

# 基于多目标检测与跟踪的密集客流检测

祖克举 刘富强 李志鹏

(同济大学嵌入式系统与服务计算教育部重点实验室, 上海 201804)

**摘要** 针对密集交通场景中的客流检测问题,提出了基于支持向量机(SVM)多目标检测与 Mean Shift 跟踪相结合的方法。首先采用自适应检测窗口提取梯度方向直方图,经过 SVM 分类和聚类算法,得到头部图像初始假设。然后采用 Mean Shift 算法,对头部假设进行跟踪,得到连续的头部图像序列。通过 SVM 分类器对序列图像进行整体判断,得到客流信息。实验结果表明,自适应滑动窗口的方法减少了特征提取阶段的处理时间,提高了检测速度;同时,通过对得到的跟踪序列进行整体判别,客流量的检测精度得到了提高。

**关键词** 梯度方向直方图 支持向量机 Mean Shift 跟踪 客流计数

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2009)10-2106-04

## Counting People Based on Multiple Targets Detection and Tracking

ZU Ke-ju, LIU Fu-qiang, LI Zhi-peng

(The Key Laboratory of Embedded System and Service Computing Supported by Ministry of Education, Tongji University, Shanghai 201804)

**Abstract** In this paper we proposed a method for determining the passenger flux based on SVM (support vector machine) detection and Mean Shift tracker. With adaptive detection window, the histogram of the gradient orientations is extracted through the detection region. After classification and clustering, the initial head hypotheses are obtained. Then, the Mean Shift tracker is used to track them, and image sequences of the head are achieved. By the whole decision on the consecutive head sequences using SVM, the number of people is obtained. Experimental results show that the adaptive sliding window method reduces the time consumption, and the accuracy of detection is improved by the combination with head tracking.

**Keywords** histogram of the gradient orientations, SVM (support vector machine), Mean Shift tracking, counting people

## 1 引言

视频监控在公共安全管理与安全防范中起着重要的作用,特别在地铁、机场等公共交通场所的进出口处,需要对监控视频进行处理,检测监控区域内的客流信息,为管理部门提供可靠的数据,辅助进行交通管理,同时也可以作为规划交通设施的重要参考依据。然而,在这些交通场景中客流密集,给视频图像处理技术的应用带来了极大的挑战。主要难点在于:监控视频质量差、目标尺度变化较大,以及行人

间的相互遮挡。因此在客流密集的情况下,难以提取判别性的特征来描述目标。

目前,国内外的文献中提出不同的方法估计客流密度或者检测客流量。比如在文献[1]中提取检测区域内的图像特征来对密度的等级进行分类,最后给出客流的密集程度。然而这些方法只能粗略估计密度等级,不能得到准确的人数信息。文献[2]、[3]则通过对单个行人的检测或分割,检测区域内的行人数目。然而它们对摄像机的高度和角度都有一定的要求,需要采用侧面或者顶部进行安装摄像机。文献[4]中采用特征跟踪和聚类的方法,实现

基金项目:上海市自然科学基金(08ZR1420600)

收稿日期:2009-07-15; 改回日期:2009-07-28

第一作者简介:祖克举(1976~),男。同济大学控制理论与控制工程专业博士研究生。主要研究方向为数字图像处理、计算机视觉、多媒体信息处理。E-mail:zukeju@126.com

了密集客流的计数,但是当人群具有相同的运动方向时,也不能给出满意的结果。

针对密集地铁监控场景的应用,文中采用自适应检测窗口对检测区域进行搜索,提取梯度方向直方图,通过支持向量机(SVM)分类器进行分类,从而得到大量的行人头部图像假设。然后对检测到的头部假设进行跟踪,得到连续的头部图像序列,通过分类器进行整体判别提高客流检测精确。

## 2 基于线性 SVM 的头部检测

在地铁等密集交通场景中,行人之间存在严重的遮挡,只有部分身体部位可见。对于这类场景,背景建模或常用的行人检测算法不能得到满意的检测结果。本文结合梯度方向直方图和 SVM 分类器的方法,检测接近正面的行人头部。

### 2.1 特征提取

在行人检测中,构造具有判别力的特征是检测的基础。尽管尺度不变特征变换(SIFT)/梯度直方图(HOG)特征在行人检测中取得了良好的效果<sup>[5-6]</sup>,但是这些方法不能直接用于本文的场景中。主要由于以下两方面原因:(1)行人是非刚体目标,在连续视频帧中不能维持一致的特征点;(2)监控视频的分辨率低,行人的表面特征不明显。因此,与SIFT特征描述不同,本文在整个行人的头部区域建立梯度方向直方图,通过局部梯度方向和位置的分布来描述头部的表面和形状,并且允许梯度位置在小区域内的适度偏移。图1左侧给出了地铁交通视频中典型的行人头部与非头部图像,将其划分为 $4 \times 4$ 的不重叠小区域,在每个小区域内定义8个方向bin,用 $b$ 表示。为了计算梯度方向直方图,首先计算头部图像区域内梯度 $G_n$ 和方向 $A_n$ ,然后在每个小的区域内部累积方向直方图 $v_{ij}$ ,如式(1)所示。

$$v_{ij} = \sum_{n=1}^N G_n \delta(b - \lfloor A_n \rfloor) \quad b = 1, \dots, 8 \quad (1)$$

式中,每个方向的大小为区域内所有接近该方向的梯度值总和。最后将所有小区域内方向直方图串接,得到 $4 \times 4 \times 8 = 128$ 维头部区域特征向量,这样提取的特征向量隐式地包含了特征分布的空间信息。

图1右侧分别给出了头部与非头部对应的特征向量。从特征向量的分布可以看出,头部图像的特

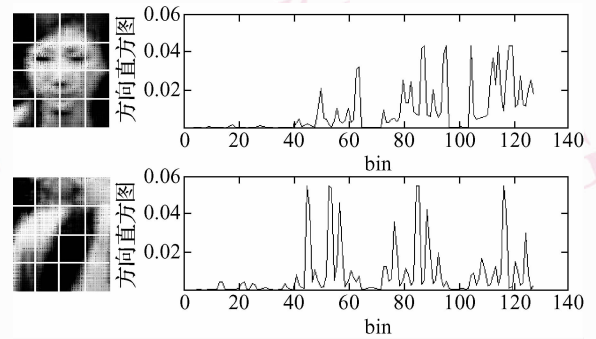


图1 头部与非头部图像及对应的梯度方向直方图

Fig. 1 Head and non-head images with their histogram of the gradient orientations

征分布更加集中,而非头部的特征分布比较分散。

### 2.2 自适应滑动窗口

在地铁监控视频中,行人的尺度变化较大,特别是在远离镜头的区域,行人的可见部分变得很小,难以提取出有效的目标特征。在现有的行人检测算法中,普遍采用多尺度重复扫描机制。然而,这一方法在取得较好检测结果的同时,严重降低了检测的速度,限制了实时性应用。为解决这一问题,文中采用自适应检测窗口,某一位置 $(x, y)$ 的窗口的高度 $h$ 如下:

$$h = \frac{y - y_0}{y_1 - y_0} (h_1 - h_0) + h_0 \quad (2)$$

式中, $y_0$ 和 $y_1$ 分别为处于检测区域上部和下部的行人头部位置的纵坐标, $h_0$ 和 $h_1$ 则分别为该位置对应的头部区域的高度。根据人头部的比例特征,定义窗口的宽高比为 $1 : 1.25$ 。检测窗口以8像素的步长在检测区域内滑动,密集提取检测区域内的图像特征。

### 2.3 学习 SVM 分类器

线性 SVM 由于其简单、有效,而被广泛应用于特征分类。文中采用线性 SVM 分类器,对提取的行人头部区域特征进行分类。

学习两类分类器问题可以归结为寻找分类函数 $f: \mathbf{R}^n \rightarrow \pm 1$ ,建立模式 $x_i \in \mathbf{R}^n, i = 1, \dots, n$ ,与其对应类别 $y_i \in \{-1, 1\}$ 间的映射关系,即 $y = f(x)$ 。在 SVM 情况下,分类函数 $f$ 的形式为

$$f(x) = \sum_{i=1}^N y_i a_i k(x, x_i) + b \quad (3)$$

式中, $N$ 为训练的样本数, $(x_i, y_i)$ 为训练模式及其类别, $a_i$ 和 $b$ 分别为需要学习的权值和分类阈值,

$k(\cdot)$  为核函数。可采用径向基核函数  $k(x, x_i) = e^{-\|x-x_i\|^2/2\sigma^2}$ , 对于  $a_i > 0$  的模式定义为支持向量。函数  $f(x) = 0$  定义了一个特征空间的超平面。选择权值  $a_i$  和  $b$ , 使训练样本的错分数最小, 同时使得超平面到支持向量的距离最大。这一过程可以转化为求解下面的优化问题:

$$\begin{aligned} \max \sum_{i=1}^N a_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N y_i y_j a_i a_j k(x_i, x_j) \quad (4) \\ \text{s. t. } 0 \leq a_i \leq C, \quad \sum_{i=1}^N y_i a_i = 0 \end{aligned}$$

式中, 常数  $C$  为惩罚参数, 对错分进行约束, 模型参数的选择, 是支持向量机中的典型问题, 没有统一的选择标准和理论, 一般根据具体的应用问题选取。这里为了减小对经验误差的惩罚, 参数  $C$  取 0.1。

在本文的应用, 分别选取 500 个正样本和负样本, 图 2 给出了密集客流视频中头部图像正样本, 经过预处理后进行特征提取, 而负样本则为对应视频序列中不含头部图像区域。对所有样本图像进行归一化处理, 并分配类别标签, 对正样本  $y_i = 1$ , 负样本  $y_i = -1$ 。



图 2 监控视频中头部图像样本

Fig. 2 Positive head samples in the surveillance video

在初始帧的检测区域内, 采用自适应滑动窗口密集提取特征向量。经过 SVM 分类和聚类后, 能够在头部区域附近给出初始头部假设。

### 3 头部跟踪

通过 SVM 分类虽然可以得到较高的目标检测率, 但是对连续视频进行检测不但非常耗时, 而且会引入多目标的对应问题。为了提高实时性和降低误检率, 采用多目标跟踪算法对初始帧中的头部假设进行跟踪, 得到头部假设在连续视频帧中的运动序列, 通过对图像序列的整体判别, 能够给出更加可靠的检测结果。

为了得到头部运动序列, 首先将初始头部假设区域转换到 HSV 颜色空间, 计算区域内的颜色直方图, 采用 Mean Shift 对头部假设区域进行跟踪<sup>[7]</sup>。从进入检测区域到离开检测区域, 得到头部图像序

列。图 3 给出了采用 Mean Shift 算法的头部跟踪结果。



图 3 Mean Shift 跟踪结果

Fig. 3 The results of Mean Shift tracking

## 4 实验与分析

实验采用上海地铁的监控视频, 对提出的系统进行测试。图 4 中的圆形指出了 SVM 对梯度方向直方图分类结果, 矩形框为经过聚类后得到的头部假设。从图中可以看出, 对于接近正面的头部进行了有效的检测, 同时对于一些非头部区域, SVM 分类器做出了错误的检测。这样的检测结果说明, 对于低分辨率的监控视频, 只采用梯度方向直方图和 SVM 分类器的方法还不能给出满意的检测结果。因此, 需要对连续视频中的头部图像进行跟踪, 利用多帧的检测信息进行判断。



图 4 SVM 初始检测结果

Fig. 4 The initial detection by SVM

图 5 给出了头部跟踪的部分序列,白色的矩形框为跟踪结果,矩形框内部为 SVM 分类结果。可以看出,对于包含头部图像的初始假设,如图 5 的顶部一排所示,通过对连续视频帧的整体判别,检测的结果得到进一步提升。而对于非头部假设,如图 5 的下排所示,分类器对跟踪序列图像给出了更多的否定检测结果。因此,综合跟踪序列的整体判别,能够去除那些由于误检产生的头部假设,从而提高检测率。

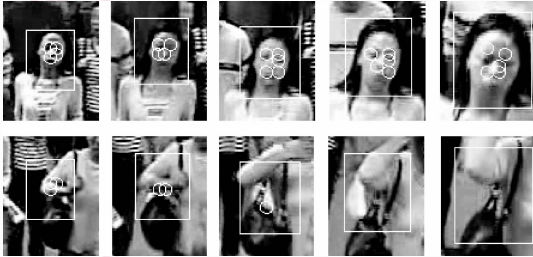


图 5 跟踪得到的头部假设序列

Fig. 5 Head hypotheses sequences associated by tracking

图 6 为经过连续视频序列跟踪判别后的结果。去除了初始假设中的非头部区域,而真正的头部区域得到了保留。在当前的视频图像中,检测区域内包含的完整头部信息的行人数为 11,其中漏检的一个行人是由于在初始帧没有获得正面的图像。最终的检测率为 91%。在 Matlab 仿真实验中,每秒可以处理 5 到 7 帧的视频图像。



图 6 最终的检测结果

Fig. 6 The final result of detection

## 5 结论

本文提出了基于多目标检测和跟踪的客流量检测方法。通过自适应检测窗口提取判别性的图像特征,结合 SVM 分类器,得到头部图像的初始假设。然后,多目标跟踪算法被用来跟踪连续图像序列中的头部假设,通过对整个序列的判别检测精度得到了提高。采用实际地铁监控录像的测试表明,该方法可以准确地检测客流量信息。如何有效地提取特征,提高算法的实时性是下一步的重点研究内容。

## 参考文献 (References)

- 1 Marana A N. Real-time crowd density estimation using images [J]. *Advances in visual computing*, 2005, **3804**: 355-362.
- 2 Bastian Leibe, Edgar Seemann, Bernt Schiele. Pedestrian detection in crowded scenes [A]. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*, San Diego, CA, USA, 2005: 878-885.
- 3 Dan Kong, Doug Gray, Hai Tao. A viewpoint invariant approach for crowd counting [A]. In: *Proceedings of the 18th International Conference on Pattern Recognition [C]*, Hong Kong, China, 2006: 1187-1190.
- 4 Rabaud V, Belongie S. Counting crowded moving objects [A]. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*, New York, NY, USA, 2006: 705-711.
- 5 David G. Lowe. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004: 91-110.
- 6 Navneet Dalal, Bill Triggs. Histograms of oriented gradients for human detection [A]. In: *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]*, San Diego, CA, USA, 2005: 886-893.
- 7 Dorin Comaniciu, Visvanathan Ramesh, Peter Meer. Kernel-based object tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564-577.