

基于多尺度方向特征的快速鲁棒人体检测算法*

叶齐祥¹⁺, 焦建彬¹, 蒋树强²

¹(中国科学院 研究生院, 北京 100049)

²(中国科学院 计算技术研究所, 北京 100190)

Fast and Robust Pedestrian Detection Algorithm with Multi-Scale Orientation Features

YE Qi-Xiang¹⁺, JIAO Jian-Bin¹, JIANG Shu-Qiang²

¹(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

²(Institute of Computing Technique, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: qxye@gucas.ac.cn

Ye QX, Jiao JB, Jiang SQ. Fast and robust pedestrian detection algorithm with multi-scale orientation features. Journal of Software, 2011, 22(12): 3004-3014. <http://www.jos.org.cn/1000-9825/3987.htm>

Abstract: The multi-scale orientation (MSO) features for pedestrian detection in still images are put forwarded in this paper. Extracted on randomly sampled square image blocks (units), MSO features are made up of coarse and fine features, which are calculated with a unit gradient and the Gabor wavelet magnitudes respectively. Greedy methods are employed respectively to select the features. Furthermore, the selected features are inputted into a cascade classifier with Adaboost and SVM for classification. In addition, the spatial location of MSO units can be shifted, are used to the handle multi-view problem and assembled; therefore, the occluded features are completed with average features of training positives, given an occlusion model, which enable the proposed approach to work in crowd scenes. Experimental results on INRIA testset and SDL multi-view testset report the state-of-arts results on INRIA include it is 12.4 times the faster than SVM+HOG method.

Key words: object detection; pedestrian detection; multi-view; occlusion

摘要: 提出一种多尺度方向(multi-scale orientation, 简称 MSO)特征描述子用于静态图片中的人体目标检测. MSO 特征由随机采样的图像方块组成, 包含了粗特征集合与精特征集合. 其中, 粗特征是图像块的方向, 而精特征由 Gabor 小波幅值响应竞争获得. 对于两种特征, 分别采用贪心算法进行选择, 并使用级联 Adaboost 算法及 SVM 训练检测模型. 基于粗特征的 Adaboost 分类器能够保证高的检测速度, 而基于精特征的 SVM 分类器则保证了检测精度. 另外, 通过 MSO 特征块的平移, 使得所提算法能够检测多视角的人体. 通过对于 MSO 特征块的装配, 使得算法能够检测人群中相互遮挡的人体目标. 在 INRIA 公共测试集合及 SDL 多视角测试集合上的实验结果表明, 算法具有对视角与遮挡的鲁棒性和较高的检测速度.

关键词: 目标检测; 人体检测; 多视角; 遮挡

中图法分类号: TP391 文献标识码: A

* 基金项目: 国家自然科学基金(61039003, 60872143); 国家重点基础研究发展计划(973)(2011CB706900, 2010CB731800)

收稿时间: 2010-04-06; 定稿时间: 2011-01-20

CNKI 网络优先出版: 2011-05-26 13:18, <http://www.cnki.net/kcms/detail/11.2560.tp.20110526.1318.002.html>

近年来,受到机器人视觉、智能家居、安全监控等应用的驱动^[1-6],人们越来越重视对于图像与视频帧中的人体目标检测.尽管人们已经采用各种方法进行研究,人体目标检测仍然受到视角、遮挡、效率等问题的挑战.

在图像目标检测中,首先需要考虑的问题是特征表达.早期的研究中,人们广泛采用 Haar 小波特征来描述图像目标^[7,8].然而,Haar-Like 特征容易受到目标形态、光照条件及视角的影响.近年来,在目标检测,尤其是人体目标检测中,轮廓特征表达受到了重视.文献[9]中,Dalal 等人提出了梯度方向直方图(HOG)特.HOG 特征是图像块内的方向统计,对于目标的颜色、局部形变与视角变化具有一定的鲁棒性.除了 HOG 以外,变尺度 HOG (v-HOG)^[10]、局部二值模式(LBP)特征^[11]、HOG-LBP^[12]及边缘集特征(Edgelet)^[13]都是具有代表性的轮廓特征.它们通常与一个统计分类器结合,用于人体目标检测.尽管上述特征都有较强的表达能力,它们一般都是在固定尺度上计算,没有考虑人体目标轮廓的多尺度特性.文献[14]仿照人类大脑皮层的工作方式提出一种多尺度的特征表达,并且基于 Gabor 小波进行实现.文献[15]提出了灰度协方差(COV)特征,通过选择不同位置的多尺度特征,COV 实现了多尺度轮廓表达.在文献[16]中,Sabzmeydani 等人也研究了多尺度的块方向特征,并且将多个块组合起来对人体目标进行表达.文献[17]对 HOG 进行了多尺度扩展,其实验结果表明,多尺度的表达能够提高性能.

在基于机器学习方法的目标检测中,级联 Adaboost 与 SVM 是人们较多采用的分类器^[18,19].其中,级联分类器的主要优点是其高的检测速度^[7],SVM 的主要优点是对目标模式变化的鲁棒性.也有部分研究者采用模板匹配的方法^[20,21].Munder 等人^[18]对比了几种特征及分类器,结论是:局部轮廓特征是最有效特征表达,而 SVM 分类器与级联 Adaboost 是最有效的分类器.在最近的综述论文中^[22],作者的结论是:HOG 特征加 SVM 分类器获得了最好的性能;而 Haar 小波特征加级联 Adaboost 分类器性能稍差,但是速度高出其他方法很多.

除了特征表达和分类器设计,还需要考虑视角、姿态及遮挡等具体问题.在文献[19,23]中,Wu 等人提出了 Tree Adaboost 分类器用以处理多视角问题,文献[20,21]通过多个模板解决多视角检测问题.在文献[12]中,Wang 等人基于局部检测器组合提出了一种解决人体目标遮挡的方法.文献[24-26]分别采用跟踪、分割、信息融合、多摄像机等策略来处理形变与遮挡问题.

在国内,中国科学院研究生院^[27]、清华大学^[28,29]、中国科学院计算技术研究所^[30]、中国科学院软件研究所^[31]对人体目标进行了深入研究.其中,文献[28,30]研究了一种新的方向特征空间或特征描述,而文献[30,31]分别侧重分类器及人体检测系统.

在人体目标检测研究中,全面考虑视角、遮挡及检测速度的方法较少.本文研究了一种新的多尺度局部轮廓(MSO)特征提取方法,与现有方法的不同是采用了尺度/方向竞争及块装配机制.其中,基于 Gabor 小波幅值竞争机制的局部轮廓特征是一种新的人体多尺度特征,而块装配机制则使得所提出的算法能够处理遮挡问题.算法根据块方向描述的内在关联特性对视角变化和遮挡进行了有针对性的处理,保证了鲁棒性.算法是一个由粗到精的过程,如图 1 所示.在粗检测过程中使用了块方向特征及级联 Adaboost 分类器,这使得算法有较高的速度;在精检测过程中使用了方向直方图特征及 SVM 分类器,保证了检测精度.

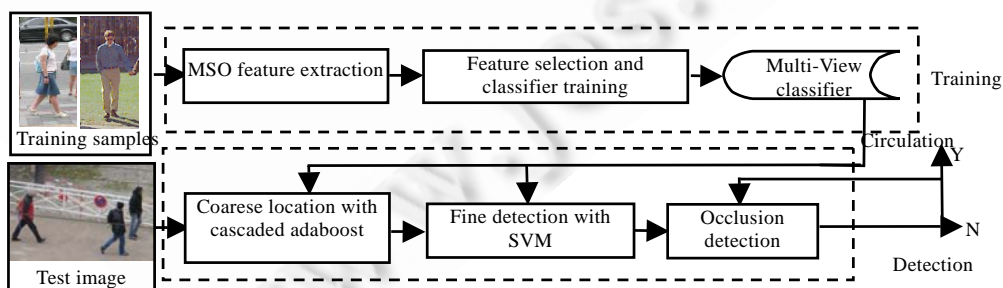


Fig.1 Training and detection process

图 1 训练及检测流程

本文第 1 节表述 MSO 特征提取及选择.第 2 节是人体检测算法.第 3 节探讨遮挡人体目标处理.第 4 节是实

验部分.第5节给出本文的结论.

1 多尺度方向特征提取与选择

为了实现人体目标的紧致表达及快速检测,作者采用了由粗到精的特征表达,并且提取了一个较完备的特征集合,然后对其中的特征进行了关联选择.

1.1 特征提取

对人体的观察表明,腿部、上身等部位有较大的轮廓,而头部、脖子等部位则有较小的轮廓.基于这些观察,我们认为,人体轮廓可以由适当位置及尺度的多个特征单元装配形成.这里的特征单元为方块,而非像 Haar-Like 特征那样的矩形块.其原因是,方块上可以计算整块方向,而且任何形状的矩形都可以由方块组合而成.仅选择方块,将大大降低特征的数量.

MSO 粗特征:粗特征的计算包含了如下两个步骤:1) 计算单元块方向;2) 将方向数值量化为离散特征.如图 2(b)所示,将特征块划分为上下左右子块,并分别计算子块上的颜色积分值,进而计算水平梯度:

$$Dh = \left| \sum_{X \in \text{Left subunit}} I(X) - \sum_{X \in \text{Right subunit}} I(X) \right| \tag{1}$$

其中, $I(X)$ 表示 X 像素的颜色值.同样可以计算块的竖直梯度 Dv .式(1)的计算可以通过“积分图”方法加快计算效率^[7],接着就可以计算块的方向并量化为特征:

$$f_i = Q(\arctan(Dv/Dh)) \tag{2}$$

其中, Q 是将 $0^\circ \sim 180^\circ$ 的连续方向转化为离散特征值 $\{0,1,\dots,7\}$ 的函数.

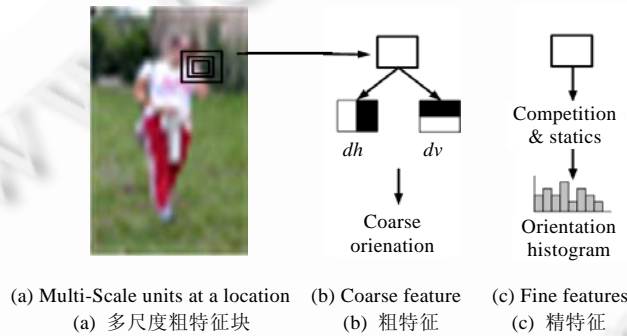


Fig.2 MSO features
图 2 MSO 特征

MSO 精特征:MSO 精特征基于 Gabor 小波计算.在图像处理与模式识别领域,人们广泛采用 Gabor 小波进行特征提取与目标表达,其滤波器特性决定了它对于带有结构重复性的纹理或者轮廓表达特别有效.Gabor 小波滤波器定义如下:

$$G_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{u,v}z} - e^{-\sigma^2/2}] \tag{3}$$

其中, $z = \begin{pmatrix} x \\ y \end{pmatrix}$, $k_{u,v} = \begin{pmatrix} k_{jx} \\ k_{jy} \end{pmatrix} = \begin{pmatrix} k_v \cos \phi_u \\ k_v \sin \phi_u \end{pmatrix}$, $k_v = \pi/2^{v/2}$, $\phi_u = u \frac{\pi}{8}$, $v = 0, \dots, v_{\max} - 1$, $u = 0, \dots, u_{\max} - 1$, v 表示频率, u 表示方向, $v_{\max} = 5, u_{\max} = 8, \sigma = 2\pi$.图 3 是滤波器 8 个方向、5 个尺度的幅值 ($\|G_{\sigma,\theta}(x,y)\|, v=1, \dots, 5, u=0, \dots, 7$) 可视化结果.

在可视化的 Gabor 小波幅值上可以直观地看到, Gabor 小波滤波器能够捕捉不同尺度的方向特征.对于样本上的每个像素,在滤波后获得 $5 \times 8 = 40$ 个幅值系数;但是考虑到方向与尺度之间的关系,同一个像素在不同的尺度上可以有不同的方向值.作者认为,较大尺度上的方向能够反映较大的轮廓,反之则反映较小的轮廓.这样,像

素的方向最终通过 Gabor 滤波结果的幅值与方向竞争获得.方向竞争保证每个像素只有一个方向,尺度竞争保证方向的平滑性.对于不同大小的特征块,计算方向时要选择不同的尺度,如图 4 所示.

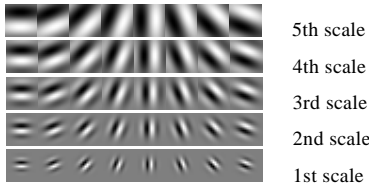


Fig.3 Visualization of Gabor filters' magnitudes
图 3 Gabor 小波幅值可视化结果

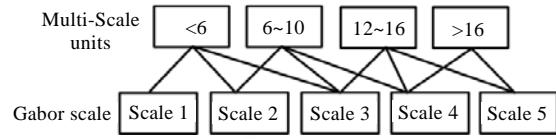


Fig.4 Units categories and their related Gabor scales
图 4 不同大小的特征块对 Gabor 尺度的选择

如图 4 所示,对于 64×128 像素的样本,特征块分成 4 类(<6 像素, $6 \sim 10$ 像素, $12 \sim 16$ 像素以及 >16 像素).每一类特征块和 3 个相邻的 Gabor 尺度相关.大小为 S 的特征块内的像素 $\theta_s(x, y)$ 的方向计算如下:

$$\theta_s(x, y) = \arg \max_u \|G_{v,u}(x, y)\|, u \in \{0, 1, \dots, 7\}, v \in \{\text{Scales of } S\} \quad (4)$$

举例解释:当计算 8×8 大小的特征块内的像素方向时,使用 Gabor 的尺度 2~4 的幅值进行竞争.最终,此过程将使得像素方向与其所在特征块的尺度关联起来,导致所提的特征是显著的,并且对噪声不敏感.

在确定像素方向以后,对于任何一个特征块都可以统计其所包含像素的方向,计算 8 方向的直方图特征:

$$h_\theta(n) = r_\theta(n) / \sum_{\theta=0}^7 r_\theta(n) \quad (5)$$

其中, $h_\theta(n)$ 是特征块 n 内方向为 θ 的像素个数; $f_\theta(n)$ 是其直方图数值,即特征块 n 内方向为 θ 的像素百分比.通过对特征块的随机采样,一个 64×128 的样本图像上可以获得 11 180 个特征块,进而提取了 11180×8 维的特征.对于如此高的维数,通过特征选取进行降维.

1.2 特征选择

MSO 粗特征选择:在粗特征块的选择中使用了 Adaboost 算法.

首先需要为 Adaboost 构建一个弱分类器:

$$C(n) = \begin{cases} 0, & \text{if } T_l(n) \leq f_n \leq T_u(n) \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

其中, $C(n)$ 是使用弱分类器对一个粗特征分类的结果; T_l 与 T_u 是弱分类器的上下两个阈值(角度值),在分类器训练过程中通过贪心搜索获得.构造公式(6)作为弱分类器的原因是,由于人体轮廓的一部分都具有一定的角度范围,每个 f_n 所代表的方向特征对人体轮廓具有较强的表达能力.一个弱分类器的含义就是:“如果一个特征块在一定范围内,则仅就此特征块而言,样本是正例(人体)样本”.在此基础上,采用一个标准的 Adaboost 算法^[7]选择 MSO 粗特征(如图 5 所示).

MSO 精特征选择:MSO 精特征选择是基于贪心搜索算法进行的.在第 t 轮特征选取过程中,新特征 u_{t+1} 的加入应该使得分类精度的提高最大.在此准则下,循环从待选特征集合 A 选出特征加入到已选特征集合 S .

$$u_{n+1} = \arg \max_{u_i \in A_n} \min_{u_j \in S_n} (I(u_i, u_j; C) - I(u_j; C)) \quad (7)$$

其中, $I(u; C) = \sum_c p(C, u) \log(p(C|u))$ 是特征块 u 的条件熵, $P(C, u)$ 表示类别 C (正例与反例)样本正确的概率.在 t 次循环选择一个特征块后,待选特征块集合 A 与已选特征集合 S 更新如下:

$$A_{n+1} = A_n \setminus u_k, S_{n+1} = S_n \cup \{u_{t+1}\} \quad (8)$$

这一特征选取过程直到分类精度不再增长时结束.假如最终获得了 N 个特征块,可以提取共 $8 \times N$ 维精特征.如图 6 所示(左侧是 8×8 ,右侧是 16×16 的特征块,4 幅图像显示了特征逐步增加的过程),选择的特征关联在一起构成了粗略的人体轮廓,且特征块之间有一定的重合,这保证了特征对于轮廓噪声的鲁棒性.根据经验,将特征块之间的最大重合度选为 50%.也就是说,对于一个待选特征而言,如果在已选特征集中存在的某个特征与其

空间重合度大于 50%,此待选特征不予考虑.

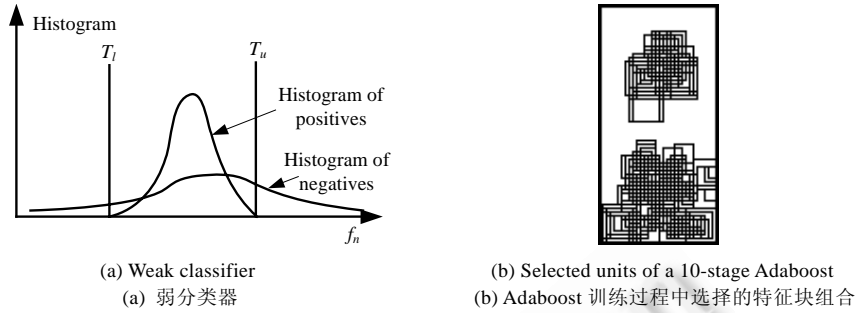


Fig.5 Unit selection for coarse features

图 5 粗特征块选择

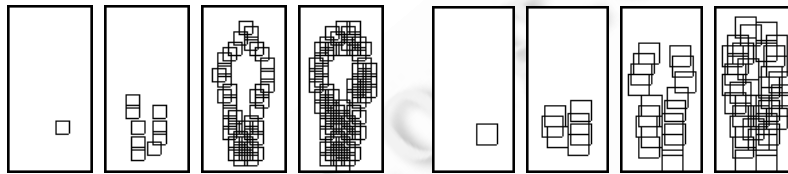


Fig.6 Fine MSO feature unit selection procedure

图 6 MSO 精特征块选取过程

2 多视角人体目标检测

在训练分类器时,采用了级联的 Adaboost 与 SVM 进行训练,其分类器框架如图 7 所示.分类器分成 5 个层.在级联分类器的最前端把所有视角样本放在一起训练一个模型,然后是 3 个、5 个到 7 个视角.在分类器的前 4 层使用了 Adaboost,最后一层使用了线性 SVM.

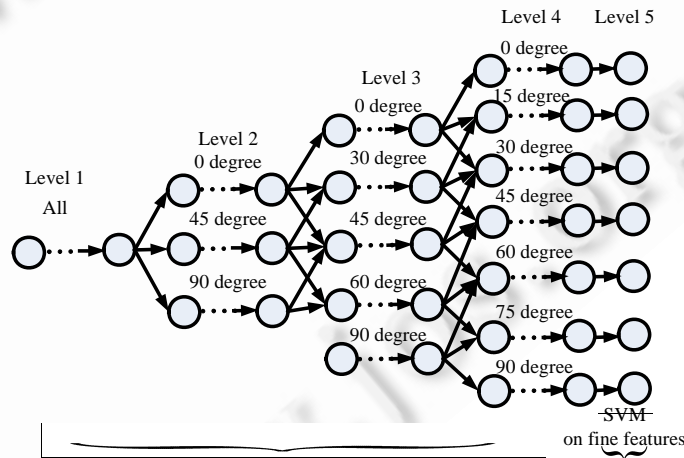


Fig.7 Cascade classifier for multi-view pedestrian detection

图 7 多视角人体目标检测的分层框架

检测时,通过逐个对图像窗口进行分类的方法判定图像中有人体目标,并确定其位置.如果一个测试窗口通过了所有的 5 层分类器,则被判定为人体目标,否则为反例.

实验结果表明,对前 4 层中的每层选择 2~3 个级联(总共 8~12 个级联)即可达到较高的检测速度及较高的查全.对于最后一层的 SVM,在使用了 255 个精特征块(255×8=2040 维特征)的情况下,线性 SVM 能够达到

99.95%的分类精度.

2.1 基于特征块位移的多视角样本生成

在图 8 中,需要用到 $0^\circ\sim 90^\circ$ 之间各个角度的样本(图 8(a)~图 8(c)为 MSO 特征合成,图 8(d)~图 8(f)为 0° 、 90° 及中间视角的 MSO 特征块位置分布). 0° 与 90° 的样本可以较精确地获取,而中间角度的样本获取较为困难.以下提出了一种特征块位置线性平移的方法对中间视角样本进行合成.

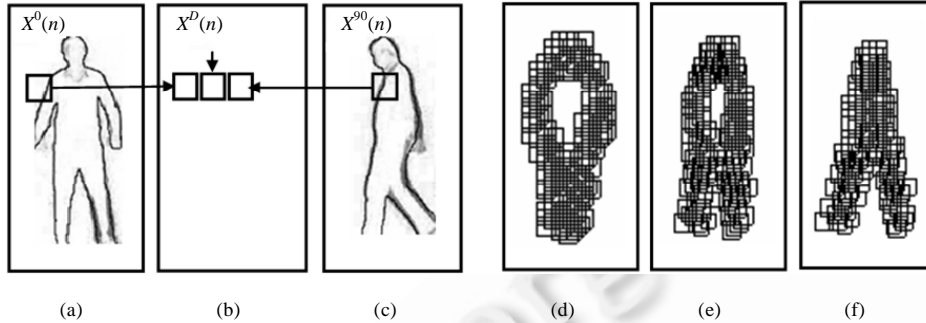


Fig.8 Sample composition

图 8 多视角的样本合成

在实际场景中,因摄像机视角的变化人体可以在 $0^\circ\sim 360^\circ$ 视角内旋转.考虑到人体的左右前后的基本对称性,人体目标视角可以简化到 $0^\circ\sim 90^\circ$ 之间.因为不进行精确的姿态识别,前后左右视角的微小差别可以忽略.假定一个 D 度视角样本的特征块 n 的位置为 $X^D(n)$,其可以由 0° 与 90° 的样本线性合成:

$$X^D(n) = \alpha(n) \cdot X^0(n) + (1.0 - \alpha(n)) \cdot X^{90}(n) \quad (9)$$

其中, $\alpha(n)$ 是合成系数.在假定人体为一个圆柱体的前提下,根据简单的三角函数可以计算:

$$\alpha(n) = \cos(D) / (\sin(D) + \cos(D)).$$

- 当 $\alpha(n)=1.0$ 时,样本就是正视角(0°)样本;
- 当 $\alpha(n)=0.0$,样本就是侧视角(90°)样本.

与位置计算方法相同,块内的特征也可以由下式合成:

$$F^D(n) = \alpha(n) \cdot F^0(n) + (1.0 - \alpha(n)) \cdot F^{90}(n) \quad (10)$$

其中, $F^D(n)$, $F^0(n)$ 和 $F^{90}(n)$ 分别表示第 n 个特征块在 D° 、 0° 与 90° 的特征.

通过上式,利用正视角与侧视角样本可以合成出中间视角(15° , 30° , 45° , 60° , 75°)的训练样本.

3 遮挡处理

人体目标检测和其他目标检测问题一样会受到遮挡的严重影响.当人体目标出现在人群中,相互遮挡问题尤为突出.一旦发生遮挡,所提取的轮廓特征就会发生局部缺失,进而导致错误的分类.人体目标之间的遮挡模型如图 9 所示.为了处理人与人之间的遮挡,我们提出特征装配补全的思路.

假定 $\tilde{f}_o(n) \in F_o$ 是第 n 个被遮挡特征块的一个特征,它可以由训练样本中的相同特征块的特征合成:

$$\tilde{f}_o(n) = \frac{1}{M} \sum_{m=1}^M w^m \cdot f_o^m(n) \quad (11)$$

其中, $f_o^m(n)$ 是训练样本集合中第 m 个样本的第 n 个特征块; M 是训练样本中的正例样本个数; w^m 表示第 m 个样本的权值向量,权值的大小表示在合成过程中第 m 个样本的重要性.计算 w^m 时定义如下规则:在未遮挡特征上与遮挡人体相似度高的样本在遮挡特征合成时具有较大的权值,公式描述如下:

$$w^m = \frac{1}{M} \sum_{m=1}^M \left(\prod_u \exp(-\|f_u(n) - f_u^m(n)\|_2) \right) \quad (12)$$

其中 $f_u(n)$ 表示待分类目标第 n 个未遮挡特征块的特征, $f_u^m(n)$ 表示样本 m 的第 n 个未遮挡块的特征.一旦确定了 w^m 并且将其进行归一化,就可以基于公式(11)计算所有遮挡特征,进而将遮挡特征补全进行分类.

如图 9(b)所示如果人体目标被多个其他人体目标遮挡,需要基于公式(11)计算多个遮挡特征集合.另外,在检测时需要首先检测前方的人体目标,然后递归的搜索被其遮挡的后面的人体目标.

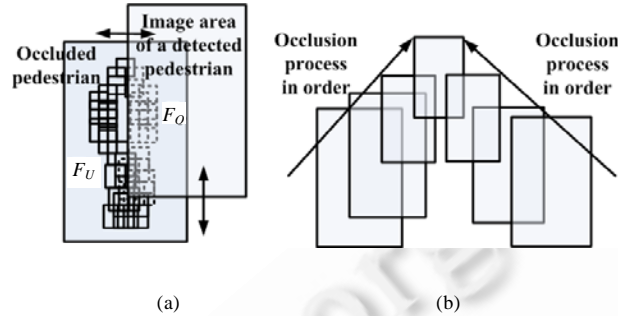


Fig.9 Pedestrian-to-Pedestrian occlusion model

图 9 人体目标之间的遮挡模型

4 实验分析

在本节中,作者对本文提出的算法在两个数据集上进行了验证,并且与现有方法进行了对比分析.

4.1 数据集

目前的训练集合主要分成了正面(0°)人体数据集合 3 500 幅及侧面(90°)人体数据 3 051 副,反例样本 5 000 幅.正例样本的收集来自 MIT 人体数据集合^[8]、INRIA 数据集合^[9]及 SDL 数据集合^[27].其正面、侧面视角的划分通过人工实现,中间视角的人体样本通过第 2.1 节的算法合成.测试集合 INRIA 是公共测试集合,SDL 多视角人体集合^[27]也是网上公开的数据集合.以下是训练和测试集合信息及示例.

Table 1 Description for training set

表 1 训练集合说明

View	Positives	Negatives
Frontal view	3500, MIT, INRIA, SDL Sets ^[27]	5000, obtained by boosting
Side view	3051, INRIA+SDL Sets ^[27]	5000, obtained by boosting

Table 2 Description for test set

表 2 测试集合说明

Name	Count	Description
INRIA Dataset ^[9]	288	Pedestrian dataset of France National Informatique and Automation Insitute. Samples in this dataset have occlusion and multi-views
SDL Dataset ^[27]	110	Pedestrian dataset of the Chinese Academy of Sciences. Samples in this dataset have multi-views



Fig.10 Training positives and negatives

图 10 正例与反例训练本举例

4.2 性能与比较

图 11 给出了一些经过多检测结果的例子.从图 11(a)~图 11(j)中可以看出,本文算法正确检测了大部分人体目标.在图 11(g)的 4 个人体目标中,漏检了复杂背景下的两个带有多姿态的人体.在图 11(j)的 3 个人体目标中,漏检了一个“侧视角两腿合并”的人体.这种人体轮廓非常容易和圆柱体形状的目标(如树干、电线杆等)混淆,不具有人体目标的典型模式.图 11(d)~图 11(f)中包含遮挡人体,采用了遮挡处理策略后,这些人体目标都被正确地检测到.

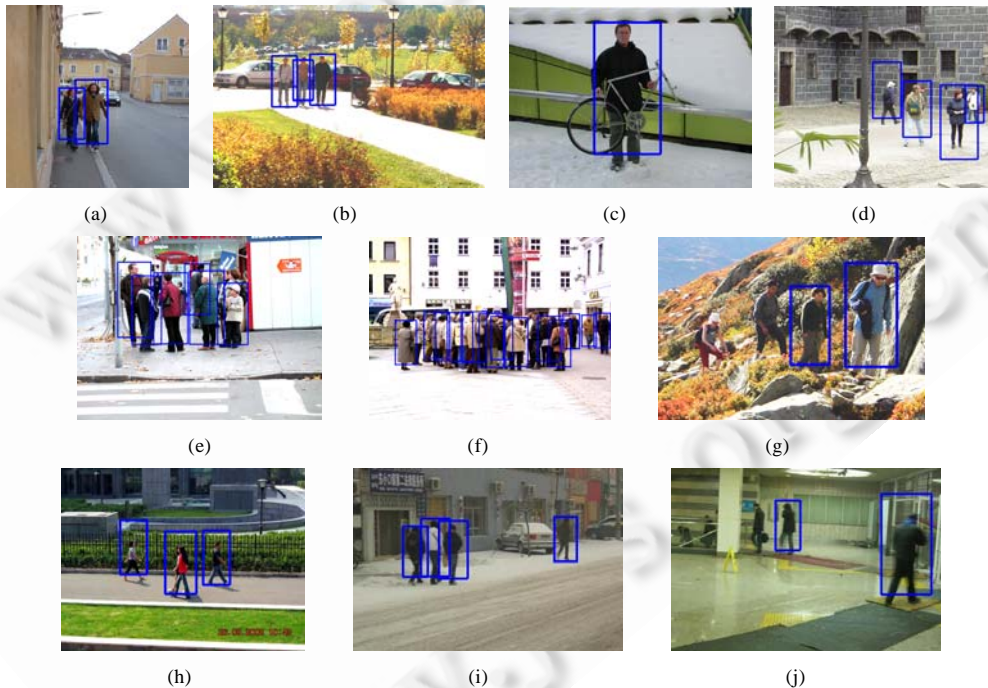


Fig.11 Detection examples

图 11 检测实例

本文用查全率(recall)、单位窗口误检率(false positives per window,简称 FPPW)作为性能指标来评测算法.

与代表性算法的比较:将本文算法与近年来的一些具代表性的人体目标检测算法进行了比较,包括 HOG 特征+SVM 分类器方法^[9]、Edgelet+COV+HOG 特征+Adaboost 分类器方法^[32]、COV 特征+Logiboost 分类器方法^[14].其中,文献[9]中的方法是根据原文献编程实现的,而其他方法都是参照其发表的文献上的结果.如图 12(a)

所示,本文算法取得了较好的效果.在 FPPW 为 10^{-4} 时,本文算法达到了 98% 的查全率(recall),比 HOG+SVM 方法高出约 8%,比 Edgelet+Adaboost 及 COV+Logiboost 方法高出约 5%.在 FPPW 为 10^{-6} 时,可以看出本文算法仍保持较高查全率.

遮挡检测:如图 12(a)所示,所提出的遮挡处理策略提高了检测性能(比不采用遮挡处理事提高大约 1.5% 的查全率).1.5% 并不是一个显著的提高,原因是 INRIA 测试集中包含了非常少的遮挡人体样例.在实验中观测的 20 个遮挡人体样例中,有 15 个被正确检测到了,而在无遮挡处理时只能检测到其中的 4 个.

多视角检测:据我们所知,目前仍然没有针对多视角人体目标检测算法的测试集合.本文作者收集了一个包含 110 幅多视角样本的测试集合^[27].在此测试集合上,把所提出的多视角检测算法与 Tree Adaboost 方法^[19]及多 HOG+SVM 方法^[9](分别对 0,30,60,90 度样本训练分类器)进行了比较.在图 12(b)可以看出,本文提出的算法是最有效的.

尺度参数选择:在特征提取过程中,实验验证了多尺度的机制会提高性能.在图 12(c)中使用了 5~7 尺度的 Gabor 滤波器效果较好,而使用 1~3 尺度时性能就会下降.这一点是本文所提的特征与 HOG 特征的不同,前者认为尺度不是决定特征性能的关键因素,而我们的特征中尺度是一个重要因素.

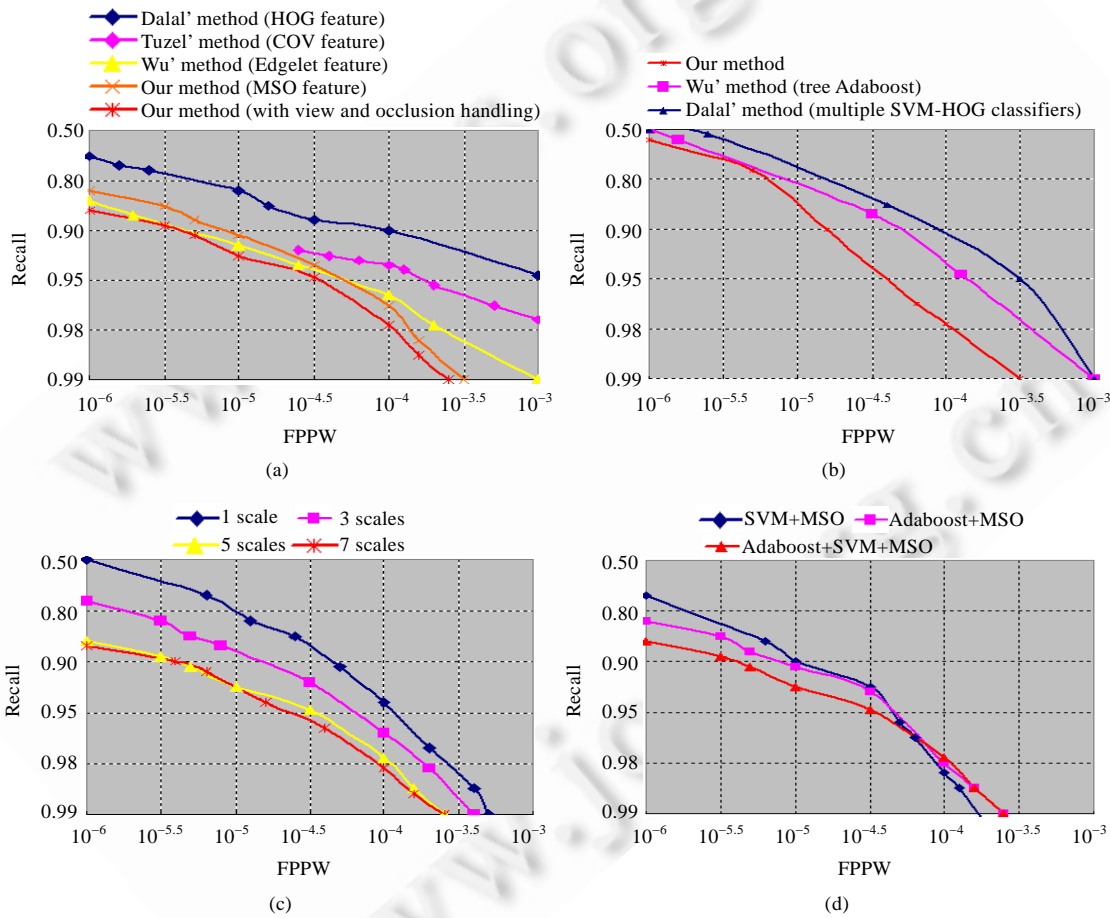


Fig.12 Performance and comparisons

图 12 检测性能及比较

分类器比较:图 12(d)比较了所提出的特征在不同分类器下的性能.可以看出,单独的级联 Adaboost 或者 SVM 分类器都比本文所采用的分类器性能低.尽管 SVM 分类器的查全率在 $FPPW > 10^{-4.5}$ 时超过了本文算法,

这对于实际应用意义不大.因为实际应用中大都要求较低的误检测率($FPPW < 10^{-5}$).

检测速度:在实际应用中,检测速度是衡量一个算法的重要指标.因为采用了由粗到精的级联分类器,许多反例窗口很快就会判别成非人体,只有极少数窗口会提取精特征并送入精分类器.算法在 Pentium IV 3.0 GHZ CPU 上对 400×500 像素大小的图像检测的平均时间大约为 0.5s,这个速度大约是 SVM+HOG 方法的 12.4 倍,印证了算法的快速性.

5 结 论

本文提出了一种用于人体目标检测的新特征表达,在此基础上研究了一种由粗到精的人体目标快速检测方法,并且对视角、遮挡问题进行了研究.其中,基于特征平移的样本合成为多视角目标检测提供了一个新的研究视点,基于特征装配补全的遮挡处理解决了人群中的目标相互遮挡问题.在大量训练样本及公共评测集合上的结果表明,所提出的算法在精度与速度上都达到了当前代表性人体目标检测算法的水平,对于复杂背景、多视角、遮挡等情况具有较好的鲁棒性.算法的不足在于不能处理任意目标对人体的遮挡,而多姿态的人体目标检测问题也没有考虑.这些问题将在未来的研究中涉及.

References:

- [1] Li L, Leung MKH. Unsupervised learning of human perspective context using ME-DT for efficient human detection in surveillance. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2008. [doi: 10.1109/CVPR.2008.4587725]
- [2] Dai CX, Zheng YF, Li X. Pedestrian detection and tracking in infrared imagery using shape and appearance. Computer Vision and Image Understanding, 2007,106(2-3):288–299. [doi: 10.1016/j.cviu.2006.08.009]
- [3] Wren CR, Azarbayejani A, Darrell T, Pentland AP. Pfinder: Real-time tracking of pedestrian body. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997,19:780–785. [doi: 10.1109/34.598236]
- [4] Ogale N. A survey of techniques for human detection from video. Technical Report, MARILAND University, <http://www.cs.umd.edu/Grad/scholarlypapers/papers/neetiPaper.pdf>
- [5] Wojek C, Walk S, Schiele B. Multi-Cue onboard pedestrian detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2010. [doi: 10.1109/CVPR.2009.5206638]
- [6] Xu R, Zhang B, Ye Q, Jiao J. Cascade L1-norm classifier for pedestrian detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2010.
- [7] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2001. [doi: 10.1109/CVPR.2001.990517]
- [8] Mohan A, Papageorgiou C, Poggio T. Example-Based object detection in images by components. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(4):349–361. [doi: 10.1109/34.917571]
- [9] Dalal N, Triggs B. Histograms of oriented gradients for pedestrian detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2005. [doi: 10.1109/CVPR.2005.177]
- [10] Zhu Q, Avidan S, Yeh MC, Cheng KT. Fast human detection using a cascade of histograms of oriented gradients. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2006. [doi: 10.1109/CVPR.2006.119]
- [11] Mu Y, Yan S, Liu Y, Huang T, Zhou B. Discriminative local binary patterns for human detection in personal album. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2008. [doi: 10.1109/CVPR.2008.4587800]
- [12] Wang X, Han TX, Yan S. An HOG-LBP human detector with partial occlusion handling. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2009. [doi: 10.1109/ICCV.2009.5459207]
- [13] Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of Edgelet part detectors. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2009. [doi: 10.1109/ICCV.2005.74]
- [14] Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Object recognition with cortex-like mechanisms. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(3):411–426. [doi: 10.1109/TPAMI.2007.56]
- [15] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on riemannian manifolds. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008,30(10):1713–1727. [doi: 10.1109/TPAMI.2008.75]
- [16] Sabzmeydani P, Mori G. Detecting pedestrians by learning Shapelet features. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2007. [doi: 10.1109/CVPR.2007.383134]
- [17] Bileschi SM. A multi-scale generalization of the HoG and HMAX image descriptors for object detection. Technical Report, Cambridge: MIT, 2008. <http://dspace.mit.edu/bitstream/1721.1/41093/1/MIT-CSAIL-TR-2008-019.pdf>

- [18] Munder S, Gavrilu D. An experimental study on pedestrian classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(11):1863–1868. [doi: 10.1109/TPAMI.2006.217]
- [19] Wu B, Nevatia R. Cluster boosted tree classifier for multi-view, multi-pose object detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2007. [doi: 10.1109/ICCV.2007.4409006]
- [20] Ye QX, Jiao JB. Multi-Posture pedestrian detection in video frames by motion contour matching. In: Proc. of the IEEE Asia Conf. on Computer Vision. 2007. [doi: 10.1007/978-3-540-76386-4_85]
- [21] Lin Z, Davis L, Doermann D, DeMenthon D. Hierarchical part-template matching for pedestrian detection and segmentation. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2007. [doi: 10.1109/ICCV.2007.4408975]
- [22] Enzweiler M, Dariu M. Gavrilu monocular pedestrian detection: Survey and experiments. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(12):2179–2195. [doi: 10.1109/TPAMI.2008.260]
- [23] Wu B, Nevatia R, Yuan L. Segmentation of multiple, partially occluded objects by grouping, merging, assigning part detection responses. In: Proc. of the CVPR 2008. [doi: 10.1109/CVPR.2008.4587750]
- [24] Andriluka M, Roth S, Schiele B. People tracking by detection and people detection by tracking. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2008. [doi: 10.1109/CVPR.2008.4587583]
- [25] Sharma V, Davis JW. Integrating appearance and motion cues for simultaneous detection and segmentation of pedestrians. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2007. [doi: 10.1109/ICCV.2007.4409035]
- [26] Eshel R, Moses Y. Homography based multiple camera detection and tracking of people in a dense crowd. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2008.
- [27] <http://coe.gucas.ac.cn/SDL-Homepage/resource.html>
- [28] Gao W, AI HZ, Lao SH. Adaptive contour features in oriented granular space for human detection and segmentation. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2009. [doi: 10.1109/CVPR.2009.5206762]
- [29] Li TZ, Ding XQ, Wang SJ. Human detection with a coarse-to-fine cascade linear SVM. Journal of Image and Graphics, 2008,13(3):566–570 (in Chinese with English abstract).
- [30] Liu YZ, Shan SG, Zhang WC, Chen XL, Gao W. Granularity-Tunable gradients partition (GGP) descriptors for human detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2009. [doi: 10.1109/CVPR.2009.5206724]
- [31] Sun QJ. Study on Human Detection in a Static Image [Ph.D. Thesis]. Beijing: Graduate University, the Chinese Academy of Sciences (Institute of Software), 2004 (in Chinese with English abstract).
- [32] Wu B, Nevatia R. Optimizing discrimination-efficiency tradeoff in integrating heterogeneous local features for object detection. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. 2008. [doi: 10.1109/CVPR.2008.4587749]

附中中文参考文献:

- [29] 李同治,丁晓青,王生进.利用级联 SVM 的人体检测方法.中国图像图形学报,2008,13(3):566–570.
- [31] 孙庆杰,吴恩华.静态图像中人体检测技术研究[博士学位论文].北京:中国科学院软件研究所,2004.



叶齐祥(1978—),男,河南虞城人,博士,副教授,主要研究领域为模式识别,图像处理.



蒋树强(1977—),男,博士,副教授,主要研究领域为多媒体检索技术.



焦建彬(1969—),男,博士,教授,博士生导师,主要研究领域为模式识别,智能系统.