

中图分类号: TP391.4 文献标志码: A 文章编号: 1006-8961(2011)09-1729-07

论文索引信息: 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

基于 AdaBoost 和帧间特征的人数统计

文嘉俊¹⁾, 徐勇¹⁾, 战荫伟²⁾

¹⁾(哈尔滨工业大学深圳研究生院计算机科学与技术学院, 深圳 518055) ²⁾(广东工业大学计算机学院, 广州 510006)

摘要: 视频监控中出入口人数统计的难点在于人流密集时对每个人体的准确分割。通过学习的方法得到人头检测的分类器,并在垂直拍摄图像中提取人头候选区域,以分离相互靠近的人体目标,进而根据人头的运动特征剔除静止误检区域,根据误检目标检测频率低及其响应位置不连续的特征剔除动态误检区域。最后提出一种简易可行的过线跟踪方案以完成计数。实验中对各种复杂情况的过线视频进行测试,正确率能够达到95%以上。与传统方法相比,本文方法解决了多人过线或搬货物过线时传统方法难以准确完成的行人分割和计数,更适于实际情况的人数统计。

关键词: AdaBoost; 人头识别; 帧差法; 检测响应特征; 目标跟踪

People counting based on AdaBoost and inter-frame features

Wen Jiajun¹⁾, Xu Yong¹⁾, Zhan Yinwei²⁾

¹⁾(School of Computer Science and Technology, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055 China)

²⁾(Faculty of Computer, Guangdong University of Technology, Guangzhou 510006 China)

Abstract: The difficulty of counting the number of people using video surveillance lies in the correct segmentation of each person in crowded situations. For this problem, we train a classifier to detect head candidates so as to distinguish between people walking closely under the overhead mounted camera. The static false detected candidates are removed according to the motion feature of the head, while the dynamic ones are eliminated due to their low detected frequency and discontinuous response positions. Finally, an easy and effective crossing tracking scheme is proposed for counting. We conducted experiments in various complex situations. It shows that the counting rate is over 95%. Compared to the traditional method, the proposed means can segment and count the people more accurately in cases where multiple people are crossing or crossing with items. It is therefore much more suitable for people counting in real life situations.

Keywords: AdaBoost; head recognition; inter-frame differencing; detection response features; object tracking

0 引言

人数统计应用范围广泛,譬如,关键区域人数统计、百货商场出入口人数统计、电梯口人数统计、地铁口人数统计、公交车上下人数统计等。这些区域的人数或客流量对公共安全防范、市场决策和交通

配置都有着重要的参考价值。相比传统的人数统计方式,基于视觉的人数统计方法具有统计精度高、实施便利的特点,因此越来越受到人们的重视。

实际应用场合环境复杂多变,视觉人数统计方法统计精度的提高受到以下因素影响:运动人体遮挡、随身物干扰和光线变化。为了避免发生严重遮挡,研究人员在实验中多采用垂直摄像头布控。所

收稿日期:2011-01-04;修回日期:2011-04-11

第一作者简介:文嘉俊(1983—),男,哈尔滨工业大学深圳研究生院计算机科学与技术专业博士研究生,主要研究方向为图像处理与模式识别。E-mail:enjoy_world@163.com。

使用的方法主要分为 3 类:运动前景分割方法^[1-5],轮廓匹配方法^[6-7]和运动矢量统计方法^[8]。

基于运动前景分割的人数统计方法在行人与背景颜色相似、光照变化、多人发生分裂与合并时性能较差。文献[1]结合背景减法和帧差法,在目标与背景颜色相似时,仍可继续提供较完整的前景区域。并根据帧间图像对应 4 个角区域内的亮度变化对背景进行快速光照补偿。文献[2]使用目标轮廓作为前景更新对象,由于仅使用了边缘信息,因此能有效抵抗光照的影响。文献[3]根据行人在图像中所占面积对前景进行分裂与合并,在理想的前景分割下能够获得很好的计数结果。文献[4]在此基础上提出丢失查找算法,能够还原前景分割失效而丢失的目标。文献[5]使用了先验知识和 k 均值聚类算法分离运动区域。然而这些依赖于前景分割的人数统计方法在行人携带大物件通过时会失效。

此外,部分学者在运动前景提取的基础上使用轮廓匹配方法定位目标,文献[6]使用半圆模型搜索前景中人头肩所在位置,然后使用蛇模型跟踪目标,然而当行人发生重叠时,定位精度大为下降;文献[7]使用圆环模型定位人头,但实际中的人头形状较复杂,非圆环模型所能适应。针对运动分割方法存在难以正确分离、合并前景目标的问题,文献[8]使用运动矢量描述过线行人,通过具有运动矢量的特征点的密度和方向聚类性可以估计出行人的个数和运动方向。但实际中过线行人可能会受到干扰和发生重叠的情况,造成计数精度下降。

垂直摄像头可以降低行人之间的遮挡程度,但是避免不了行人相互靠近时的遮挡问题。从实际环境和大多数实验视频的观察可知,行人在相互靠近时,只有人头之间基本不发生遮挡。因此可以从这个角度出发分割运动目标。与轮廓匹配方法不同,通过样本训练的方法,学习一个良好的人头检测分类器,然后根据运动人头的特征去除误检区域,最后配合过线跟踪实现出入口人数统计。

1 人数统计方案

为进行出入口人数统计,采用在天花板安装垂直摄像头的方式进行实验。首先在离线状态下对垂直人头模型进行训练,得到人头检测的分类器;然后设计并实现了人数统计系统,具体流程如图 1。对于输入图像,系统同时进行人头检测和帧差法。当

前得到的候选人头区域,必须达到一定的检测强度和符合运动人头的特征才识别为人头。运动强度判决模块利用帧差法判断所得到的人头检测区域是否具有运动趋势,并剔除没有运动趋势的检测区域。同时人头大小判决模块采用迭代的方式估计场景中的人头大小,并把过大和过小的检测区域去掉。最后,当检测到人头与虚拟线相交时开始对目标跟踪,实现计数过程。

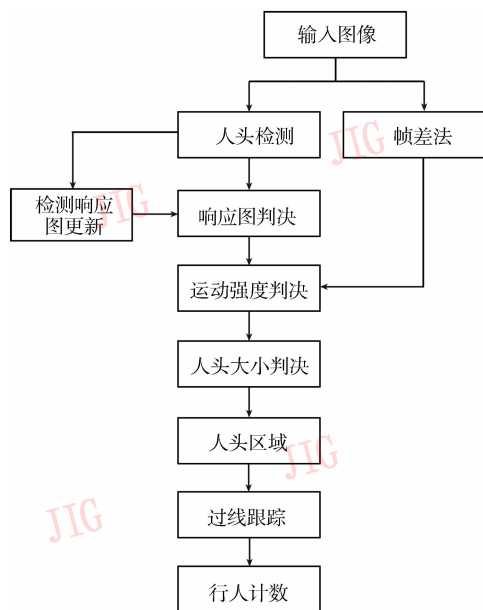


图 1 人数统计流程图

Fig. 1 Flow chart of people counting

该方案的优势有两点:1)能够在行人相互靠近情况下分割单个目标,并且在多人一起搬运大件货物时正确分割行人;2)通过训练的方式检测人头对光照变化具有较强的适应性。

2 基于 AdaBoost 的人头检测

在垂直摄像头环境下,不同人头有近似圆的形状,但也存在差异,而同一人头在画面中的不同位置,形状会发生相应变化。使用基于轮廓的方法难以对人头进行精确定位。因此使用基于学习的方法对人头进行检测,步骤如下:1)在分辨率为 20×20 的人头样本中使用积分图方法提取扩展 Haar 特征^[9],如图 2 所示,每一个特征对应一个弱分类器;2)使用 AdaBoost 算法^[10-11]训练弱分类器获得强分类器,并把多个强分类器串联得到分级分类器;3)利用最终得到的人头检测分类器检测输入图像

产生候选人头区域。

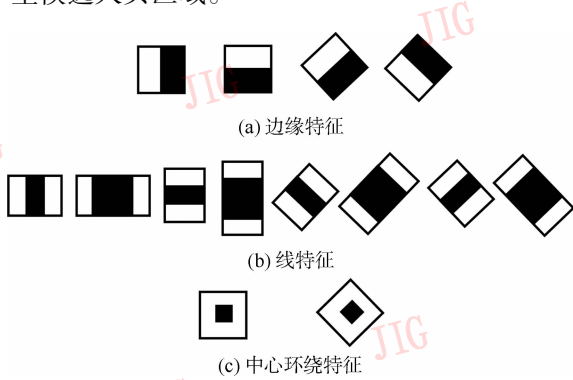


图2 Haar 特征

Fig.2 Haar features

在不同环境下采集了 1 000 张人头正样本和 1 957张负样本,图3 为部分样本示例。从图中可看出 Haar 中心环绕特征非常适合于对正样本的描述。而负样本包括了室内外各种光照环境下的场景。然而通过分类器检测出的区域会有误检发生,必须进一步剔除误检目标。

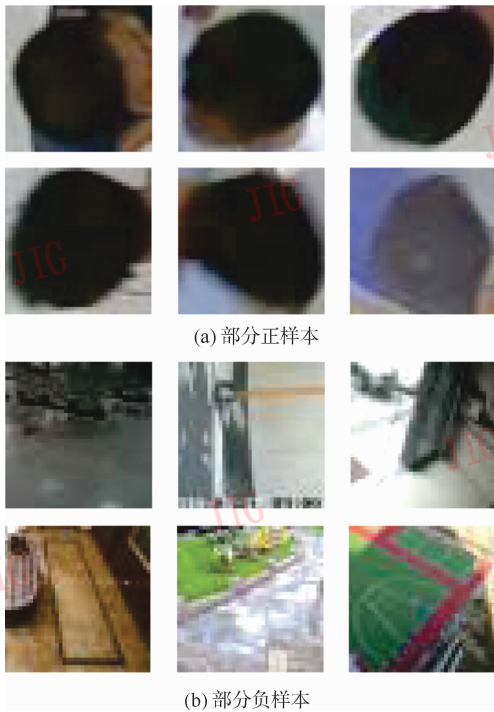


图3 部分样本

Fig.3 Part of the sample images

3 帧间特征人头识别

对输入图像进行人头检测所产生的误检主

要由以下 3 方面造成:1) 训练的正负样本种类(包括各种形状、背景和光照环境下的人头等)和个数不够充分;2) 背景中存在着与人头相似的目标;3) 运动前景在运动变化过程中出现了与人头相似的目标。因此诸如地面的斑点,场景中的凳子,行人身体的某些部位会给人头检测与识别带来一定影响。

3.1 人头区域检测响应特征模型

实验中发现,视频中出现的人头能够被正常检测,同时运动前景中有时会产生误检目标,然而所检测到的人头和动态误检目标在检测频率和帧间检测位置的连续性上存在着较大差别。往往人头在出现时间段内的空间位置上能够被连续检出,而上面所提到的误检目标仅在少数时候检测出来。由于该类目标处于运动状态,因此其检测位置呈现不连续的状况。根据这个特征,可识别真正的人头而去除误检目标。算法中定义了人头检测响应图,用来记录目标被人头检测分类器成功检测的响应强度。

假设在帧数区间 $[F_1, F_2]$ 内出现的检测目标总共有 m 个,第 $j(1 \leq j \leq m)$ 个检测目标出现的帧数区间为 $[F'_j, F''_j]$ 。令 $P_{ij}, R_{ij}, N_{ij}, S_{ij}(P_{ij}, R_{ij}, N_{ij})$ 分别为第 j 个目标在第 i 帧时的位置、响应值、空响应值和状态。若目标 j 在第 i 和 $i+1$ 帧都被检测出来则称 S_{ij} 和 $S_{(i+1)j}$ 关联。令 $C(S_{ij}, S_{(i+1)j})$ 为前后两帧间的状态关联匹配函数,

$$C(S_{ij}, S_{(i+1)j}) = \begin{cases} 1 & S_{ij} \text{ 与 } S_{(i+1)j} \text{ 关联} \\ 0 & \text{其他} \end{cases} \quad (1)$$

目标的响应值、空响应值和位置按下式更新

$$R_{(i+1)j} = \begin{cases} R_{ij} + 1 & C(S_{ij}, S_{(i+1)j}) = 1 \\ R_{ij} & C(S_{ij}, S_{(i+1)j}) = 0 \text{ 且 } R_{ij} \geq T_1 \\ 0 & \text{其他} \end{cases} \quad (2)$$

$$N_{(i+1)j} = \begin{cases} 0 & C(S_{ij}, S_{(i+1)j}) = 1 \text{ 或 } N_{ij} \geq T_2 \\ N_{ij} + 1 & C(S_{ij}, S_{(i+1)j}) = 0 \end{cases} \quad (3)$$

$$P_{(i+1)j} = \begin{cases} P_{(i+1)j} & C(S_{ij}, S_{(i+1)j}) = 1 \\ P_{ij} & C(S_{ij}, S_{(i+1)j}) = 0 \text{ 且 } N_{ij} < T_2 \end{cases} \quad (4)$$

检测响应图在人数统计中的作用有以下 3 个方面:1) 累加连续被检测到的目标的响应强度,只有当响应强度值大于一定阈值的目标才符合人头区域的假设;2) 当目标的响应强度值达到一定阈值,且在当前帧中漏检,则为该目标延续几帧的强度值和位置,以保持对该目标检测的连续性;3) 空响应值达到一定阈值的目标则被剔除。检测响应图的更新流程(见图4)。

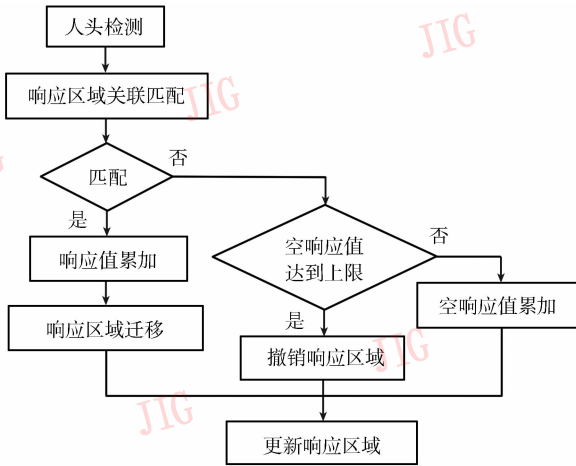


图4 检测响应图更新模块

Fig. 4 The update module of detection response image

3.2 人头区域运动特征模型

在人头检测中可能存在误检的静止目标。该类目标往往具有与人头相似的形状,但没有运动趋势。然而在正常情况下,越过虚拟线的行人头部会有运动发生,即使行人暂时停下来,在相隔较远的两帧中人头部位总有运动趋势。因此通过定义人头区域的运动强度,可以剔除这类误检静止目标。

帧差法是常见的目标检测方法,可以快速反映视频画面中的运动变化,使用相隔 n 帧的图像之差来定义人头区域的运动强度。设 $I_k(x, y)$ 为第 k 帧获取的灰度图像, $D_k(x, y)$ 为帧间差的绝对值, $F_k(x, y)$ 为二值差分图, (x_t, y_t) 、 w_t 、 h_t 和 E_t 分别为当前帧第 t 个人头检测框的左上角坐标、宽度、高度和运动强度,则

$$D_k(x, y) = | I_k(x, y) - I_{k-n}(x, y) | \quad (5)$$

$$F_k(x, y) = \begin{cases} 0 & D_k(x, y) \leq T_0 \\ 1 & D_k(x, y) > T_0 \end{cases} \quad (6)$$

$$E_t = \frac{1}{w_t \times h_t} \sum_{y_t}^{y_t+h_t-1} \sum_{x_t}^{x_t+w_t-1} F_k(x, y) \quad (7)$$

若 $E_t \geq T$,则符合过线行人的实际情况,否则剔除该区域中的误检目标,实验中 $n = 3, T = 0.05$ 。

3.3 人头大小估计

在垂直摄像头环境下,人头大小恒定在某一范围之内。因此让人头大小为每个人头检测框大小的加权平均,可以估计人头包围框的大小。令出现的人头检测框大小为 B_k ,人头大小估计值为 H_k ,有

$$H_k = \alpha H_{k-1} + (1 - \alpha) B_k \quad (8)$$

式中 α 为学习率。当算法运行一段时间后,凡是大

小符合 $| H_k - B_k | > d$ 的人头检测框为过小或过大检测框,需要剔除。实验中取 $\alpha = 0.3, d = H_k / 3$ 。

经过静、动态误检目标剔除和非人头大小区域的过滤后,最终得到可信度较高的候选人头区域。

4 过线跟踪与计数

当人头检测框与预先设置的虚拟线相交时,开始执行跟踪。算法中的跟踪对象为检测框与虚拟线的相交线段。设第 i 个人头检测框在过线时与虚拟线的相交次数为 U_i ,检测框与虚拟线相交的线段为 $p_{ij}q_{ij}, j = 1, 2, \dots, U_i$,检测框左上角点到虚拟线的垂直距离为 $d_{ij}, j = 1, 2, \dots, U_i$ 。图5为第 k 帧和第 $k + 1$ 帧时3个人头检测框与虚拟线相交的情况。在前后帧间,检测框与虚拟线相交的线段在长度和位置上均发生一定的变化,如第 k 帧的跟踪线段 $p_{11}q_{11}$ 在第 $k + 1$ 帧变为 $p_{12}q_{12}$ 。

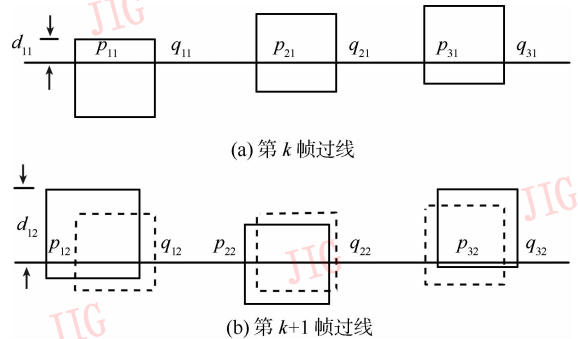


图5 过线跟踪

Fig. 5 The tracking of line crossing

对于过线目标 i 使用近邻法可以实现其与虚拟线的相交线段在帧间的关联和跟踪。同时通过对序列 $d_{ij}, j = 1, 2, \dots, U_i$ 分析可以判断目标的进出方向。当行人从虚拟线上方走向虚拟线下方时,序列 d_{ij} 的值由大变小,反之,序列 d_{ij} 的值由小变大。因此当式(9)成立时,可以判断目标方向往下,否则目标方向往上。

$$\frac{1}{\left\lfloor \frac{U_i}{2} \right\rfloor} \sum_{j=1}^{\left\lfloor \frac{U_i}{2} \right\rfloor} d_{ij} > \frac{1}{\left\lceil \frac{U_i}{2} \right\rceil} \sum_{j=\left\lceil \frac{U_i}{2} \right\rceil}^{U_i} d_{ij} \quad (9)$$

5 实验结果与分析

为了验证提出的人数统计方法,使用多段不同类型的视频进行测试,包括双向同时过线,搬货物过线等。并在光照变化情况下对人头检测进行了测试。实验在装有 Windows 操作系统的环境下运行,计算机处理器为英特尔酷睿 2,主频 2.2 GHz,内存 2 G。当图像大小为 352×288 时,算法对实际的视频场景每秒处理 30 ~ 40 帧。

5.1 误检滤除

图 6(b) 中的每一个检测目标均有一个响应序列,其中序列越长代表该目标的检测响应值越大,序列灰度由浅到暗,在灰度最暗的位置为目标当前响应最强烈的位置。

根据响应序列的长度,可以去除部分行人在过线时的误检目标。由实验可知,这些误检目标包括俯视下的肩膀、脚、衣服图案等,如图 7(a) — (c)。它们所表现出来的特征是被检测到的次数低且在空间上的检测位置不连续,所以其对应的响应序列长度往往较短。图 7(d) — (f) 显示滤除低响应误检目标后的结果。

除了在运动行人上会产生误检目标,地面的一些场景也会引起误检。但这类目标较容易去除,它们往往处于静止状态。因此通过引入运动特征可以提高检测的正确性。图 8(a) 中右边有一误检的凳子,在引入运动特征后,如图 8(c) 所示,可以去掉凳子的影响,图 8(b) 为最终效果。

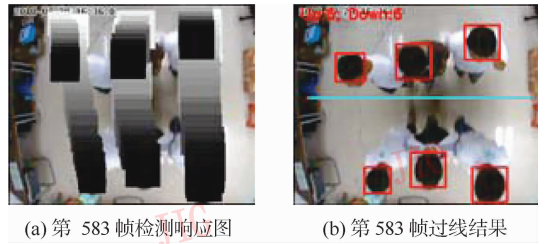


图 6 检测响应图

Fig. 6 Detection response image

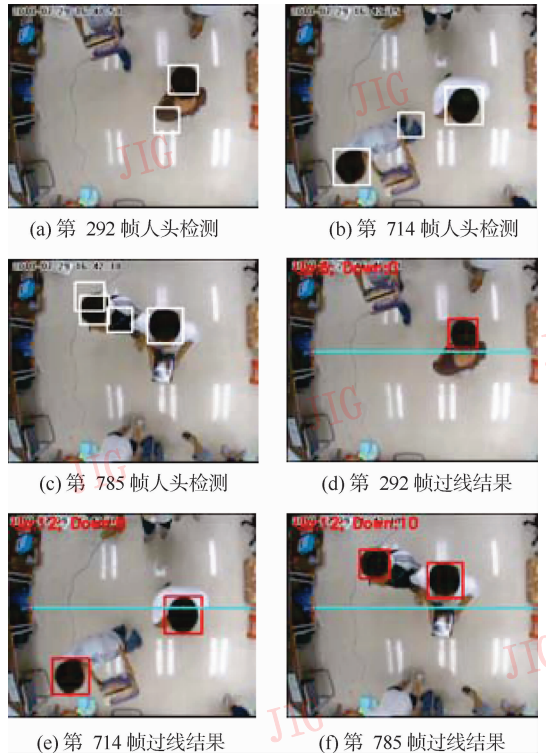


图 7 低响应目标滤除

Fig. 7 Filtering of low response target



图 8 静态误检目标滤除

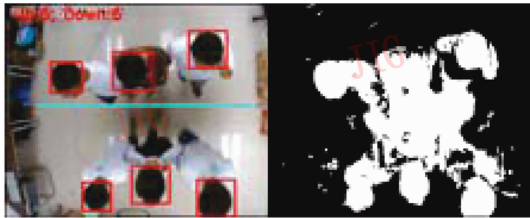
Fig. 8 Filtering of static false detected object

5.2 过线多目标分割

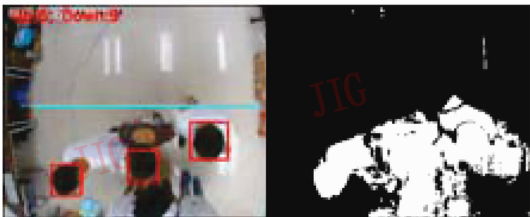
过线时行人的合并与分离是难点。传统的运动

前景分割方法进行人数统计往往依赖于前景分割的正确性和完整性。图 9 为人群过线时本文方法的正

确跟踪结果和使用高斯背景建模方法获得的前景图像。由前景图像观察可知高斯背景方法难以正确分割目标,这主要受到行人衣服与背景色相似以及阴影的影响。使得行人区域没有完整提取,而部分阴影区域由于背景更新速度赶不上却被分割为前景区域。因此若使用单个行人在前景中所占面积来估计前景中的人数,必定会带来误差。



(a) 第 584 帧过线与 GMM 法前景



(b) 第 603 帧过线与 GMM 法前景

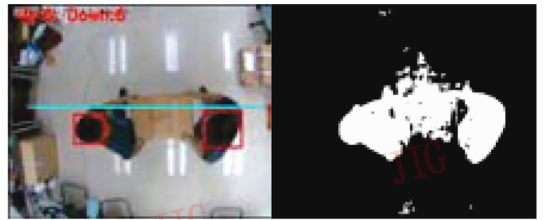
图 9 多人过线

Fig. 9 Multiple people cross the line

回顾以往的研究成果,很少有研究人员考虑行人在过线同时搬货物的情况。然而这也是常发生的事情。因此使用运动分割得到的前景区域中并非都有行人。还必须引入对前景区域的识别,判断过线的运动区域中有多少行人。图 10 显示了人头检测方法对行人个数的正确判断,完成了搬货物过线的正确人数统计。

为了测试人头检测在各种光照环境下的稳定性,对拍摄视频进行人为的光照干预。从图 11 的 4 种光照环境的测试结果来看,人头检测对光照变化具有相对稳定的适应性。

实验中算法对多段视频进行测试,并使用文献[5]的方法评价本算法的性能。定义 TP 、 FP 和 FN 分别为过线的正确检测个数、误检个数和漏检个数定义 PR 、 RE 和 F 分别为正确率、查全率和调和平均,其中调和平均表示融合正确率与查全率所得到的评价,则有 $PR = TP / (TP + FP)$, $RE = TP / (TP + FN)$ 和 $F = 2 \cdot PR \cdot RE / (PR + RE)$ 。



(a) 第 804 帧过线与 GMM 法前景



(b) 第 812 帧过线与 GMM 法前景

图 10 搬货物过线

Fig. 10 Cross the line with items



(a) 暗光照检测



(b) 亮光照检测



(c) 偏亮光照检测



(d) 昏暗光照检测

图 11 光照变化人头检测

Fig. 11 Head detection under lighting changes

根据表 1 的测试结果,算法的计数精度达到 95% 以上。双向过线的视频之所以有误检产生,主要是由于行人穿深色衣服时,其肩膀或者衣服上的某些图案在俯视角度下与人头相似,而且持续出现在画面中。算法暂时还不能排除所有的误检情况。而视频中出现漏检的情况主要是由于行人在过线前其人头检测失效,而当重新被检测时,对应的检测响应值没有达到一定阈值,因此在过线时漏掉了该目标。另外,当过线的行人遮挡了头部,算法不能够识别出人头,也会出现漏检情况。这些问题将在以后的工作中继续深入研究。

表1 人数统计结果

Tab.1 Result of people counting

视频名	实际过线进+出	过线计数进+出	正确检测	误检	漏检	正确率/%	查全率/%	调和平均/%
双向1	10+10	10+10	20	0	0	100	100	100
双向2	8+8	8+8	16	0	0	100	100	100
双向3	15+15	14+17	29	2	1	94	97	95
双向4	12+12	12+12	24	0	0	100	100	100
双向5	18+18	19+18	36	1	0	97	100	98
多人	23+26	23+26	49	0	0	100	100	100
携随身物	28+28	28+27	55	0	1	100	98	99
搬货物	10+10	8+10	18	0	2	100	90	95
光线变化	2+10	2+10	12	0	0	100	100	100

6 结论

提出使用基于 AdaBoost 的人头检测的方法检测过线的行人目标,检测性能在各种光照环境下获得良好的效果。同时定义检测响应图,根据部分动态的误检目标检测响应值低及其在帧间检测位置不连续的特点,滤除该类误检目标。引入运动人头特征剔除静态的误检目标,并通过估计视频中人头的大小去掉过大和过小的误检目标。算法解决了多人同时过线或搬货物过线时传统方法难以准确完成的行人分割和计数。经实验测试,人数统计正确率达到95%以上。

本文算法仍存在着误检和漏检的情况,因此在以后的研究中,将考虑结合前景运动分析和人头识别来解决当前存在的问题。

参考文献 (References)

- [1] Kim J W, Choi K S, Park W S, et al. Robust real-time people tracking system for security [J]. IBS Journal, 2002, 2(3): 184-190.
- [2] Yu Shengsheng, Chen Xiaoping, Sun Weiping, et al. A robust method for detecting and counting people [C]//Proceedings of International Conference on Audio, Language and Image Processing. Piscataway, NJ, USA; IEEE Press, 2008; 1545-1549.
- [3] Chen Thouho, Hsu Chewei. An automatic bi-directional passing-people counting method based on color Image processing [C]//Proceedings of 37th IEEE International Carnahan Conference on Security Technology. Piscataway, NJ, USA; IEEE Press, 2003; 200-207.
- [4] Septian H, Tao Ji, Tan Yappeng. People counting by video segmentation and tracking [C]//Proceedings 9th International Conference on Control, Automation, Robotics and Vision. Piscataway, NJ, USA; IEEE Press, 2006; 1-4.
- [5] Antić B, Letić D, Čulić D, et al. K-means based segmentation for real-time zenithal people counting [C]//Proceedings of 16th IEEE International Conference on Image Processing. Piscataway, NJ, USA; IEEE Press, 2009; 2565-2568.
- [6] Jaijing K, Kaewtrakulpong P, Siddhichai S. Object detection and modeling algorithm for automatic visual people counting system [C]//Proceedings of 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. Piscataway, NJ, USA, IEEE Press, 2009, 2; 1062-1065.
- [7] Gardel A, Bravo I, Jimenez P, et al. Real time head detection for embedded vision modules [C]// International Symposium on Intelligent Signal Processing. Piscataway, NJ, USA; IEEE Press, 2007; 1-6.
- [8] Albiol A, Albiol A, Silla J. Statistical video analysis for crowds counting [C]//Proceedings of 16th IEEE International Conference on Image Processing. Piscataway, NJ, USA; IEEE Press, 2009; 2569-2572.
- [9] Rainer L, Jochen M. An extended set of Haar-like features for rapid object detection [C]//Proceedings of 16th IEEE International Conference on Image Processing. Piscataway, NJ, USA; IEEE Press, 2002; 900-903.
- [10] Paul V, Michael Jones. Rapid object detection using a boosted cascade of simple features [C]//Proceedings of the IEEE Computer Vision and Pattern Recognition. Piscataway, NJ, USA; IEEE Press, 2001; 151-155.
- [11] Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting [J]. Journal of Computer and System Sciences, 1997, 55(1): 119-139.