

中图法分类号: TP391 文献标志码: A 文章编号: 1006-8961(2011)05-0813-08  
论文索引信息: 何鹏, 麻文华, 黄磊, 刘昌平. 实时人数计数系统 [J]. 中国图象图形学报, 2011, 16(5): 813-820

# 实时人数计数系统

何鹏<sup>1),2)</sup>, 麻文华<sup>1),2)</sup>, 黄磊<sup>1)</sup>, 刘昌平<sup>1)</sup>

<sup>1)</sup>(中国科学院自动化研究所, 北京 100190) <sup>2)</sup>(中国科学院研究生院, 北京 100190)

**摘要:** 描述一个实时在线人数计数系统, 该系统采用检测加跟踪的方法来实现人数计数功能。在检测阶段, 采用MBLBP(multi-scale block LBP)特征, 从运动区域上检测出行人。该特征速度快, 并且在归一化下, 能够适应多尺度的应用; 在跟踪阶段, 通过一个概率模型, 将对行人的跟踪转化为对特征点的跟踪, 并且在将检测目标和跟踪目标进行一一对应时, 进一步利用各个目标内的特征点来完成相应的操作。最后用实际中不同场景下的视频, 对系统的性能进行测试, 同时还在一段公开的视频上进行了测试, 实验结果表明, 该系统能够在不同场景下较准确地实现人数计数功能。

**关键词:** 人数计数; 行人检测; 目标跟踪; 特征点

## Real-time people counting system

He Peng<sup>1),2)</sup>, Ma Wenhua<sup>1),2)</sup>, Huang Lei<sup>1)</sup>, Liu Changping<sup>1)</sup>

<sup>1)</sup>(Institute of Automation, Chinese Academy of Sciences, Beijing 100190 China)

<sup>2)</sup>(Graduate University of Chinese Academy of Sciences, Beijing 100190 China)

**Abstract:** This paper describes a real-time people counting system, which combines the pedestrian detection and tracking method together for people counting. The normalized MB-LBP feature, which can be easily calculated and suitable for multi-scale detection, is adopted for detection in the foreground at the detecting stage. The points are grouped together with a probability model to track the pedestrian at the tracking stage. At the same time, these points are employed to map the detected pedestrian to the tracked pedestrian. And in order to verify the performance of the system, three experiments are designed for testing. The first two videos are captured at different locations with different backgrounds; the third video, which has been used for testing in many other papers, is used for comparison. The experiments results demonstrate that this system performs the function of counting well at different background situations.

**Keywords:** people counting; pedestrian detection; object tracking; feature points

## 0 引言

随着计算机视觉的发展, 人数计数已成为视频监控领域中的研究热点之一。重要区域中人流量的统计结果, 可以为安全管理提供直接的指导。一些国家重要场所, 如, 鸟巢体育馆、奥林匹克公园等, 都需要对其区域内的人流量有准确的统计。参照这

些统计信息, 可以为未来节假日安全隐患控制管理做好准备。

一个完整的人数计数系统需要考虑如下的问题:

1) 实时性 人数计数系统必须能够实时地工作, 这样才能真正地应用到实际中;

2) 鲁棒性 在实际应用中, 通常不同的场景其背景图像有着非常大的差别, 并且在不同的时刻表

收稿日期: 2010-03-03; 修回日期: 2010-04-01

第一作者简介: 何鹏(1987—), 男, 中国科学院自动化研究所硕士, 主要研究方向为行人检测、行人跟踪、人数计数。

E-mail: peng.he@ia.ac.cn.

现出不同的光照特性,这些因素都对系统提出了较高的稳定性要求;

3) 准确性 人数计数系统在有着较强的鲁棒性情况下,应当提供准确度较高的结果,否则其计数结果就失去了参考的意义。

目前,很多论文已经提出了不同的人数计数方法,根据计数采用的摄像头安装位置,可以将这些方法归纳为基于垂直摄像头和基于倾斜摄像头的人数计数。基于垂直摄像头的系统,其优点是没有遮挡现象,但是摄像头的监控范围太小,不适合室外的场景。因此本文的人数计数仅针对基于倾斜摄像头的人数计数方法。

当前人数计数方法可以划分为三大类:

第 1 类方法采用一种统计的方法来估计某个区域的人数。该类方法通常是将运动区域的像素或者其他特征和该区域具体人数关联起来,然后训练一个函数来估计运动区域的人数。如 Kim 等人<sup>[1]</sup>和 Lee 等人<sup>[2]</sup>利用底层特征——前景像素累计和与运动向量来进行人数统计,像素累计和与人数关联,运动向量用于区分两个方向,该文中使用了较简单的特征来统计得到核函数。而 Chan 等人<sup>[3]</sup>则进一步使用多种特征来训练回归函数。使用的特征可以分为两大类:区域相关特征和区域内部边缘相关特征。区域相关特征,如区域面积、周长、区域外沿方向直方图、周长面积比等。区域内部边缘相关特征,包括区域内部边沿像素和值,内部边缘直方图和纹理信息等。总共使用了 28 种与该运动区域相关的特征。得到了较满意的计数效果。但是核函数的方法和具体场景关联性太大。

第 2 类方法采用检测加跟踪相结合的方法。该类方法通常都有一个预处理过程,在预处理过程中,提取出运动区域,然后在运动区域上检测行人。常用的检测方法有基于投影直方图的分割方法<sup>[4-5]</sup>,基于模板匹配的方法<sup>[6]</sup>和基于统计分类<sup>[7]</sup>(训练一个分类器)的方法。基于投影直方图的方法,如 Zhang 等人<sup>[4]</sup>首先采用高斯背景建模提取出前景区域,进一步阈值化得到前景二值图像,然后对此图像进行竖直方向投影,得到一个关于图像宽度和像素累积值得直方图,然后根据直方图的极值点来分割行人。Ma 等人<sup>[5]</sup>进一步地对图像在两个方向上均进行投影,解决同一竖直方向多人重合的情况。这种分割方法的优点是简单快速,但缺点也很明显,在人群较为密集的情况下,得到的结果可信度不高;基

于模板匹配的方法,如由 Gavrilu 等人<sup>[8]</sup>基于轮廓的分层匹配算法,为了解决行人姿态问题,构造了接近 2 500 个轮廓模板对行人进行匹配,采用由粗到细的匹配策略来提高速度;基于统计分类的方法通过对训练数据的学习,得到一个分类器,检测就转化为一个分类问题。当前的行人检测热门方法几乎都属于此类。如 Dalal 等人<sup>[9]</sup>提出了梯度边沿方向直方图(HoG),将图像分不同大小的块,对每一个块,统计其梯度方向直方图,然后再把这些直方图串接起来得到图像的特征,在行人检测中取得了较好的结果。但是由于快的划分需要分多种尺寸,因此最后得到的特征维数通常都比较高,这给分类器的训练提出了挑战,通常的方法是使用降维处理,将高维数据去除相关性转化到低维空间中处理,然后再把得到的低维空间中的数据用于训练分类器。如文献[10]中,将 HoG 和纹理信息特征(基于共生矩阵提取的 12 个描述子)加上 RGB 颜色信息,组合成一个高达 17 万维的特征,然后采用偏最小二乘法进行降维,最后用降维后的数据训练分类器。搜索过程中采用了多尺度搜索,最后得到了较好的效果。但是由于计算量大,耗时多,这些方法目前很难应用到实时的计数系统中。

第 3 类方法采用特征点跟踪,得到点的路径信息,然后对特征点路径信息进行聚类分析实现人数计数,这种处理方法是针对密集人群下的实时计数需求提出的。Vincent 等人<sup>[11]</sup>采用 KLT(Kanade-Lucas-Tomasi)跟踪器获取特征点的轨迹信息,然后对这些轨迹信息使用一些基本规则,如轨迹之间运动相似度,轨迹之间距离等,将点归属于不同的目标,后采用 RANSAC 算法合并得到最终目标数目。Sugimura<sup>[12]</sup>在使用特征点聚类的基础之上,增加了单个个体的步态信息和局部时域一致性特征来更好地将不同的行人之间的特征点划分开。这类方法是当前一种较为新颖的方法,其优点是能够在人群拥挤的情况下实现计数。

## 1 本文方法

### 1.1 系统框架

采用检测和跟踪相结合的方法来实现实时人数计数,跟踪模块使用了 KLT 跟踪器完成对特征点的跟踪,利用已建立的模型,将对点的跟踪结果转化为对行人的跟踪结果。系统流程图如图 1。

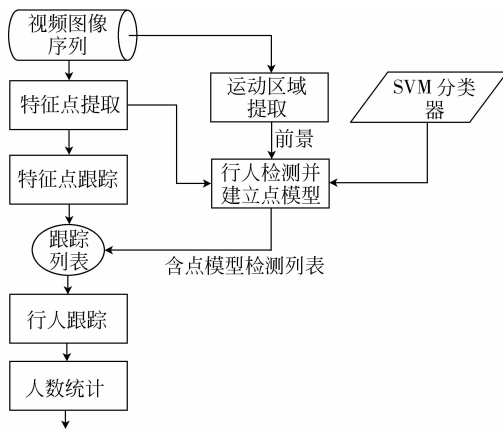


图 1 系统流程图

Fig. 1 System flow chart

在系统中存在两个重要的列表,检测行人列表和跟踪行人列表。在检测模块中,首先对输入的图像,利用帧差法得到运动区域,然后在运动区域上,进行滑动窗口搜索,对窗口用 SVM (support vector machine) 分类器进行判定,判定为行人的窗口,则添加此区域内的特征点并保存下来,当遍历整个运动区域后,需要做一个后处理,将那些属于同一人的多个窗口进行合并,此处的合并也是一个关键,本文利用特征点集合相似度来合并。合并后得到的行人加入检测列表中,并同时将这些检测区域内的特征点作为此区域内行人的一种属性特征一并保存。将检测列表中那些已经存在于跟踪列表中的物体,合并到跟踪列表中,对剩下的物体作为新目标,加入到跟踪列表中。对物体的跟踪,采用一种新颖的方法,利用特征点的跟踪来完成对行人的跟踪。行人区域内的每个特征点以不同的概率归属于该物体,对该物体在下一帧中的位置预测做出不同的贡献。最后,将跟踪列表送到计数模块,使用绊线技术实现计数统计。具体方法将在下面进行描述。

### 1.2 行人检测

本系统中使用 MBLBP 特征,此特征是 Liao 等人<sup>[13]</sup>首先提出的。在原始 LBP 中,其模式的计算是通过比较中心像素的值和其周围相邻像素值来决定,原始 LBP 计算起来非常简单,且有着较好的区分性,但是由于像素的值很可能受到噪声的影响而导致最后的模式不一样,因此 Liao 用其周围点的  $K$  邻域的平均值来代替本身值用于计算最后的模式,思想很简单,但相对于原始 LBP 而言,有着更好的鲁棒性。如图 2 为 LBP 和 3 邻域 MBLBP 的计算。

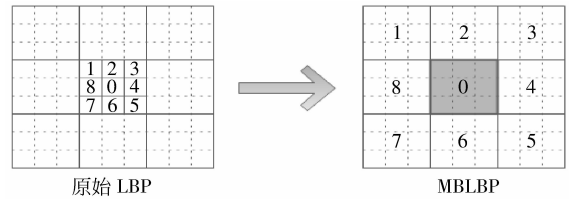


图 2 原始 LBP 和 MBLBP 的计算

Fig. 2 Comparison between original LBP and multi-scale block LBP

对检测的结果进行一个后处理的合并,得到最后的行人目标,然后将每个行人区域内部的特征点作为一种属性,添加到物体特征点属性表中。如果滑动窗口加 SVM 作决策这个过程在每个像素上均进行,则很难做到实时处理,因此最直接的方法是采用运动区域滤掉大量的非运动窗口。系统中,采用了最常用的帧差法,对帧差后的图像做一个二值化,可得到前景的掩模图像。同时,考虑到行人作为前景目标,和背景通常存在着较大的差异性,因此特征点会聚集在行人目标周围,因此可以更进一步地只在存在特征点的运动区域内进行搜索。考虑到实际摄像头角度问题,视频图像中的人数会呈现一种远小近大的现象,因此,增加了一个透视模型,处理这种透视现象带来的影响。在这个透视模型中,最大窗口宽度和最小窗口宽度是由一个人工交互操作来得到的。典型的透视模型中,对透视区域中其他位置的检测框宽度通过线形插值得到。为了减少运算量,将透视区划分为 3 个部分:第 1 部分使用最小宽度作为检测窗口的宽度;第 2 部分使用最小宽度和最大宽度的平均值作为检测窗口的宽度;第 3 部分使用最大宽度作为检测窗口的宽度,如图 3。当检测出行人后,将检测区域内的特征点作为一种特殊属性和检测结果绑定,以便在跟踪阶段加以利用。

### 1.3 行人跟踪

行人计数系统中往往需要对多目标同时进行跟踪。多目标的跟踪存在相互遮挡的问题,因此要实现直接对一个区域较稳定的跟踪,是比较困难的。通常的跟踪方法如卡尔曼跟踪较难实现多目标的跟踪,而粒子滤波的跟踪方法又过于耗时,不便于实际应用。相反,对图像中的特征点进行跟踪,在多目标存在时则相对稳定,而 KLT tracker 就是一个能够完成对点的实时跟踪器。KLT Tracker 首先是由 Tomasi 等人<sup>[14]</sup>于 1981 年作为图像匹配方法提出



图 3 本文中使用的透视模型示意图

Fig. 3 The perspective model used in this paper

的。Tomasi 认为特征点在两帧连续图像中的位置, 可以通过一个仿射变换加一个平移变化得到, 因此, 对点的跟踪可以转化为求解最优的仿射变换矩阵和平移向量, 这种最优是定义在两帧图像的差异上, 即

$$\varepsilon = \int_w [J(Ax + d) - I(x)]^2 \bar{\omega}(x) dx$$

式中,  $I, J$  表示相邻的两帧图像,  $\bar{\omega}(x)$  为一个权重函数, 一般设为一个常数, 或者高斯函数。  $W$  表示积分窗口。上述式子表达的是两幅图像的差异性, 因此问题就是寻找参数使得这个差异性最小。在处理连续的视频图像时, 通常将仿射系数  $A$  忽略掉, 仅仅考虑平移参数。将  $J(Ax + d)$  用泰勒级数展开, 并对  $\varepsilon$  求对  $d$  的导数, 可以得到

$$Zd = e$$

式中  $Z$  是一个  $2 \times 2$  的矩阵, 而  $e$  是一个  $2 \times 1$  的向量, 即

$$Z = \int_w g(x) g^T(x) \bar{\omega}(x) dx$$

$$e = \int_w [I(x) - J(x)] g(x) \bar{\omega}(x) dx$$

$$\text{式中 } g(x) = \begin{bmatrix} \frac{\partial J(x)}{\partial x} \\ \frac{\partial J(x)}{\partial y} \end{bmatrix}.$$

本文中描述的系统, 采用 KLT tracker 来对特征点进行跟踪。在检测过程中, 属于行人检测窗口内的特征点已经加入到此目标属性列表中, 每个特征点均有一个概率, 表明此特征点属于当前行人的置信度。在跟踪过程中, 每个特征点属于此物体的概率在动态地随着此特征点和该物体所有特征点的运

动模式的差异而发生改变, 即

$$P_i = \frac{\alpha e^{-\alpha \tau_i}}{P} \quad i = 1, 2, 3, \dots, k$$

式中,  $k$  表示此物体区域内的特征点个数,  $\alpha$  是一个的衰减速度参数,  $P$  为归一化因子。而  $\tau_i$  这个参数, 反应了此特征点的运动模式与该物体所有特征点的平均值点的运动模式之间的差异。是一个动态变化的参数, 其更新机制如下:

$$\tau_i = \begin{cases} \tau_{i-1} - d_i & d_i < d_{\text{Thld}} \\ \tau_{i-1} + d_i & \text{其他} \end{cases}$$

$d_i$  是该特征点运动矢量和物体所有特征点的平均点运动模式之间的余弦距离。  $d_{\text{Thld}}$  为一个经验值。  $\tau_i$  的初始值在本系统中, 设为 0, 即初始时认为所有特征点以等概率归属于该物体。在跟踪过程中, 由于人体运动的复杂性, 个别特征点的运动可能和行人的运动并非一致, 此时这些特征点的运动矢量就会和整体运动矢量会表现出较大的差异, 运动模式一致的,  $\tau_i$  会减小, 相反则增大, 如果  $\tau_i < 0$ , 则令  $\tau_i = 0$ 。跟踪时, 窗口内的某些特征点可能并不属于此物体, 有可能属于背景, 但在初始时也被添加到物体属性列表中, 对这些特征点, 须要从属性列表中删除。因此, 在跟踪过程中, 如果“归属概率  $p$ ”小于某个预先设定好的阈值后, 则将此特征点从物体属性列表中删除。剩下的特征点, 根据其归属概率, 对物体位置的更新给出不同的加权, 即

$$R_j(x, y) = \sum_{i=1}^k p_i r_i(x, y)$$

式中  $r_i(x, y)$  代表特征点在下一帧中的位置,  $p_i$  为其对应的归属概率。

对单个个体, 利用数学概率模型将对此物体的跟踪转化为过对其内部区域的特征点的跟踪。当多个目标存在时, 检测和跟踪之间还必须存在一个对应, 以避免同一物体被多次作为新的目标添加到跟踪列表中。通常解决此问题的最简单方法是使用最近邻进行匹配, 但是当多目标之间的距离本身就非常的接近, 使用最近邻匹配, 往往会得到错误的结果。利用各个物体的特征点列表, 来快速判断两个窗口的物体是否为同一目标。在检测结束后, 每个物体增加了一个属性, 特征点列表, 该表中包含了此区域的特征点, 判断两个物体的相似性就可以转化为判断两个属性表的相似性。每一个特征点列表其内部的各个值是一些非负整数, 即特征点的序号。

此处关注的是这两个属性表(或称为集合)有多少元素是一样的,在此相似性的定义基础上,提出下面的衡量方法:

$$\text{sim}(\mathbf{A}, \mathbf{B}) = \beta \cdot \frac{N_{\text{con}}}{N_{\mathbf{A}}} + (1 - \beta) \frac{N_{\text{con}}}{N_{\mathbf{B}}} \quad (1)$$

$$N_{\mathbf{A}} = \#(\mathbf{A}), N_{\mathbf{B}} = \#(\mathbf{B}), N_{\text{con}} = \#(\mathbf{A} \cap \mathbf{B}),$$

$$0 < \beta < 1$$

式中,  $\mathbf{A}, \mathbf{B}$  表示两个如上所描述的集合,  $\#$  表示求集合元素个数的一个操作。  $\beta$  是一个参数, 本文描述的系统  $\beta = N_{\mathbf{B}} / (N_{\mathbf{A}} + N_{\mathbf{B}})$ 。此处将  $\beta$  如此取值, 是为了在两个集合元素个数差别比较大时, 能够使上述的相似性偏向于元素较少的集合。在定义完集合的相似性后, 就可以根据两个物体的特征点属性表来判断这两个物体是否是同一目标。

在系统中, 存在着3个重要的目标物体列表, 即检测物体列表、跟踪物体列表和丢失物体列表。检测物体列表和跟踪物体列表, 在上面已经有了其描述。丢失物体列表是为了保存那些在跟踪过程中, 物体由于其严重的被遮挡, 或者其他一些因素, 导致丢失的对象。这些对象, 在下次检测中, 可能被检测出来作为新对象, 因此需要保存这些跟丢的物体。跟踪过程具体做法是, 首先从检测列表中, 取出一个对象, 和跟踪物体列表中的对象依次对比, 利用式(1)计算物体之间相似性, 找到相似性最大的那一个对象, 如果此相似性的取值大于预先设定的一个阈值, 则认为这两个物体是同一个目标, 合并这两个对象, 更新跟踪列表中相应对象的特征点属性表; 如果跟踪列表中没有和检测列表中取出的对象相似的物体, 则将此对象和丢失物体列表中的对象采用最近邻依次比较, 如果能够和丢失物体列表中的某个对象匹配(文中的匹配指是否为同一目标), 则将此对象从丢失物体列表中移送到跟踪列表中; 如果均不能和当前跟踪列表和丢失物体列表中的任何对象匹配, 则将此对象作为一个新出现的目标添加到跟踪列表中。上述过程反复进行, 直到检测列表为空。算法描述如下:

定义:

Deteced\_Object\_List: 检测物体列表

Tracked\_Object\_List: 跟踪物体列表

Missed\_Object\_List: 丢失物体列表

算法:

1) 如果 Deteced\_Object\_List 不为空, 则从中取出一个对象  $\mathbf{A}$ ; 否则, 退出算法。

2) 如果 Tracked\_Object\_List 为空, 则跳转到步骤4) 执行, 否则利用式(1)从 Tracked\_Object\_List 中选取和  $\mathbf{A}$  相似度最大的对象  $\mathbf{B}$ 。

3) 如果  $\text{sim}(\mathbf{A}, \mathbf{B}) > T_1$ , 则合并  $\mathbf{A}$  和  $\mathbf{B}$ , 并更新  $\mathbf{B}$  的特征点列表 ( $T_1$  为一个预先设定的阈值); 跳转到步骤1), 否则执行下一步。

4) 如果 Missed\_Object\_List 为空, 则跳转到步骤6) 执行; 否则采用最近邻原则, 从 Missed\_Object\_List 中选取对象  $\mathbf{C}$ 。

5) 如果  $\mathbf{A}$  和  $\mathbf{C}$  的距离小于一个设定的阈值, 则将  $\mathbf{A}$  和  $\mathbf{C}$  进行合并, 更新  $\mathbf{C}$  的特征点列表, 将  $\mathbf{C}$  插入 Tracked\_Object\_List 中; 跳转到步骤1); 否则执行下一步。

6) 将对象  $\mathbf{A}$  作为一个新目标, 加入到 Tracked\_Object\_List; 转到步骤1) 执行。

## 1.4 计数统计

使用绊线计数(virtual gate)方法, 通过对跟踪物体列表中的对象, 进行绊线检测, 并根据对象的轨迹和检测线的向量叉积来得到人的运动方向信息。这种方法在基于检测加跟踪的系统中, 经常被使用。在文献[15]中, Javier Barandiaran 使用多条检测线分别进行绊线计数, 最后取计数结果最大的, 他指出采用此方法比单独使用一条检测线效果要好, 本文借鉴了这种多线检测的思想, 但同时稍作一些改变, 本文中采用三线计数, 但是并不是将三条线分别进行计数, 而是将这三线联合起来, 在对被跟踪的物体, 增加一个标志位, 初始时设为0, 当任意物体触发任何一条检测线时, 首先判断该标志为, 如过为0, 则发生一次计数, 并将标志为置为1; 否则虽触发绊线检测, 但并不发生计数。

## 2 实验

### 2.1 训练样本

目前并没有统一的人数计数公开数据库。公开的行人检测数据库, 其拍摄角度一般是正面, 摄像头和水平线之间近似平行, 而人数计数在实际应用中, 安装的摄像头一般都和地平面(水平线)存在一定的角度, 因此用公开的行人检测库进行训练并不适合, 故在实验中, 使用自己实际应用而采取的数据。针对两个不同的场景, 使用了两段不同的视频。第一段视频, 总长度为30 min, 将前20 min 的视频进行手动标注, 采集样本作为训练数据; 第二段视频, 由

于人数较多,故只采用了 10 min 的视频,前 5 min 的数据用作训练,后 5 min 的视频用于测试。对于训练数据,所有的负样本是在图像上,除掉正样本标注所在区域外,随机产生的。其中正样本共 3 000 张,而负样本 30 000 张,将所有的图像归一化到 $48 \times 96$ 的大小。部分样本如图 4。分类器的训练是采用公开的 libsvm 开发包完成的。其中核函数选择线性核函数,并对正样本系数进行一个 10 倍的加权。同时为了作对比,对一段其他论文中公开的视频,进行了测试,针对该视频没有采集任何与其相关的训练数据。

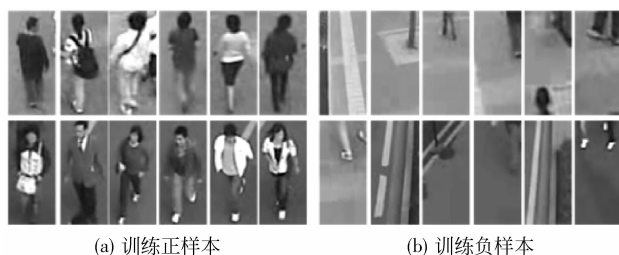


图 4 训练数据部分样本

Fig. 4 Some of the training samples

为了验证本系统中所使用的特征,具有较好的区分性,分别选取一张正样本和负样本图像,将其特征用直方图形式表达,原始 MBLBP 是 256 维的,此处为了达到一个更好的可视效果,将特征显示为 32 维的,见图 5。从图中可以看出,此特征能够较好地区分行人和其他背景物体。

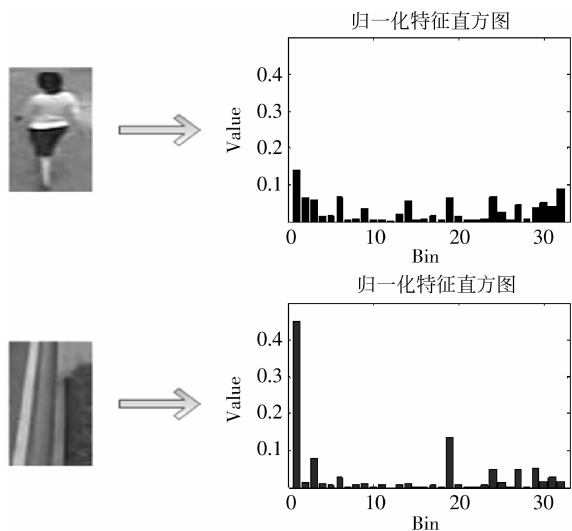


图 5 特征的直方图显示

Fig. 5 The feature displayed as histogram

## 2.2 计数实验

在实验中,分别测试了两段不同场景下拍摄的视频 Video A 和 Video B。视频大小为 $355 \times 288$ ,帧率为 25 帧/s。Video A 为一个路道的监控,视频中,行人朝两个方向行走,相对于摄像头来说,称为“进入方向”和“出去方向”。Video B 是一个入口安检棚,因此在这个视频中,人流主方向只有一个,称之为“进入方向”。每一个实验的相关参数见表 1。系统需要交互设置感兴趣区域和最大搜索窗口尺寸与最小搜索窗口尺寸(尺寸用宽 $\times$ 高来表示),此处用( $BW, SW$ )分别来代表。同时,滑动窗口仅在运动区域上进行搜索,因此需要设置一个阈值,使得当搜索窗内的前景在此窗口所占的面积大于该阈值,则在此窗口做分类判断。阈值用  $T$  来表示。

表 1 测试视频及相关实验参数

Tab. 1 Experiment parameters for testing videos

视频	( $BW, SW$ )	$T$	视频帧数
Video A	( $28 \times 56, 18 \times 36$ )	0.4	14 775
Video B	( $28 \times 56, 24 \times 48$ )	0.33	7 500

目前,针对人数计数,还没有一个公认的评价标准,不同于行人检测(注重检出率和误检率),人数计数更加注重于计数结果,因此,实验中直接采用对系统所得人数个数和真实的人数个数进行对比。为了进一步地反映在各个时刻,系统均表现出良好的性能,实验中将每隔 400 帧(约 16 s)进行一次统计,将统计结果以曲线的方式表示出来,如图 6。图 6(a)(c)给出了 Video A 和 Video B 的系统计数结果和真实数据的对比。真实数据是由人工统计得到的,统计各个时间段的计数标准为:行人中心压线则计数。同时在图 6(b)(d)中给出了两段视频按时间段进行计数统计和真实数据之间的偏移图(此处的偏移表示系统给出此时间段计数的结果减去由人工确定此时间段的行人个数)。偏移量可正可负,当偏移量为正数时,表明“计多了”,这种情况的发生,一般是在行人本身带有其他运动物体,这些物体和地上的阴影被检测为了行人并计数。当偏移量为负,表明此时间段“计少了”,这种情况一般是因为漏检了一些行人导致的。这些现象见图 7。部分跟踪结果见图 8,其中(a)~(c)对应于 Video A, (d)~(f)对应于 Video B。

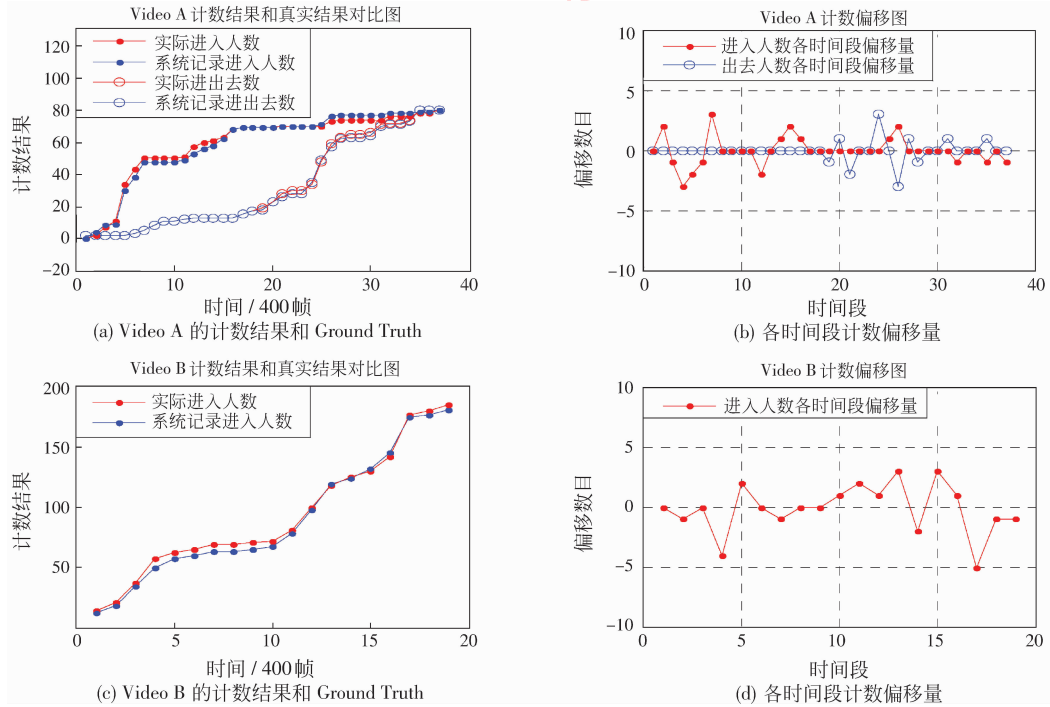


图 6 实验结果

Fig. 6 Experimental Results

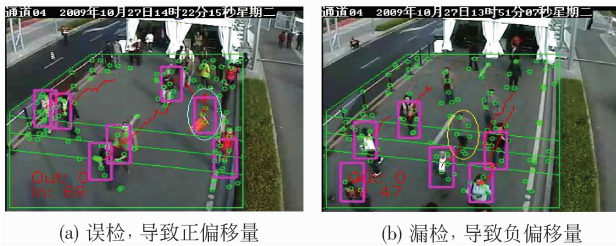


图 7 误检和漏检说明

Fig. 7 Images of false-positive detection and miss-detection

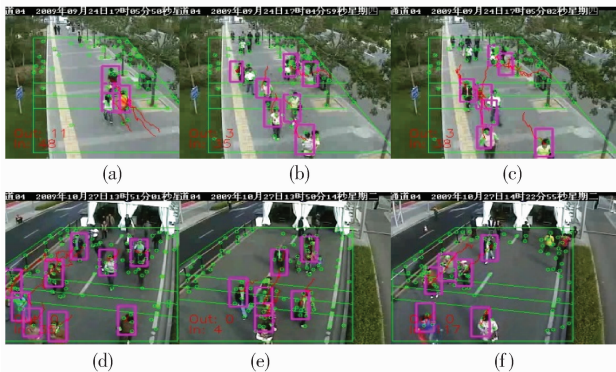


图 8 一些在 Video A 和 Video B 上计数测试结果

Fig. 8 Experimental results on video A and B

为了和其他论文中的计数做出对比,本文对文献 [16]中使用的视频,进行了实验。但并没有针对该段

视频增加额外的训练样本,直接使用已有的系统对该段视频进行测试,而其参数直接使用 Video A 的参数(实际上,使用 Video B 的参数得到了同样的结果)。该段视频共长 908 帧,进入人数为 19,出去人数为 8。针对该视频,为了更好的和其他论文做出一个对比,实验中直接将感兴趣区域设置为全图。检测线设置在中间,统计人数按压线原则进行。针对开始已经过线的人数不做统计,如图 9。系统计数结果为 19,9。在“出去方向”上多计了一个,这个可以通过图 9 (d)看出检测时,椭圆中的人,其下半身和树枝组合在一块被检测为行人进行跟踪,最后压线计数。

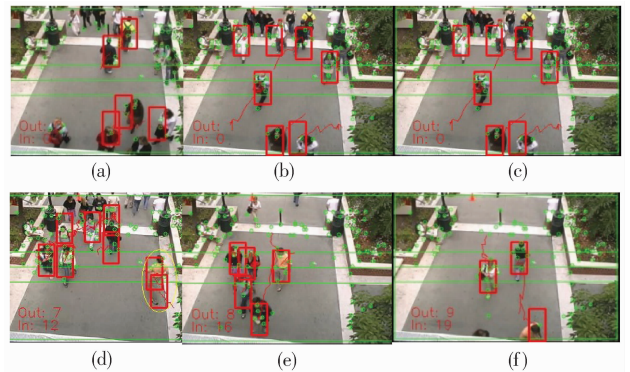


图 9 在公开视频上的计数测试

Fig. 9 Experimental results on public video

实验结果表明,本文描述的计数系统有着较好的性能。由于整个流程中,使用了简单且具有较好区分性的特征,使得特征的提取耗时非常小,使得检测非常迅速;由于视频图像是连续的,对特征点的跟踪,不需要采用仿射变换进行计算,大大减小了运算量,使得跟踪非常的迅速。特征点的检测是和行人检测同步,并且和跟踪交替进行。整个系统的运行时间和实际场景中的人数均有关系,根据测试,在实验的电脑上(电脑基本配置: Intel Pentium Dual 3.00 GHz,内存 2 G),处理标准的 CIF (352 × 288) 格式视频,每一帧大约需要 10 ms。因此能够运用到实际中去。

### 3 结 论

描述了一个实时计数系统。采用典型的检测加跟踪来实现人数计数的目的。在检测阶段,使用运算速度快的 MBLBP 特征来检测行人;在跟踪阶段,通过一个数学模型,将对目标行人的跟踪转化为对特征点的跟踪,并且在检测目标和跟踪目标进行一一对应时,再次利用特征点来进行对应,实验结果表明,该系统具有较好的鲁棒性,能够在不同的场景下实现较准确的人数计数功能。

### 参考文献 (References)

- [ 1 ] Kim Byeoung su, Lee Gwang Gook, Yoon Ja Young. A method of counting pedestrian in crowded scenes [ C ] // Proc. of the 4th International Conference on Intelligent Computing. Berlin, German: Springer, 2008: 1117 - 1126.
- [ 2 ] Lee Gwang Gook, Kim Byeoung su, Kim Whoi Yul. Automatic Estimation of Pedestrian Flow [ EB/OL ]. ( 2007-10-22 ) [ 2010-04-12 ]. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4357536](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4357536).
- [ 3 ] Chan A B, Liang Z -S J, Vasconcelos N. Privacy Preserving Crowded Monitoring: Counting People without People Models or Tracking [ EB/OL ]. ( 1008-05-08 ) [ 2010-04-12 ]. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4587569](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4587569).
- [ 4 ] Zhang Enwei, Chen Feng. A fast Robust people counting method in video surveillance [ C ] // Proc. of International Conference on Computational Intelligence and Security, Washington DC, USA: IEEE Computer Society, 2007: 339-343.
- [ 5 ] Ma Honglian, Lu Huchuan, Zhang Mingxiu. A real-time effective system for tracking passing people using a single camera [ C ] // Proc. of World Congress on Publication Intelligent Control and Automation. Washington DC, USA: IEEE Computer Society, 2008: 6173-6177.
- [ 6 ] Hsieh Jun Wei, Peng Chengshuang, Fan Kao Chin. Grid-based template matching for people counting [ C ] // Proc. of IEEE 9th Workshop on Multimedia Signal Processing. Washington DC, USA: IEEE Computer Society, 2007: 316-319.
- [ 7 ] Li Min, Zhang Zhaoxiang, Huang Kaiqi, et al. Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection [ C ] // Proc. of IEEE 19th International Conference on Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2008: 1-4.
- [ 8 ] Gavrila DM. Pedestrian detection from a moving vehicle [ C ] // Proceedings of the 6th European Conference on Computer Vision. Berlin, German: Springer, 2000: 37-49.
- [ 9 ] Dalal Navneet, Triggs Bill. Histograms of oriented gradients for human detection [ C ] // Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2005: 886-893.
- [ 10 ] William Robson Schwarz, Aniruddha Kembhavi, David Harwood, et al. Human detection using partial least squares analysis [ C ] // Proc. of IEEE International Conference on Computer Vision. Washington DC, USA: IEEE Computer Society, 2009.
- [ 11 ] Vincent Rabaud, Serge Belongie. Counting crowded moving objects [ C ] // Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2006: 705-711.
- [ 12 ] Sugimura Daisuke, Kitani Kris M. Using individuality to track individuals: clustering individual trajectories in crowds using local appearance and frequency trait [ C ] // Proc. of IEEE International Conference on Computer Vision. Washington DC, USA: IEEE Computer Society, 2009.
- [ 13 ] Liao Shengcai, Zhu Xiangxin, Stan Z Li. Learning multi-scale block local binary pattern for face recognition [ C ] // Proc. of International Conference on Biometrics. Berlin, German: Springer, 2007: 828-837.
- [ 14 ] Tomasi Carlo, Kanade Takeo. Detection and Tracking of Points Features, Technical Report CMU-CS-91-132 [ R ]. Pittsburgh: Carnegie Mellon University, 1991.
- [ 15 ] Javier Barandiaran, Fernando Boto. Real-time people counting using multiple lines [ C ] // Proc. of International Workshop on Image Analysis for Multimedia Interactive. Wahington DC, USA: IEEE Computer Society, 2008: 159-162.
- [ 16 ] Brostow G J, Cipolla R. Unsupervised bayesian detection of independent motion in crowds [ C ] // Proc. of IEEE Conference on Computer Vision and Pattern Recognition. Washington DC, USA: IEEE Computer Society, 2006: 594-601.