

# 基于改进 EOH 特征的行人检测

周千昊 戚飞虎

(上海交通大学计算机科学与技术系, 上海 200240)

**摘要** 行人检测是物体检测领域的一大难点。为了进一步提高行人检测的精度和速度,将 Kobi Levi 和 Yair Weiss 提出的边缘方向直方图特征和传统的 Adaboost 算法两者有机地结合起来,提出了一种基于改进的 EOH 特征的行人检测算法。该算法首先对原先只适用于较为简单的人脸检测的 EOH(edge orientation histogram)特征进行了改进,弥补了其对于行人的对称性特征的描述能力不足的问题,然后通过改进 Adaboost 算法中对样本权值进行调整的策略来减少 overfitting。实验证明,该方法的检测性能能够接近目前行人检测的领先水平。在误报率为 1/10000 时,该算法在一个复杂的 Inria 行人数据集上的检测率可以达到 90%。对于 640 × 480 大小的图片,该算法的检测速度可以达到 2fps。

**关键词** 边缘方向直方图 平滑的 boosting 算法

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2008)10-1861-04

## Improved EOH Based Pedestrian Detection

ZHOU Qian-hao, QI Fei-hu

(Department of Computer Science and Technology, Shanghai Jiaotong University, Shanghai 200240)

**Abstract** Pedestrian detection is a difficult problem in the field of object detection. We combine Kobi Levi and Yair Weiss's edge orientation histogram and Dalal's hog (histogram of gradients) feature and apply them to pedestrian detection. We improve the algorithm from the following aspects: first, we have changed the calculation formula of the original EOH (edge orientation histogram) to gain more descriptive ability. Second, we have changed the policy of updating the weight of the samples of the original Adaboost algorithm in order to reduce overfitting. Experiments show our method is very efficient. When the false positive rate is 1/10000, our detection rate is about 90% on Inria pedestrian dataset. The running speed is about 2 fps with 640 × 480 images on a 1.8GHz CPU.

**Keywords** EOH, smoothed boosting algorithm

## 1 引言

行人检测对计算机视觉领域的许多应用非常重要,例如视频监控、汽车驾驶员辅助系统、人体的运动捕捉系统等。尤其在汽车驾驶员辅助系统中,由于连续两帧图像的背景是变化的,使得基于运动信息的行人检测方法在这种环境下的检测性能很差,因而,应用于静止摄像机情况下的一些有效方法,如 Viola 和

Jones 提出的集成行人外貌和运动信息的行人检测方法<sup>[1]</sup>,其在移动摄像机环境中误报率非常高。目前,移动摄像机环境中的行人检测方法主要采用行人的外貌信息来在单帧图像上检测行人,然而,在单帧图像上进行行人检测存在着以下几个难点:

- (1)有效的模型;
- (2)复杂的背景;
- (3)光照对比度的差异;
- (4)速度上的要求。

本文在 Viola 和 Jones 提出的行人检测框架的基础上,做了以下两个主要的改进:

第 1 个改进是,对 Levi 和 Weiss 原先应用于人脸检测的 EOH(edge orientation histogram)<sup>[2]</sup> 特征进行改进,使其能够更好地捕捉行人的边缘信息,如肩部的对称信息等,以便能更好地适用于行人检测。

第 2 个改进是,在每一轮 boosting<sup>[3]</sup> 之后,即对样本权值进行更新之后,再对权值进行平滑(smoothing)处理,以便可以尽可能保证在进行下一轮的训练时,样本权值具有更好的泛化能力。

## 2 改进的 EOH 特征

### 2.1 用于人脸检测的 EOH 特征

EOH 在人脸检测的应用中,在同样的误报率下比 Viola 和 Jones 提出的方法的检测率高出 2 个百分点。尽管如此,由于上述提到的行人检测的复杂性,致使 EOH 特征并未得到很好的应用。

下面简单介绍一下 EOH 特征:

EOH 是边缘方向直方图的简称,因为边缘是人脸检测很重要的特征,所以要进行边缘提取,本文采用 Sobel 算子来进行边缘提取。

点  $(x, y)$  的边缘强度可定义为

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (1)$$

$G_x(x, y)$ ,  $G_y(x, y)$  分别为  $x, y$  方向上的梯度  
为了去除噪声的影响,可设定一个阈值

$$\hat{G}(x, y) = \begin{cases} G(x, y) & \text{如果 } G(x, y) \geq T \\ 0 & \text{其他} \end{cases} \quad (2)$$

实际应用中的  $T$  值通常设为 80 ~ 110。

点  $(x, y)$  的边缘方向可定义为

$$\theta(x, y) = \arctan\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (3)$$

然后将  $\theta$  ( $0^\circ \sim 180^\circ$ ) 等分为  $K$  个区间( $K$  值常取 4 ~ 8)

$$\varphi_k(x, y) = \begin{cases} \hat{G}(x, y) & \text{如果 } \theta(x, y) \in R_k \\ 0 & \text{其他} \end{cases} \quad (4)$$

现定义子窗口  $W$  中某一个方向  $k$  的边缘强度为  $E_k(W) = \sum_{(x, y) \in W} \varphi_k(x, y)$ 。由上述的定义即可得到如图 1 所示的 3 种类型的 EOH 特征:

图 1 (a) 表示水平和垂直方向的边缘强度的对比,

$$A_{i,j}(W) = \frac{E_i(W) + \xi}{E_j(W) + \xi} \quad (5)$$

图 1 (b) 表示在矩形框中水平方向的强度占主导地位,即

$$B_{i,j}(W) = \frac{E_i(W) + \xi}{\sum_j E_j(W) + \xi} \quad (6)$$

图 1 (c) 表示一组对称特征,其对称度为

$$S(W_1, W_2) = \frac{\sum_i |E_i(W_1) - E_i(W_2)|}{sizeof(W_1)} \quad (7)$$

式中,  $sizeof(W_1)$  表示子窗口  $W_1$  中包含的像素个数。

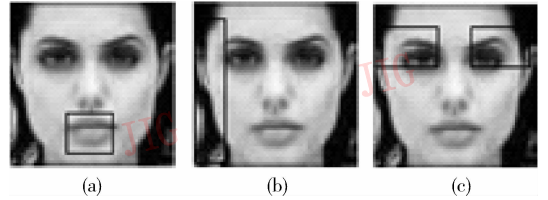


图 1 3 种不同的 EOH 特征

Fig. 1 Three different kinds of EOH features

### 2.2 EOH 特征的局限性以及改进

#### 2.2.1 除法的局限性及改进

由于除法的特殊性,使得特征值的范围太大,范围可以从 0.01 到 100。这也使得其概率密度分布建模的难度大大增加了。

虽然 EOH 特征值的取值范围非常大,但是主要集中在 0 到 20 之间,相当不均匀如图 2 所示。此时如果使用查找表(look-up table)<sup>[4]</sup> 训练弱分类器的时候,选出的分类器就缺乏代表性,难于反映大多数训练样本的特性。为了解决这个问题,可以在除法的基础上再做一次反正切的变换,这样一来,特征值将比原先更加均匀的分到值域范围之内(如图 3 所示)。

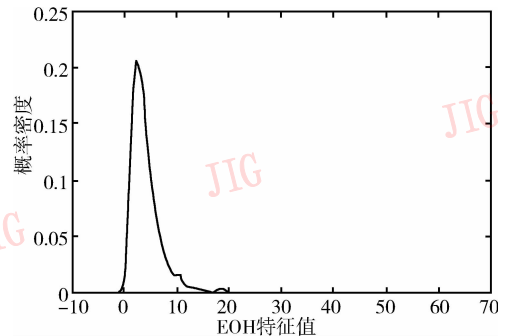


图 2 原本的 EOH 特征值的概率密度分布

Fig. 2 Probability density distribution of the original EOH feature

从上述两图可以看出,经过变换之后,特征值分布更加均匀,这与本文的预期也非常吻合。

改进后的 EOH 特征定义如下:

$$A_{i,j}(\mathbf{W}) = \arctan\left(\frac{E_i(\mathbf{W}) + \xi}{E_j(\mathbf{W}) + \xi}\right) \quad (8)$$

$$B_{i,j}(\mathbf{W}) = \arctan\left(\frac{E_i(\mathbf{W}) + \xi}{\sum_j E_j(\mathbf{W}) + \xi}\right) \quad (9)$$

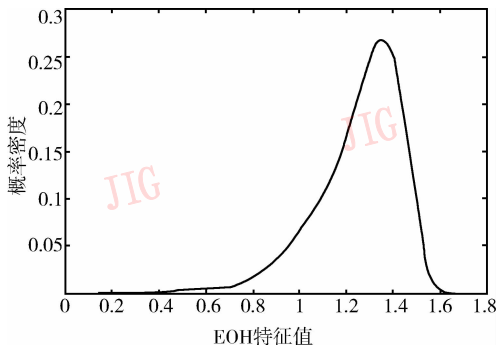


图 3 经过反正切变换后的 EOH 特征的概率密度分布  
Fig. 3 Robability density distribution of the EOH feature after arctan transformation

### 2.2.2 对称性的局限性及改进

仔细观察式(7)可见,在比较两个方向是否对称的时候,所比较的两个方向是相同的方向。但是很明显,如果两个区域中的图像是对称的,那么里面的边缘方向应该也是呈镜面对称的,也就是说,两个方向应该是互补的,而不是相等的。

而文献[2]中的对于对称特征的定义是由人脸检测的特殊性所造成的:人脸部几乎所有的对称部分,其边缘的主要方向都是垂直或者水平的,例如人的左右眼和左右脸颊。

如果直接将这样的定义应用于更加复杂的行人检测,那将是非常不适合的。例如,行人的两个肩膀虽然是对称的,但是其在方向上是互补,而非相等的。为此必须重新定义这个对称特征:

$$S_i(\mathbf{W}_1, \mathbf{W}_2) = \arctan\left(\frac{E_i(\mathbf{W}_1) - E_{p(i)}(\mathbf{W}_2) + \xi}{\sum_j (E_j(\mathbf{W}_1) + E_j(\mathbf{W}_2)) + \xi}\right)$$

$$T_i(\mathbf{W}_1, \mathbf{W}_2) = \arctan\left(\frac{E_i(\mathbf{W}_1) + E_{p(i)}(\mathbf{W}_2) + \xi}{\sum_j (E_j(\mathbf{W}_1) + E_j(\mathbf{W}_2)) + \xi}\right)$$

$$p(i) = (M - i) \% M \quad (10)$$

将  $0^\circ \sim 180^\circ$  划分为  $M$  个区间,  $M$  通常取 6。

## 3 改进的 boosting 算法

在利用 Adaboost 算法训练分类器的时候,经常会出现这样的问题,如果在训练样本中有一个或几

个样本很难被正确划分,那么在几轮的权值更新过后,这些样本将被赋予一个很高的权值。而且在之后的训练过程中,所挑选的弱分类器及其权值都仅仅是为了对这极少一部分的样本进行分类,也就是说,训练过程只是对这些样本上做过度拟合(overfitting)。

为了解决过度拟合的问题,有以下两个方法:(1)在训练达到一定时候,就可以将这些权值很高的难以被正确划分的样本从训练样本中剔除,这样做虽简单易行,但同时也限制了最终的分类器的分类能力;(2)对那些难以被正确划分的样本的权值进行调整,这也就是本文所采用的方法。

本文在每一轮的样本权值更新完之后,在进行下一轮样本权值归一化之前,要对每个权值加上一个很小的正数,本文称其为 smoothing,这样做的主要作用可以用如下一个简单的例子来证明:

$$w_0 = 0.02, w_1 = 0.001, N = 1000$$

$$w_0/w_1 = 20$$

$$\hat{w}_0 = w_0 + 1/N = 0.021,$$

$$\hat{w}_1 = w_1 + 1/N = 0.002$$

$$\hat{w}_0/\hat{w}_1 = 10.5$$

从上面的例子可以看出,  $\hat{w}_0/\hat{w}_1$  在 smoothing 之后变小了,也就是说,权值  $w_0$  对于特征选择的影响变小了。那么选择出来的特征也将能更好地反映更多样本的特性了。

上述 smoothing 是这样影响特征选择的过程的,即首先在加上一个小的正数后,看似所有的权值都增大了,其实一经归一化后,就使原先大的权值减小和使小的权值增大,这也就使得权值的分布更加均匀。这样就使得用大权值样本选出的那部分特征,其所对应的弱分类器的响应值变小了,从而也就更难达到训练的终止条件了。于是也就必须通过从特征池中去选择更多的特征来进行训练。

当然,若有更多的特征,则其表达能力就更强,但也就必须使用更多的时间进行训练,这也是这种方法所带来的额外开销。上例中,  $N$  一般取为训练样本大小的整数倍,调整  $N$  的大小即可在所选的特征数量和额外的训练时间开销之间取一个折中。

## 4 实验结果

训练样本中均来自 INRIA 的行人数据库,共包

含 2 416 个正样本和 4 000 多个负样本。在正样本图像中,行人穿着不同颜色、式样和花纹的服装,有着不同的动作,而且面对摄像机的角度变化很大,是在不同的光照条件和季节拍摄的。负样本图像都是自然场景图像。训练样本的尺寸为  $32 \times 64$  pixels。

实验时,本算法用 C++ 实现,并使用了公开的 opencv 图像处理库和 Visual Studio。NET2005 进行开发,并在 P4 2.0G CPU, 2G RAM 和 80G/7200rpm Hard Disk 的计算机上进行测试。测试集为标准的 INRIA 测试集,测试速度达到  $1 \sim 2$  fps。

表 1 与如今其他一些主流算法的比较结果

Tab.1 Comparison with state-of-art methods

方法	误报率 (%)	检测率 (%)
本文算法	0.01	90
	0.001	82
文献[5]算法	0.01	88
	0.001	80
文献[6]算法	0.01	94
	0.001	81
文献[7]算法	0.01	88
	0.001	<80

从表 1 可见,文献[5]的 SVM + HOG 和文献[6]的 CBT + Edgelet,在较高的误报率(false alarm rate FAR)  $10^{-4}$  下,具有较高的检测率(detection rate DR),比本算法都高出了 1 至 2 个百分点,但是在 FAR 下降时,上述两种算法的 DR 下降也比较快。而当 FAR 达到  $10^{-5}$  的时候,本文算法对 boosting 算法进行改进的优越性也体现出来了,由于加入了 smoothing 步骤,使得训练时提取了更多的特征,因此不仅增强了其泛化能力,而且使本文算法在误报率降低的时候,检测率下降更加缓慢(图 4)。

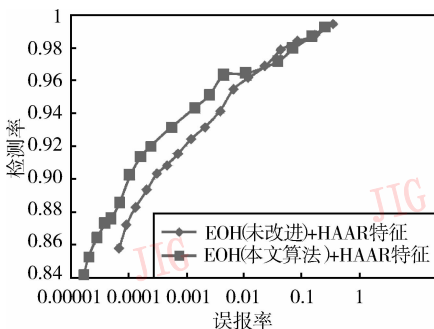


图 4 本文中所提算法与未经改进的算法比较

Fig.4 Comparison between improved method and the original one

## 5 结 论

本文提出了一种对 EOH 特征和 boosting 算法进行改进的算法,并在 Viola 和 Jones 提出的行人检测框架下,提出了基于新的 EOH 特征的行人检测算法,在实际应用中也取得不错的检测效果。

本文对 EOH 特征进行的改进不仅能使其更好地从人脸检测应用到行人检测中,而且能更好地捕捉行人的对称信息。而对于 boosting 算法的改进,在花费一定额外的时间之后,则能够很大程度上减少 overfitting 的发生。

本文的行人检测方法对一些杂乱的场景图像中的行人进行检测的实验结果表明,这种算法在移动摄像机环境下,对自然场景中的行人进行检测非常有效,同时也达到了如今一些主流算法的性能水平。

**致 谢** 本实验得到了日本 Omron 公司的支持和赞助,在这里对他们的支持表示真挚的感谢。

## 参考文献 (References)

- Viola Paul, Jones Michael. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137 ~ 154.
- Levi Kobi, Weiss Yair. Learning object detection from a small number of examples the importance of good features [A]. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04) [C], 2004.
- Freund Y, Schapire R E. 1995. A decision-theoretic generalization of on-line learning and an application to boosting [A]. In: Computational Learning Theory: Eurocolt [C], 1995: 23 ~ 37.
- Wu Bo, Ai Hai zhou, Huang Chang. Fast rotation invariant multi-view face detection based on real adaboost [A]. In: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C], Seoul, Korea, 2004: 79 ~ 84.
- Dalal Navneet, Triggs Bill. Histograms of oriented gradients for human detection [A]. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C], San Diego, CA, USA, 2005, 1: 886 ~ 893.
- Wu Bo, Nevatia Ram. Cluster boosted tree classifier for multi-view multi-pose object detection [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Rio de Janeiro, Brazil, 2007: 1199 ~ 1200.
- Wu Bo, Ram Nevatia. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors [A]. In: Proceedings of IEEE International Conference on Computer Vision [C], Beijing, China, 2005, 1: 90 ~ 97.