

Research on Pedestrian Detection Algorithm Based on BING and HOG-LSS Feature

Chaohua Zhao^{1,2}

¹School of Computer Science, Sichuan University, Chengdu Sichuan

²The National Key Discipline Laboratory of Visual Synthesis Graphics Technology, Chengdu Sichuan

Email: 853906141@qq.com

Received: Dec. 17th, 2016; accepted: Dec. 31st, 2016; published: Jan. 5th, 2017

Copyright © 2017 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In recent years, pedestrian detection, based on computer vision, has been one of the hottest topics in the field of intelligent transportation. The pedestrian detection algorithm, based on HOG and local self-similarity (LSS) feature fusion, is better than the traditional HOG detection algorithm, and also it has the following challenges: 1) low efficiency; 2) failing to effectively handle the occlusion problem. Aiming at these challenges, this paper proposes a pedestrian detection optimization framework based on BING feature, HOG-LSS feature and data trajectory fusion. It is proved that the detection result is superior to the HOG-LSS pedestrian detection method.

Keywords

Pedestrian Detection, BING Feature, HOG-LSS Feature, Data Association

基于BING和HOG-LSS特征的行人检测算法研究

赵朝华^{1,2}

¹四川大学计算机学院, 四川 成都

²视觉合成图形图像技术国家重点学科实验室, 四川 成都

Email: 853906141@qq.com

收稿日期：2016年12月17日；录用日期：2016年12月31日；发布日期：2017年1月5日

摘要

近年来，基于计算机视觉技术的行人检测方法一直是智能交通领域研究的热点问题之一。基于HOG和局部自相似(LSS)特征融合的行人检测算法在检测效果上优于传统HOG特征的行人检测算法，但是同时也存在如下挑战：1) 算法检测的速度不够快；2) 在遮挡面积过大的情况下，无法有效地进行处理。针对这些挑战问题，本文提出了一种使用BING特征、HOG-LSS特征和数据轨迹融合的行人检测优化框架，并通过实验结果进行验证可知，检测效果优于HOG-LSS特征的行人检测方法。

关键词

行人检测，BING特征，HOG-LSS特征，数据轨迹融合

1. 引言

行人检测，顾名思义，就是把图像或视频中出现的行人从背景中分割出来，并精确定位。近年来，基于计算机视觉技术的行人检测方法一直是智能交通领域研究的热点问题之一，除此之外，它还广泛的应用于智能监控领域，对视频监控场景区域内的可疑人物或事件进行检测、识别和报警，可以有效的阻止犯罪和异常事件的发生。另外，行人检测在人机交互、虚拟现实、医学图像、机器人视觉导航等也应用非常广泛。

目前，行人检测方法主要有背景差分和帧间差分的行人检测方法、模板匹配的行人检测方法、基于光流的行人检测方法以及基于机器学习的行人检测，由于利用机器学习的思想来解决由环境变化、行人衣着和形态变化给检测带来的影响，所以这种行人检测方法已经成为视频行人检测领域中最活跃的主流研究方向，它主要包括特征提取和机器学习两部分。2005年，Dalal [1]等提出了梯度方向直方图(histograms of oriented gradient, HOG)特征和支持向量机(support vector machine, SVM)分类器的行人检测方法，并试验中取得了很好的表现，但是对于遮挡和复杂背景下的行人检测效果差，所以随之又有一些新的论文的出现[2] [3] [4]，比如局部自相似(local self-similar, LSS)特征[5]就是通过捕捉图像颜色自相似性、边缘特征和重复模板以及复杂的纹理来计算两幅图像的相似性。在行人检测上的分类器主要有 SVM 和 AdaBoost 两种，Maji [6]等采用了一种新的核 SVM 方法来检测行人，叫做直方图交叉核 SVM (histogram intersection kernel support vector machine, HIKSVM)，这种方法不但可以快速检测到行人，并且检测精度还有提升。朱文佳等提出了一种 Gentle Adaboost 的分类方法，且成功地应用到行人检测上。种衍文等[7]提出的采用多特征结合 Adaboost 和 SVM 的两级行人检测算法，也取得了不错的效果。

但是，由于原有的算法存在速度不够快且无法有效的处理遮挡等问题，所以本文提出了基于 BING (binarized normed gradients, BING)特征[8]和 HOG-LSS 特征的数据轨迹融合的行人检测优化框架。实验证明，该框架不仅能改进现有的行人检测算法的检测性能，还能大大提高检测速度。

2. 本文的行人检测框架与以前框架对比

如图 1 所示，传统的行人检测框架是直接利用行人检测算法(HOG-LSS)对输入图片进行检测，再输出结果。但是，如图 2 所示，对于使用 BING 特征和数据融合的行人检测优化框架针对 HOG-LSS 速度不够快的问题，可以优化成三个步骤：1) 首先，使用 BING 进行粗检测，在保证精度的前提之前快速过滤

掉大部分候选窗口；2) 其次，再使用 HOG-LSS 进行精确检测，在去掉大量候选窗口之后，能够减少窗口检测数量；3) 最后，使用轨迹融合的方法，将属于同一个目标的多个轨迹片段融合在一起，并且依据完整轨迹更新检测结果。

3. 本文的行人检测特征

3.1. BING 特征的使用

传统方法的目标检测和行人检测大都是使用滑动窗口(图 3)，先学习大量的样本，以获得学习结果，然后用不同的框遍历需测试的图片，并将遍历的框中内容依次与学习结果比对，再确定框中是否存在此物体。对于一副大小为 $N \times M$ 的图片来说，可能的子窗口就有 $N^2 \times M^2$ 个，从算法本身来说，这就极大地限制了目标检测系统的实时性。但是如果基于一种先验知识——人是一种具有封闭性特点的“物体”，就可以快速的从 $N^2 \times M^2$ 个窗口中过滤掉绝大部分，然后再用特征和分类器对剩余的窗口进行行人检测，就可以在保证精度的同时，提高检测的速率。

人们观察事物都有一个习惯：先粗看，再细看。粗看是分辨视野中是否有物体，快速过滤掉视野中大部分的其他部分窗口，得到一个相对较小的窗口集合，细看则是仔细分辨物体是什么，并且在图片或视频帧中，物体对象是具有良好的封闭边界，而背景则是杂乱无章的。行人检测就是在图片或视频帧中检测哪些位置出现了人(如果把人也看作是一种“物体”)。BING 检测就是这样一种“粗略检测”的方法，先将目标的大概位置检测出来，这是提速的过程。在 Cheng [8]提出的 BING 目标检测中，将训练图像采



Figure 1. The traditional pedestrian detection framework

图 1. 传统的行人检测框架

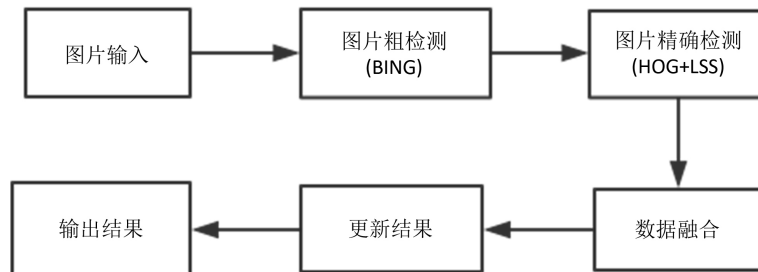


Figure 2. The pedestrian detection framework of this paper

图 2. 本文的行人检测框架

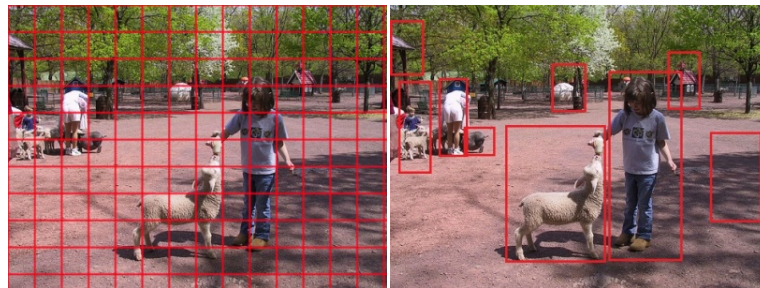


Figure 3. The left picture shows the sliding window method, and the right picture shows the general object detection method

图 3. 左图为滑动窗口方法，右图为一般普通物体检测方法

集若干个窗口并调整为 8×8 的大小，提取 BING 特征，然后将 8×8 的矩阵向量化为 64 维的特征向量，用 SVM 分类器进行训练。

Cheng [8]的方法之所以会有如此快速的检测表现，主要是由于有以下三个特点：1) 大胆地将图像缩小为 36 种大小不同的图像，然后用 8×8 的模板进行匹配，使其在滑动图像匹配时少了数倍，并且物体的结构轮廓信息存在，所以不会影响检测结果；2) 特征维数低，只有 64 维；3) 将特征二值化，64 维的特征值刚好可以使用一个 64 位(int64)的整型数据结构来存储，从而可以使用二进制的位处理大大加速运算。

本文的行人检测算法使用 Cheng [8]的 BING 特征对输入图片进行粗检测，与 Cheng [8]的“一般物体”的检测有所区别：1) 训练的正样本集不同，本文训练的正样本集只包含行人的图片集；2) 行人具有固定的高宽比，因此不必考虑过多的高宽比变化。

3.2. HOG-LSS 特征的使用

HOG 特征是一种常用于计算机视觉与图像处理中进行物体检测的特征描述器，它的步骤是 1) 首先将图像分成联通的区域(cell)；2) 然后收集 cell 中每个像素的梯度构成方向直方图；3) 再将直方图组合成特征描述子。

LSS 特征是用来捕捉图像间的自相似性，当存在小规模形变的时候，LSS 特征通过捕捉颜色、边缘、重复图样和复杂纹理的自相似性，从而匹配图像。

利用 LSS 特征的这种自相似性，再结合 HOG 特征，来辅助区分行人和非行人目标，从而可以降低误检漏检率。本文采用串行特征组合方法，形如式子(1)所示：

$$R = \{\alpha, \beta, \alpha \in X, \beta \in Y\} \quad (1)$$

式中， α 表示 HOG 特征， β 表示 LSS 特征。将 LSS 特征作为辅助检测，与 HOG 特征在降维后融合，从而在不增加计算量的基础上能降低传统算法的误检漏检率。

4. 数据融合

本文的数据融合也称为轨迹融合，它是为了解决这样的问题：对于一组轨迹片段，这组轨迹片段长度各不相同，且没有目标的外观信息，我们解决的目标是将属于同一图片目标的轨迹融合在一起。此处有三个挑战：1) 目标遮挡；2) 缺乏外观信息；3) 目标和摄像机的移动。为了解决这些挑战，我们可以将数据融合问题看成是一个广义线性分配的问题(Generalized Linear Assignment, GLA)，使用动态相似性来求解。

4.1. 广义线性分配(GLA)问题

假如给定 N 个轨迹片段 $\{\alpha_1, \dots, \alpha_N\}$ ，线性分配(LA)可以看作是如下的优化问题：

$$\max_X \sum_{i=1}^N \sum_{j=1}^N P_{ij} X_{ij}$$

其中，

$$\sum_{i=1}^N X_{ij} = 1; \sum_{j=1}^N X_{ij} = 1; X_{ij} \in \{0, 1\} \quad (2)$$

其中， P_{ij} 表示轨迹片段 α_i 和 α_j 之间的某种相似度(后面介绍具体的相似性计算方法)，所以 P 是一个前驱-后继矩阵，表明其中相邻片段之间的相似性。例如，当 α_j 不可能是 α_i 的后继片段时， P 可以表示为 $-\infty$ ；当 α_j 可能是 α_i 的后继片段时， P 表示为大于 0 的数字。优化变量 X_{ij} 表示当 $X_{ij} = 1$ 时， α_i 是 α_j 的前驱轨迹，这样的轨迹将会合并在一起。

由于式子(2)中使用的条件[9]是必须使得每个轨迹片段分配一个前驱轨迹和一个后继轨迹，为了避免

这个要求，本文使用 GLA 的另一个式子(3)，如下：

$$\max_X \sum_{i=1}^N \sum_{j=1}^N P_{ij} X_{ij}$$

其中，

$$\sum_{i=1}^N X_{ij} \leq 1; \sum_{j=1}^N X_{ij} \leq 1; X_{ij} \in \{0,1\} \quad (3)$$

GLA 与 LA 的主要区别就是，在 GLA 中，并没有强制规定每个片段必须与其他轨迹片段结合起来，但是却导致了两个重要结果：1) 避免设置源节点和学习轨迹的入口和终止概率；2) GLA 是一个 NP 完全问题，解起来比 LA 复杂。但是，在非常弱的约束条件下，这个问题可以用文献[4] [10]中的 softassign 算法来求解。

4.2. 基于动态的相似性计算

轨迹片段 α 由一组有序序列 y_k 组成， $s \leq k \leq e$ 。其中， y_k 表示候选行人窗口中心点的坐标 s 表示轨迹片段的开始时间， e 表示轨迹片段的结束时间，并且轨迹片段的动态性可以用线性回归来表示：

$$y_k = \sum_{i=1}^N a_i y_{k-i}, k \geq s + N \quad (4)$$

在没有噪声的情况下， $N = \text{rank}(H_a^m)$ ，其中， $\text{rank}(\cdot)$ 表示矩阵的秩， H_a^m 是一个 $m \geq n$ 列的汉克尔 (Hankel) 矩阵：

$$H_a^m = \begin{bmatrix} y_s & \cdots & y_{s+m-1} \\ \vdots & \ddots & \vdots \\ y_{e-m+1} & \cdots & y_e \end{bmatrix} \quad (5)$$

所以轨迹片段 α_i 和 α_j 之间的基于动态的相似度 P_{ij} 可以用如下的公式来表示：

$$P_{ij} = \begin{cases} -\infty, & \text{相互冲突} \\ \frac{\text{rank}(H_{a_i}) + \text{rank}(H_{a_j})}{\min_{\beta_i^j} \text{rank}(H_{a_{ij}})} - 1, & \text{不冲突} \end{cases} \quad (6)$$

式中， $\alpha_{ij} = [\alpha_i \beta_i^j \alpha_j]$ 表示轨迹片段 α_i 和 α_j 之间的联合轨迹， β_i^j 表示轨迹片段 α_i 和 α_j 之间的间隔轨迹。式子(6)表示，如果两个轨迹片段 α_i 和 α_j 是属于同一个目标的轨迹，则他们的联合轨迹可以通过低秩的线性回归来表示；相反，如果两个轨迹片段属于不同目标的轨迹，则他们的联合轨迹必须使用高秩的线性回归来表示。因此，如果 $\text{rank}(H_{a_i}) = r_i$ ， $\text{rank}(H_{a_j}) = r_j$ ，则 $\text{rank}(H_{a_{ij}}) = r_{ij} \leq r_i + r_j$ 。相应地，如果两个轨迹片段 α_i 和 α_j 同属于一条轨迹，则有 $r_{ij} = r_i = r_j$ ， $P_{ij} = 1$ ；同时，相反地如果两个轨迹片段 α_i 和 α_j 不属于一条轨迹，则有 $r_{ij} = r_i + r_j$ ， $P_{ij} \approx 0$ 。

5. 更新结果

通过轨迹融合得到每个行人的完整轨迹之后，就可以用完整轨迹更新检测结果，这其中就包括推测行人位置以及修正检测得分两部分。

5.1. 推测行人位置

图 4 所示是某个行人的完整轨迹，由于遮挡等原因，在 $t_2 \sim t_4$ 期间行人检测器没有检测到该行人，

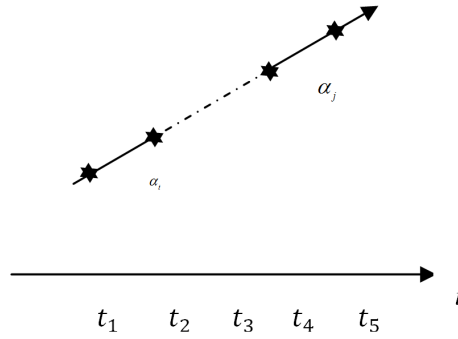


Figure 4. Pedestrian location speculation on t_3
图 4. t_3 时刻行人位置推测

但是通过轨迹融合可以得到轨迹片段 α_i 和 α_j 属于同一条轨迹，因此我们可以依据在 t_2 和 t_4 时刻该行人的位置推测出 t_3 时刻该行人的位置。

$$r(t_3) = \frac{(t_4 - t_3) * r(t_2) + (t_3 - t_2) * r(t_4)}{t_4 - t_2} \quad (7)$$

其中， $r(t)$ 表示 t 时刻该行人的矩形框。

5.2. 修正检测得分

如果一条检测轨迹上的候选窗口的平均得分越高，就越有可能包含行人，因为整体上来说，包含行人目标的窗口的检测得分比背景窗口上的检测得分要高。因此，可以使用如下的公式来修正检测得分：

$$s_f(i) = \alpha s(i) + (1 - \alpha) \bar{s} \quad (8)$$

其中， $s(i)$ 表示候选行人窗口的原始检测得分， $s_f(i)$ 表示候选窗口中行人窗口修正后的检测得分， \bar{s} 表示轨迹中所有候选行人窗口检测得分的平均值， α 是一个权重系数，这里取值为 0.75。

6. 实验结果与分析

本文实验数据集采用的数据库是 TUD-Brussels 数据库和 ETH 数据库，这两个数据库都是视频中提取出来的连续帧，也是行人检测常用的数据集，吻合本文的需求。同时，本文采用评价标准是使用误检率和漏检率曲线，来衡量检测器的性能。其中，误检率表示分类结果为正例行人目标，而实际上为负例行人目标；而漏检率为测试集为正例但是判别为负例的数目/测试集中正例的数目。在误检率一定的情况下，漏检率越低检测性能越好；相反，漏检率越高检测性能越差。为简单起见，本文选取误检测率为 10^{-1} 时的漏检率值来作为检测器性能的衡量。

表 1 是原始的 HOG-LSS 方法和本文的方法的检测性能和速度的对比。其中，“HOG-LSS”表示原始的检测器，“BING + HOG-LSS + TRACK”表示本文的方法。从实验结果可以看出，对比于原始的 HOG-LSS 算法，本文的算法不仅可以大大提高检测速度，在提升检测性能上还有不错的表现。

7. 总结

本文在 HOG-LSS 基础上，提出了一种基于 BING 和数据轨迹融合的行人检测框架优化方案。该检测框架主要针对于 HOG-LSS 检测速度不够快而引入一个粗检测的过程，这种粗检测可以在保证检测精度的前提之下，快速的过滤掉大部分候选行人窗口，从而有效的减少 HOG-LSS 算法进行检测的时间。同时，针对检测算法无法有效地解决遮挡的问题，利用数据轨迹融合的思想，将属于同一个目标的多个轨迹片段融合在一起，依据完整的轨迹更新检测结果。从结果来看，实验证明本文的方法不仅可以大大提

Table 1. Comparison of experimental results for each classifier
表 1. 各分类器实验结果对比

检测算法	TUD-Brussels 数据集		ETH 数据集	
	速度(s)	漏检率(%)	速度(s)	漏检率(%)
HOG-LSS	2.06	67.71	2.56	50.16
BING + HOGLSS + TRACK	0.14	64.32	0.15	45.21

高检测的速度，还可以提升检测性能。在未来的研究中，将更加侧重于研究基于行人的运动轨迹建模，在进一步降低误检率和漏检率的基础上，进行行人识别跟踪和行人姿态方面等的研究。

参考文献 (References)

- [1] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 *IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, 886-893. <https://doi.org/10.1109/CVPR.2005.177>
- [2] 王孝艳, 张艳珠, 董慧颖, 等. 运动目标检测的三帧差分算法研究[J]. 沈阳理工大学学报, 2011, 30(6): 82-85, 91.
- [3] Yan, J.J., Zhen, L., Dong, Y. and Li, Z.S. (2012) Multi-Pedestrian Detection in Crowded Scenes: A Global View. 2012 *IEEE Conference on Computer Vision and Pattern Recognition*, Rhode Island, 3124-3129.
- [4] Shao, H., Chen, S., et al. (2015) Face Recognition Based on Subset Selection via Metric Learning on Manifold. *Frontiers of Information Technology & Electronic Engineering*, **16**, 1046-1058. <https://doi.org/10.1631/FITEE.1500085>
- [5] Yao, S.H., Pan, S.M., et al. (2015) A New Pedestrian Detection Method Based on Combined HOG and LSS Features. *Neurocomputing*, **151**, 1006-1014. <https://doi.org/10.1016/j.neucom.2014.08.080>
- [6] Maji, S., Berg, A.C. and Malik, J. (2013) Efficient Classification for Additive Kernel SVMs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 66-77. <https://doi.org/10.1109/TPAMI.2012.62>
- [7] 种衍文, 匡湖林, 李清泉. 一种基于多特征和机器学习的分级行人检测方法[J]. 自动化学报, 2012, 38(3): 375-381.
- [8] Cheng, M.M., Zhang, Z., Lin, W.Y., et al. (2014) BING: Binarized Normed Gradients for Objectness Estimation at 300 fps. 2014 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3286-3293. <https://doi.org/10.1109/CVPR.2014.414>
- [9] Zhang, L., Li, Y. and Nevatia, R. (2008) Global Data Association for Multi-Object Tracking Using Network Flows. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, 24-26 June 2008, 1-8. <https://doi.org/10.1109/cvpr.2008.4587584>
- [10] Gold, S. and Rangarajan, A. (1996) Softmax to Softassign: Neural Network Algorithms for Combinatorial Optimization. *Journal of Artificial Neural Networks*, **2**, 381-399.

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: jisp@hanspub.org