法设定一个直方图非零区间的平均阈值  $\mu$ , 本文  $\mu$  的参考范围为  $280 \sim 350$ , 式(6)  $\beta_1$  的取值范围为  $0.95 \sim 0.985$ ,  $\beta_2$  的取值范围为  $0.90 \sim 0.95$ 。

#### 3.2 选取头肩目标特征

用于检测的训练样本包括 1780 个行人头肩正样本和 2000 个人头负样本, 其中正样本截取于不同场景下拍摄行的人视频帧, 负样本来源于不存在行人的视频帧。根据测试视频的参数以及摄像头的拍摄距离, 本文的训练样本采用  $32 \times 32$  像素, 如果视频参数不同, 选择样本的尺寸也不一样。在测试视频中本文选取了不同块大小、块滑动距离、胞元大小(单位均为像素), 分别取第 1, 100, ..., 1000 帧共 10 张图像在 1.1 节得到的行人头肩目标前景区域中进行检测算法测试,

平均正确率以及算法效率如表 1 所列。

表 1 不同 HOG 参数的检测结果

块尺寸	滑动距离	胞元尺寸	准确率/%	时间/帧
8 * 8	8 * 8	4 * 4	88.7	26
16 * 16	8 * 8	8 * 8	84.8	42
8 * 8	4 * 4	4 * 4	93.5	33

根据表 1 实验结果得出, 综合考虑, 在均衡性能与准确率的情况下, 块尺寸  $8 \times 8$ 、滑动距离  $4 \times 4$ 、胞元尺寸  $4 \times 4$  为最优选择。

#### 3.3 算法分析

图 7 给出了本文的目标检测方法传统的 HOG+SVM 算法的效率对比。选取图书馆拍摄视频的 1-1200 帧, 对传统的 HOG+SVM 检测头肩目标和本文提出的前景区域目标检测算法做比较。经实验验证, 本文目标检测算法明显减少了检测算法的运行时间, 其平均检测时间比未优化的目标检测算法少 100ms 左右。

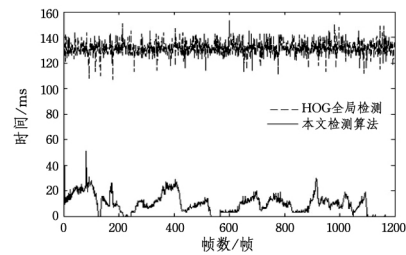


图 7 目标检测算法效率的比较

不同数量的目标同时进入检测带时, 滑动窗口的时间复杂度以及目标跟踪需要的总时间复杂度如表 2 所列, 其中滑动窗口参数为: 块尺寸  $8 \times 8$ , 滑动距离  $4 \times 4$ , 胞元尺寸  $4 \times 4$ 。

表 2 算法时间效率

目标数量/个	检测时间/ms	跟踪总时间/ms
1	11	26
2	20	60
3	32	78
4	39	95

图 8 给出了在第 274-280 帧间, 两个行人运动过程中两种比较主流的跟踪算法与本文跟踪算法效果的比较。从中可以明显看出, 采用 MeanShift 和 PF 跟踪时, 由于 1 号目标和 2 号目标产生遮挡情况, 目标跟踪丢失; 而本文提出的跟踪算法在目标遮挡的情况下有着不错的效果。

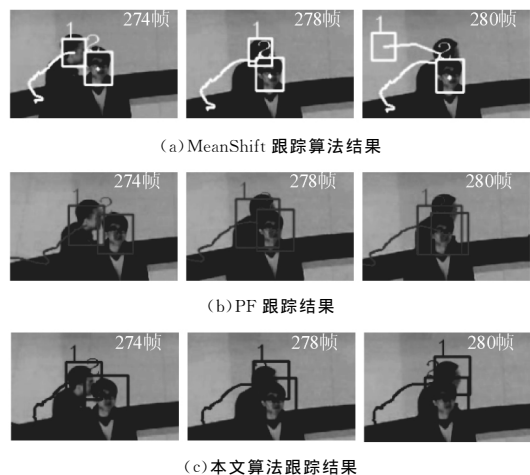


图 8 多目标跟踪效果的比较

表3给出了上述两者间的头肩目标检测正确率的对比结果。为了更加客观地与本文提出的目标检测方法进行比较,比较同一段视频中在相同的视频帧和检测区域内检测的目标个数,统计每帧中实际出现的目标个数和检测目标个数。视频选取1500—2000帧,每隔5帧检测一次,一共100帧作为测试视频帧,统计检测到的目标数、漏检目标数以及误检目标数。

表3 视频的检测结果

目标检测 算法	检测 目标	实际 目标	漏检 目标	误检 目标	准确率 /%
HOG+SVM	54	36	3	11	61.11
本文检测算法	32	36	4	1	86.11

表3中的误检目标是指将非头肩目标误判为头肩目标的情况,如图9(a)所示;漏检目标指本应检测到正确头肩目标而实际并未检测到的目标,如图9(b)所示。准确率=(实际目标数-误检目标数-漏检目标数)/实际目标数。由表3可以明显看出,本文所提目标检测算法有效减少了误检目标数。

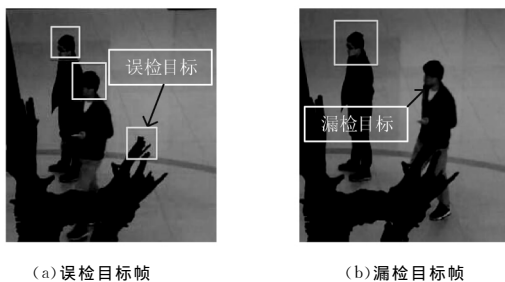


图9 误检和漏检目标

在同一段视频1—2400帧中,使用本文目标检测算法检测到头肩目标并利用模板目标匹配算法跟踪,在加入与未加入KCF跟踪算法时的行人统计与正确率的对比如表4所列。

表4 跟踪算法比较结果

跟踪算法	实际人数	统计人数	准确率/%
未加 KCF	36	49	73.47
增加 KCF	36	37	97.30

## 4 方法讨论

根据实验,本文提出的检测方法具有一定的局限性,在出现下列问题目标检测效果并不理想。如图10所示,当行人并肩进入检测区域内时,由于视频采用的是单目摄像,两个行人前景融合在一起,切割困难,计算得到的目标检测范围为图10(a)中的白色框,前景如图10(b)所示,靠近图像下方的目标没有被计算到检测范围内,产生了漏检问题。

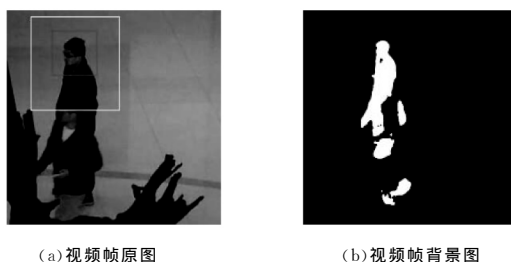


图10 漏检情况

当增加运动目标前景的检测范围(扩大图10中的白框)时,检测的速度就会降低,此类情况在本算法中未能得到有效的解决。

结束语 本文提出的基于区域定义的人数统计方法有如下的创新点:

(1)针对行人目标检测,传统的全局检测方法如HOG+SVM检测率低,基于部件检测行人模型如DPM,模型复杂,不能满足实时性的问题,提出了自适应学习速率提取运动目标前景作为目标检测的检测区,不仅有效减少了检测时间,而且降低了目标检测的误检率。

(2)提出了一种基于KCF的目标匹配跟踪算法,解决了KCF不能及时更新模型而引起的预选框误差偏大的问题;此外,KCF算法可以弥补由于头肩目标检测算法在遮挡情况下漏检而产生的目标匹配失败的短板,极大地提高了统计的准确率。

(3)针对传统过线统计人数的方式,本文提出了一种基于热点区域定义的人数统计方法,其具有更广泛的应用场景,不仅可以统计4个方向上的人流量,而且通过计算可以大概估计一段时间内停留在该区域内的人数,从而可以判断该区域的热度。

本文给出的基于区域定义的人数统计算法兼顾了算法的实时性和准确性,同时适用于多方向人流量以及封闭区域内的人数统计,适用场景灵活,并且具有广泛的应用前景和商业参考价值。不过在目标检测算法上,目标的前景分割算法和其他特征的选取缺少进一步的优化和分析,之后可以结合图像深度和机器学习来分割目标前景。

## 参考文献

- [1] HASHEMZADEH M, FARAJZADEH N. Combining keypoint-based and segment-based features for counting people in crowded scenes[J]. Information Sciences An International Journal, 2016, 345(C): 199-216.
- [2] 常庆龙,夏洪山,黎宁.一种基于归一化前景和角点信息的复杂场景人数统计方法[J].电子与信息学报,2014,36(2):312-317.
- [3] CONG Y, GONG H, ZHU S C, et al. Flow mosaicking: Real-time pedestrian counting without Scene-specific learning[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2009:1093-1100.
- [4] CHEN C L, CHEN K, GONG S, et al. Crowd Counting and Profiling: Methodology and Evaluation[M]// Modeling, Simulation and Visual Analysis of Crowds. Springer New York, 2013: 347-382.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005: 886-893.
- [6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object Detection with Discriminatively Trained Part-Based Models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 32(9): 1627-1645.
- [7] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2016: 770-778.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// IEEE Conference on Computer Vision & Pattern Recognition. 2016: 779-788.

(下转第201页)

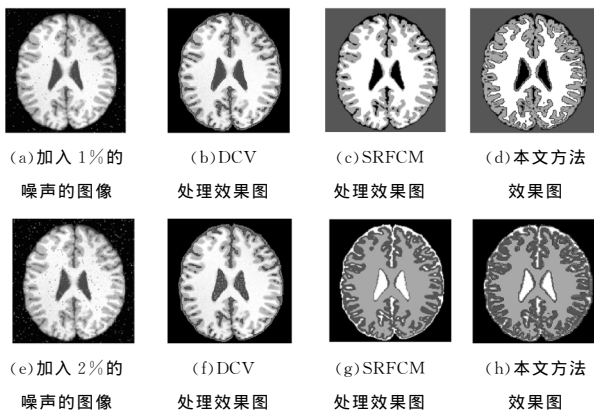


图 4 DCV 模型和本文模型分割对比图一

图 5 是另一幅医学图像的分割效果图,在图像信息量增大并且有噪声干扰的情况下,本文模型依旧可以完全分割出多目标区域。实验表明,SRFCM+DCV 算法模型可行,既能够抑制噪声,又能分割多目标,在性能上优于传统的单一分割算法。

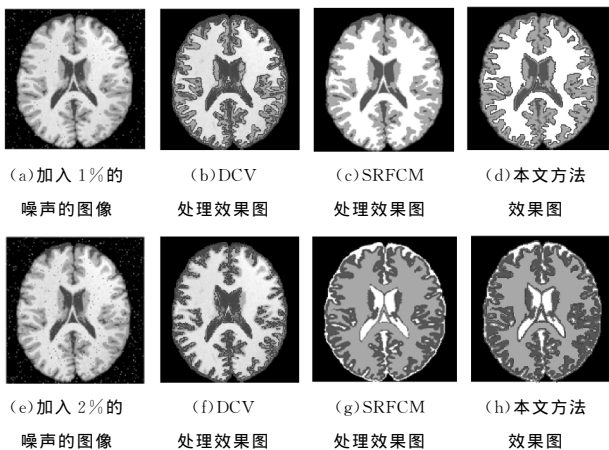


图 5 DCV 模型和本文模型分割对比图二

**结束语** 本文提出的基于抑制式模糊聚类算法的改进型双水平集模型,比传统的水平集模型对伴有噪声的多目标医学图像的分割效果较明显。通过 SRFCM 聚类算法进行图像降噪,降低算法的敏感性,根据 NMI 准则的值来判定聚类效

果,采用增加惩罚项的方法改进 DCV,使得新双水平集模型无需重新初始化,最终实现多相分割。实验表明,采用该算法分割的效果图边缘清晰、内部组织被完全分割,算法可行高效,可对其做进一步分析研究。

## 参考文献

- [1] 方江雄. 变分和偏微分方法在图像分割中的应用[M]. 北京:中国石化出版社,2015:3-9.
- [2] OSHER S, SETHIAN J A. Fronts propagating with curvature dependent speed: Algorithms based on the Hamilton-Jacobi formulation [J]. *Journal of Computational physics*, 1998, 79: 12-49.
- [3] CHAN T, VESE L. Active contours without edges [J]. *IEEE Transactions on Image Processing*, 2001, 10(2): 266-276.
- [4] VESE L A, CHAN T F. A multiphase level set framework for image segmentation using the Mumford and Shah model [J]. *International Journal of Computer Vision*, 2002, 50(3): 271-293.
- [5] LI C, XU C, GUI C, et al. Level set evolution without re-initialization: A new variational formulation [C] // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005: 430-436.
- [6] FAN J L, ZHEN W Z, XIE W X. Suppressed fuzzy C-means clustering algorithm [J]. *Pattern Recognition Letters*, 2003, 24(9/10): 1607-1612.
- [7] CAI W, CHEN S, ZHANG D. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation [J]. *Pattern Recognition*, 2007, 40(3): 825-838.
- [8] 范九伦. 抑制式模糊 C-均值聚类研究综述 [J]. *西安邮电大学*, 2014, 3(1): 2095-6533.
- [9] KNOPS Z F, MAINT J B, VIERGVER M A. Normalized mutual information based registration using k-means clustering and shading correction [J]. *Medical Image Analysis*, 2006, 10(3): 432-439.
- [10] LI C M, XU C Y. Distance Regularized Level Set Evolution and Its Application to Image Segmentation [J]. *IEEE Transactions on Image Processing*, 2010, 19(12): 3243-3253.
- [11] 唐春晖. 一种基于梯度方向直方图的俯视行人的检测方法 [J]. *模式识别与人工智能*, 2015, 28(1): 19-26.
- [12] 文嘉俊, 徐勇, 战荫伟. 基于 AdaBoost 和帧间特征的人数统计 [J]. *中国图象图形学报*, 2011, 16(9): 1729-1735.
- [13] 徐超, 高梦珠, 查宇锋, 等. 基于 HOG 和 SVM 的公交乘客人流量统计算法 [J]. *仪器仪表学报*, 2015, 36(2): 446-452.
- [14] 周治平, 许伶俐, 李文慧. 特征回归与检测结合的人数统计方法 [J]. *计算机辅助设计与图形学学报*, 2015, 27(3): 425-432.
- [15] LIU J, ZHANG M. A detection and tracking based method for real-time people counting [C] // *Chinese Automation Congress*, 2013: 470-473.
- [16] 郑翔翔, 张怡. 基于头部检测和跟踪的人数统计方法 [J]. *计算机应用与软件*, 2011, 28(2): 44-46.
- [17] 顾德军, 伍铁军. 一种基于人头特征的人数统计方法研究 [J]. *机械制造与自动化*, 2010, 39(4): 134-138.
- [16] 夏菁菁, 高琳, 范勇, 等. 基于骨架特征的人数统计 [J]. *计算机应用*, 2014, 34(2): 585-588.
- [17] HENRIQUES J F, RUI C, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 37(3): 583-596.
- [18] ZHANG K, ZHANG L, YANG M H. Real-time compressive tracking [M] // *Computer Vision-ECCV 2012*. Springer Berlin Heidelberg, 2012: 864-877.
- [19] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.
- [20] HARE S, SAFFARI A, TORR P H S. Struck: Structured output tracking with kernels [C] // *International Conference on Computer Vision*. *IEEE Trans Pattern Anal Mach Intell*, 2011: 263-270.

(上接第 178 页)