

# 嵌入式系统中实时行人检测方法<sup>①</sup>

康伟德, 高通

(内蒙古大学 电子信息工程学院, 呼和浩特 010021)

通讯作者: 高通, E-mail: [super369tong@163.com](mailto:super369tong@163.com)

**摘要:** 随着计算机技术的不断发展, 智能自动化的人数检测系统不断产生. 人数检测对于企业或机构的信息化管理至关重要. 传统人数检测方法因为肢体遮挡以及光照变化导致准确率较低. 提出了针对人头特征的垂直检测方法, 该特征可以保证在人流密度大的情况下无法被遮挡. 该方法首先提取前景图像的梯度方向直方图特征, 并通过 SVM 检测人头目标, 利用头部的颜色特征, 在相邻帧中使用 MeanShift 算法跟踪人头目标. 根据人头目标轨迹进行过线检测, 算法在嵌入式系统上进行了应用与测试, 实验表明算法有较好的实时性与准确率.

**关键词:** 行人检测; HoG 特征; MeanShift 算法; 嵌入式系统

引用格式: 康伟德, 高通. 嵌入式系统中实时行人检测方法. 计算机系统应用, 2018, 27(4): 94-99. <http://www.c-s-a.org.cn/1003-3254/6304.html>

## Real-Time Pedestrian Detection Method in Embedded System

KANG Wei-De, GAO Tong

(College of Electronic Information Engineering, Inner Mongolia University, Hohhot 010021, China)

**Abstract:** With the development of computer technology, the intelligent and automatic pedestrian detecting system continues to emerge. The pedestrian detecting system is an important part of the information management for enterprises and institutions. The traditional method has a lower accuracy due to limb occlusion and light changes. This study proposes a vertical detection method for the head features, which can ensure that the features cannot be blocked even in high flow density. Firstly, the gradient histogram feature of the foreground image is extracted and the head target is detected by SVM. The MeanShift algorithm is used to track the head target in the adjacent frame by using the color feature of the head. The pedestrian is detected according to the target track. The algorithm is applied and tested in the embedded system. The experiment results show that the method is effective and accurate.

**Key words:** pedestrian detection; HoG feature; MeanShift algorithm; embedded system

计算机视觉是一门多领域的交叉学科, 它与图像处理、计算机图形学、数学、信号处理、模式识别、心理学、物理学、人工智能息息相关. 人数统计对于城市信息化建设起着重要推动作用, 对于资源调配、数据分析都有巨大意义<sup>[1]</sup>.

基于视频图像人数统计分为区域式系统与断面式系统, 区域式系统用来统计目标区域的总人数, 而断面式系统用于统计所有进或出的人数. 本文所研究的对

象是断面式人数统计. 利用摄像头垂直拍摄, 对视频中通过断面的人数进行统计.

早期的人数统计利用红外线阻断, 该方法在单个目标时很精确, 但对于连续目标时会误判为一个目标, 这是由于其检测机制决定的. 对于高人流密度或是需要准确人数信息的场景中, 应该采用更高级的方法. 在视频图像人数统计算法中主要分为针对特征点、颜色与形状、模板匹配等方法.

<sup>①</sup> 收稿时间: 2017-07-29; 修改时间: 2017-08-17; 采用时间: 2017-08-22; csa 在线出版时间: 2018-03-31

Kim<sup>[2]</sup>等人利用图像的低层特征:前景像素积累与运动向量来进行人数统计,针对人数较多的情况,采用像素累计和的加权比例估算人数,达到了95%的准确度,但未进行行人与其他目标的区分,当前景图像存在大量非行人目标时,准确度大大降低。

Sugimura<sup>[3]</sup>等人提出一种在人群中跟踪个体的方法。在基于特征点聚类的方法上,增加了单个个体的步态信息,更好的区分行人,尽管行人之间存在部分遮挡,也在相对拥挤的环境取得较好的结果。

Dapeng Luo<sup>[4]</sup>等人提出了一种在视频图像中检测头部的在线学习框架,基于在线 boosting 算法,将检测到的对象作为新的样本训练分类器,并利用粒子滤波得到新样本的权重并标记样本。该方法减少了对原始样本的依赖,并可以在检测中渐渐提高检测精度。

本文所设计方法为了在人流密度下人保持较高检测精度,使用运动目标中的人头特征,该特征满足在人流密度仍能不遮挡。首先利用梯度方向直方图特征提取前景图像中的特征,利用 SVM 检测前景图像中的人头,并利用人头的颜色特征利用 MeanShift 算法在相邻帧中跟踪人头目标,并对跟踪到的人头进行标记,最后对人头移动轨迹进行过先检测。该方法保证了多个行人前景团块聚合时仍然可以识别单个个体,保证了人流密度仍具有较高准确度,同时使用前景图像作为感兴趣区域 (ROI) 提取疑似目标,减少了虚警的产生,并且加快了检测效率。最后,本文所述方法在嵌入式系统进行了实验测试,实验显示本文算法具有较高准确率与实时性。

## 1 背景建模与前景检测

### 1.1 背景模型的建立

嵌入式系统的计算性能较于个人计算机仍具有一定差距,实时人数统计要求算法具有较少的计算量与一定得准确度。为了检测到视频图像的行人目标,可以使用滑动窗口对窗口内是否为行人进行判断,但该方法的执行效率极低,不满足实时性的要求,同时容易在视觉显著性较低的背景中带来虚警的干扰。为了在图像中提取疑似区域,本文采用背景模型前景检测的方法。

常用的背景模型有背景差分、编码本、单高斯背景模型以及混合高斯背景模型。在视频图像中背景并非一程不变的,光照、阴影、气流、传感器噪声都会使背景发生变化,背景差分法等简单的背景模型达不

到这种准确度要求。人群移动过程中会造成人影以及气流的变化,高斯背景模型<sup>[5]</sup>利用视频图像像素点的时间序列分布,并用高斯函数逼近像素序列分布,而背景模型的性能取决于高斯函数对像素序列的逼近情况,对于单高斯背景模型无法达到复杂环境中背景模型的性能要求。混合高斯背景模型使用数个不同权值的高斯函数逼近训练过程的像素序列分布,而对于新产生的像素,可以通过假设检验验证该像素是否属于背景,若该像素出现的概率为一小概率事件,可以认为该像素不属于背景,为前景像素。具体定义见公式(1)。

$$\begin{cases} P(x_t) = \sum_{i=1}^k w_{k,t} * \eta(x_t, \mu_{i,t}, \Sigma_{i,t}) \\ \eta(x_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{i,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - \mu_{i,t})^T \Sigma_{i,t}^{-1} (x_t - \mu_{i,t})} \end{cases} \quad (1)$$

该式中  $\eta$  为高斯分布,  $w$  为高斯分布的权重。让概率分布不断逼近背景分布。前景、背景像素的检测通过计算像素在任意高斯分布的概率最大值是否大于阈值。高斯背景模型的参数更新见公式(2)。

$$\begin{cases} w_{i,t} = (1-a)w_{i,t-1} + a \\ \mu_{i,t} = (1-p)\mu_{i,t-1} + px_t \\ \sigma_{i,t}^2 = (1-p)\sigma_{i,t-1}^2 + p(x_t - \mu_{i,t})^2 \end{cases} \quad (2)$$

式中  $a$  为自定义学习率,  $p$  为参数学习率,  $a$  值越大说明背景更新越快。  $P$  值由自定义学习率  $w$  与权值  $w$  共同决定。高斯分布个数选择上可以根据不同实验环境特征进行修正。一般情况3个高斯函数可以保证背景模型的准确性,本算法使用3个高斯函数构建混合高斯背景模型。

混合高斯背景模型需要耗费大量计算资源,为了增强实时处理速度,程序在执行开始的100帧建立背景模型,随后在每个间隔帧中更新背景模型。

### 1.2 团块提取与 ROI 区域建立

视频图像中运动的行人目标,可以认为是前景目标,即为新进入摄像区域的目标。该图像通过原始图像与前景掩模图像计算获得,可以看到单纯通过前景掩模图像难以获得完整的团块,原因主要是前景检测伴随小的因气流变化或阴影等原因产生的噪声干扰,以及由于人的发色较一致,在移动过程中因为背景更新的缘故导致头部前景图像具有孔洞。为了完整获取

ROI 区域, 本文提出如下方法改善团块质量.

1) 对团块进行面积滤波, 设置阈值  $th1$ , 滤除较小面积, 即不可能是行人目标的噪声干扰.

2) 为了消除背景模型更新导致团块出现的孔洞 (比如头部), 使用形态学孔洞填充法对团块进行孔洞填充.

3) 使用外接矩形逼近团块建立 ROI 区域.

图 1 为提取视频图像中一帧的处理过程, 首先进行前景团块提取, 通过形态学处理后 (面积滤波、孔洞填充) 用外界矩形圈出, 该区域是进行人头检测的目标区域.

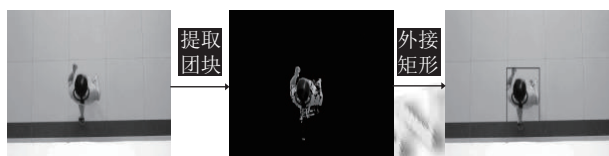


图 1 团块提取与外接矩形 ROI 建立

## 2 基于梯度方向直方图的人头检测

在单个行人通过摄像头监视区域时, 团块中只含有一个目标, 这时只要对团块跟踪计数即可获得很高的识别准确率. 但是在人流过于集中的条件下, 不同的行人有可能因为相互接触而聚集在同一团块, 这样单独对团块进行跟踪无法保证人数统计的正确性. 由于摄像头垂直监视断面时, 过大的人流密度仍然不会导致人头的重叠. 所以本文提出对人头目标检测跟踪的人数统计方法.

梯度方向直方图<sup>[6]</sup>(HoG) 特征用来对局部区域的梯度进行计算, 并用直方图来描述, 它是一种图像局部重叠区域的特征描述子. 由于其利用局部区域的梯度直方图进行特征描述, 所以能够在较小的空间域有效克服图像的光学形变与几何形变. 人的头部具有很多共性特点, 明显的颜色特征与弧度特征, 但个体之间由于拍摄角度仍具有一定差异, 对于 HoG 特征描述子来说对于微小位移和光照变化具有一定不变性, 适合人数统计系统的使用.

HoG 特征提取过程如下:

1) 归一化图像, 并计算梯度.

2) 将图像分为若干 cell(单元), 若干 cell 组合形成一个 block(块).

3) 通过将梯度方向分为若干等份, 对 cell 进行梯度分布直方图计算, 将 block 内的 cell 直方图进行归一化.

4) 移动 block, 遍历整个图像, 将所有 block 内的直方图特征组合成一个大的 HoG 向量.

HoG 特征提取较为方便, 随即需要对分类器进行训练. 本文实验样本有限, 同时在实际应用过程中, 人数统计的对象多为陌生人. 在这种训练样本远远不足的情况下适于使用结构风险最小化的支持向量机 (SVM) 分类器, 神经网络的准确性取决于样本数趋于无穷的渐进理论, 支持向量机通过约束条件, 将分类超平面置于最大类间隔内, 适合在小样本使用<sup>[7-10]</sup>.

实验通过在摄像头拍摄到的图像中选取样本. 正样本即为人头图像, 负样本即为非人头图像. 由于本文算法的检测范围是前景检测提取到的 ROI 区域, 因此负样本多为人体的其他位置, 以及人体相邻的背景. 部分训练样本见图 2.



图 2 部分训练样本

HoG 特征选取的窗口大小为  $96 \times 96$ , cell 单元大小为  $8 \times 8$ , block 大小为  $32 \times 32$ , block 的滑动距离为 8. 按照这种方式生成了 900 维特征. 支持向量机的核函数选用径向基函数, 利用训练完成的图像进行测试, 使用滑动窗口的方法在 ROI 上检测人头目标, ROI 区域滑动步长为 4. 检测结果容易出现检测窗口的重叠, 对此现象需要对检测窗口进行聚合. 本文提出如下方法聚合检测窗口.

1) 以最高 SVM 分数的一半作为阈值, 抑制部分重叠窗口.

2) 计算剩余窗口的相交矩形, 并计算矩形中心.

3) 选择窗口中心距离矩形中心最近的点为目标窗口, 并抑制其余窗口.

按照本方法的检测效果见图 3.

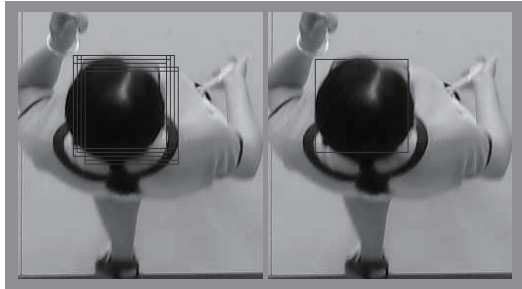


图 3 滑动窗口检测与窗口聚合

### 3 基于 MeanShift 的目标跟踪与过线检测

利用在团块 ROI 区域检测到的人头目标, 随后需要在相邻帧中进行人头目标的跟踪. 常用的方法有特征点检测 (例如: SIFT)、模板匹配、光流法、MeanShift 算法. 特征点检测算法通常利用较稳定的局部奇异点, 这种方法一般鲁棒性较强, 但不适于本文使用, 因为在人头目标中, 大部分为头部平滑的发色, 特征点数量有限. 不适合使用点追踪的方法. MeanShift 属于非参数密度估计, 具有较高的实时性, Fukunage 和 Hosteler<sup>[1]</sup> 等人在 1975 年的首次提出了 MeanShift 算法. 该算法是一种基于核密度估计的算法. 从一个起始点通过向核密度增加的方向不断的偏移迭代收敛到概率密度极大值处.

空间内的核密度估计可以写成如公式 (3) 形式.

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \quad (3)$$

利用核密度估计的表达式可以得到概率密度极大值点, 对上式求导并置 0, 得到概率密度的极值点  $x$ . 并可求得 MeanShift 向量, 它指向概率密度梯度方向. 核密度的导函数与 MeanShift 向量见公式 (4).

MeanShift 算法需要以特征直方图的形式对跟踪目标进行模型化处理, 特征模型可以是颜色、纹理、轮廓等信息. 因为人头部具有鲜明的色彩信息, 易于与背景相区分, 所以本文使用颜色目标特征模型.

$$f'(x) = \frac{2}{nh^{d+2}} \left[ \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right] \left[ \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right]$$

$$M_h(x) = \frac{\left[ \sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right]}{\left[ \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right]} - x \quad (4)$$

为了减弱光照的影响, 将人头目标的色彩空间转换为 HSV 色彩模型. 多数人的头发为黑色, 饱和度较低. 本文算法采用人头的饱和度概率分布, 统计饱和和 9 个区间的概率密度, 随后建立下一帧视频图像的反向投影图搜索头部目标, 目标的反向投影图见图 4.

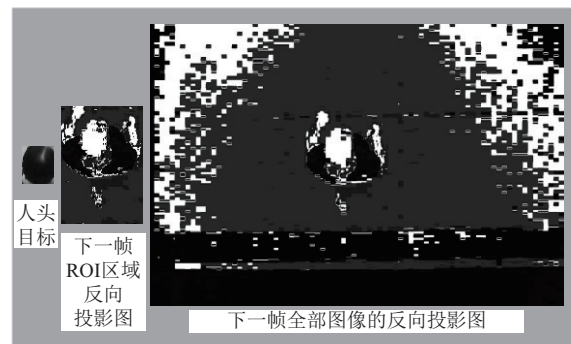


图 4 人头目标与相应反向投影图

因为人头目标在相邻帧中运动距离有限, 所以 MeanShift 算法以上一帧的位置作为迭代起点即可快速收敛到目标位置. 通过判断 MeanShift 算法的收敛位置与下一帧的人头检测位置, 若两者位置小于阈值, 可以认为该人头目标是上一帧人头目标的运动结果, 据此可以生成运动轨迹, 采集图像大小为 1280×720 分辨率, 在 400 像素高度处设置断面层, 对通过断面的目标进行计数, 以此来实现人数统计. 检测效果见图 5.

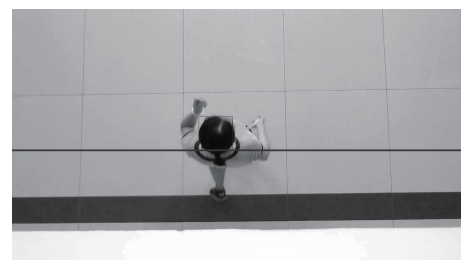


图 5 过线检测效果

### 4 算法软件设计与结果分析

本文所使用的硬件为三星公司的 S5PV210, 其内核为 Cortex-A8 架构, 主频 1 GHz. 摄像头为 300 万像素 USB 摄像头. 首先在嵌入式环境中移植 Linux 系统、OpenCV 3.0、Qt 4.8. 这为实时图像处理与图形界面显示创造了先决条件.

S5PV210 拥有 1 GHz 的主频, 并且支持 2D/3D 图形加速. 具有一定速度的图像处理能力, 可以完成本算法所需计算量. 300 万像素的 USB 摄像头, 图像清晰度具备人数统计的要求. Linux 内核已经自带该摄像头的驱动程序, 只需利用 V4l 的系统调用可以快速的采集图像. 本系统使用 OpenCV 3.0, 具备充足的中高层函数接口, 简化图像处理实用化的复杂度. 尤其是 3.0 时代甚至使用一些与零门槛入门的 Matlab 相同的函数, 大大简化了图像处理开发复杂度. 在图形用户界面使用 Qt 4.8, 可以将人数信息显示到显示屏幕上.

按照本文所述算法, 建立的软件设计流程见图 6.

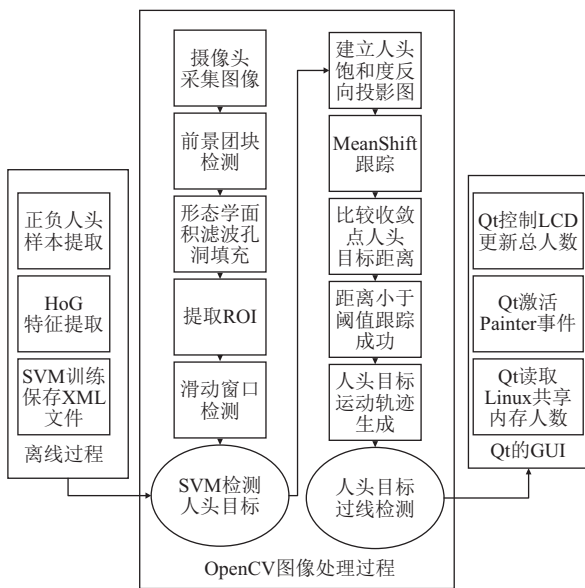


图 6 软件设计流程

在人头目标通过检测线时, 将 Linux 共享内存的数值加 1, 共享内存是进程通信的常用方法, Qt 读取共享内存的数据并将新的人数通过重画事件重新显示到屏幕上. 实验测试效果图见图 7.

通过本文所述试验方法进行测试, 分别采集图像在实验室门口与教学楼大厅, 得到实验结果见表 1. 设计中自定义记录人头相关信息的类 head\_info 类, 类中

使用 OpenCV 提供的 Rect 类记录人头目标的位置信息, 并用 Mat 类记录上一帧的人头目标图像, 方便在下一帧搜索, 将 head\_info 对象存储到 Vector 容器中, Vector 由 C++ 标准模版库提供, 是一种容器类, 用于存储某种数据结构. 通过在下一帧中跟踪上一帧所有的人头目标, 利用 MeanShift 收敛点与新检测到的目标位置比较, 生成目标轨迹, 并对通过断面层的目标进行计数.



图 7 实验测试环境

表 1 实验测试结果

测试环境	进入人数	检测人数	准确率(%)
实验室	65	64	98.5
教学楼大厅	74	65	87.8

由表 1 可见, 实验室测试效果较佳, 原因是实验室门小, 多为单人出入, 检测到的团块含有单个行人, 对于人头跟踪不容易受到其他人头的干扰. 教学楼大厅的人数较密集, 容易出现团块内含有多个目标的状况, 这种情况下算法需要较大的计算量, 致使处理一帧时间较长, 这使得目标在下一帧距离较大, 目标跟踪效果容易受到影响, 出现目标丢失的现象, 考虑调整摄像头采集图像分辨率并优化算法, 进一步提高准确率.

### 5 结语

本文提出了基于人头特征的实时人数统计方法, 该特征能够在高人流密度下不发生遮挡. 为了有效抑制人头检测虚警与加强算法执行速度, 现对视频图像提取前景, 并采用了形态学滤波等方法提取完整的 ROI, 为了得到行人的轨迹, 利用人头饱和度的特征在相邻帧中追踪目标. 最后算法进行了嵌入式系统的设计与实验, 结果表明该方案是切实可行, 同时具备较高准确率.

#### 参考文献

1 孙锐, 陈军, 高隽. 基于显著性检测与 HOG-NMF 特征的快

- 速行人检测方法. 电子与信息学报, 2013, 35(8): 1921–1926.
- 2 Kim BS, Lee GG, Yoon JY, *et al.* A method of counting pedestrians in crowded scenes. Proceedings of the 4th International Conference on Intelligent Computing: Advanced Intelligent Computing Theories and Applications with Aspects of Artificial Intelligence. Shanghai, China. 2008. 1117–1126.
  - 3 Sugimura D, Kitani KM, Okabe T, *et al.* Using individuality to track individuals: Clustering individual trajectories in crowds using local appearance and frequency trait. 2009 IEEE 12th International Conference on Computer Vision. Kyoto, Japan. 2009. 1467–1474.
  - 4 Luo DP, Sang N, Huang R, *et al.* A novel online learning method for head detection in video sequences. Proceedings of Visual Information Processing XIX. Orlando, FL, USA. 2010. 77010Q.
  - 5 Suo P, Wang YJ. An improved adaptive background modeling algorithm based on Gaussian mixture model. 9th International Conference on Signal Processing. Beijing, China. 2008. 1436–1439.
  - 6 Rossi M, Bozzoli A. Tracking and counting moving people. IEEE International Conference on Image Processing. Austin, TX, USA. 1994. 212–216.
  - 7 Ding M, Fan GL. Generalized sum of Gaussians for real-time human pose tracking from a single depth sensor. 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA. 2015. 47–54.
  - 8 Geng YY, Liang RZ, Li WZ, *et al.* Learning convolutional neural network to maximize Pos@Top performance measure. arXiv: 1609.08417, 2017.
  - 9 Tan M, Hu ZF, Wang BY, *et al.* Robust object recognition via weakly supervised metric and template learning. Neurocomputing, 2016, 181: 96–107. [doi: [10.1016/j.neucom.2015.04.123](https://doi.org/10.1016/j.neucom.2015.04.123)]
  - 10 Gerónimo D, López AM, Sappa AD, *et al.* Survey of pedestrian detection for advanced driver assistance systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(7): 1239–1258. [doi: [10.1109/TPAMI.2009.122](https://doi.org/10.1109/TPAMI.2009.122)]
  - 11 Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Transactions on Information Theory, 1975, 21(1): 32–40. [doi: [10.1109/TIT.1975.1055330](https://doi.org/10.1109/TIT.1975.1055330)]