

文章编号 1004-924X(2013)11-2906-08

多部位集合的人体检测

黄 炎, 颜 佳*, 张 虎, 邓德祥

(武汉大学 电子信息学院, 湖北 武汉 430079)

摘要:采用启发式、有监督的部位筛选方法组成了一种多部位集合的检测模型,用于缓解遮挡和形变对人体检测造成的影响。该模型通过比较人体同部位上关节间的 Procrustes 距离,在训练集中获取有着相似姿态的同部位样本;将梯度方向直方图(HOG)作为特征,由典型部位分别训练出判别模型;比较其在验证集上的检测效果,从中筛选出检测率高的部位和未检出的图片,再寻找对未检出图片检测率高的部位,由这些部位组成混合模型。用混合模型检测时,由 Kullback-Leibler 距离判断各部位在图片上的不同响应是否属于同一人,以此来确定人体的外接矩形框。在 INRIA 人体库上的测试表明,本文采用的模型在误检率(FPPI)为 0.5 时有 81% 的检测率,高于有 77% 检测率的 Poselets 算法。本文基于 Poselets,结合 HOG 的特点采用了一套有监督的部位筛选机制,使得模型成员数大幅度减少,检测时间比原始方法下降了 50%,同时取得了优于 Poselets 的检测效果。

关键词:梯度方向直方图;部位模型;支持向量机;人体检测

中图分类号:TP391 **文献标识码:**A **doi:**10.3788/OPE.20132111.2906

Poselets for pedestrian detection

HUANG Yan, YAN Jia*, Zhang Hu, DENG De-xiang

(School of Electronic Information, Wuhan University, Wuhan 430079, China)

* Corresponding author, E-mail: yanjiaapple@tom.com

Abstract: A part-based mixture model (Poselets) was utilized to alleviate the problem of pedestrian detection under occlusion and articulation. Firstly, Procrustes distances between similar pedestrian parts were calculated to gather training examples with the same configurations. Then, common parts of frontal pedestrian described by Histogram of Oriented Gradient (HOG) were trained to get discriminative models. A test was carried out on a validation set to find out which parts were more accurate, and those pictures which were not detected formed the hard set. Other complementary parts were explored on the hard set, afterwards, all these parts formed the final Poselets detector. While detecting, Poselets were clustered if the symmetrized KL-divergence between two Poselet activations were small. The bounding box of a pedestrian was inferred by the bounding boxes of the Poselets in a cluster. Based on the test of the INRIA pedestrian dataset, it is showed that the detection rate increases from 77% of Poselets to 81% in this paper while False Positive Per Image (FPPI) is 0.5. It is concluded that the part selection mechanism proposed in this paper promotes the detection rate with speedup rate about twice of the traditional method, meanwhile it reduces the number of models.

Key words: Histogram of Oriented Gradient (HOG); poselets; support vector machine; pedestrian detection

收稿日期:2013-05-13;修订日期:2013-06-18.

基金项目:国家自然科学基金面上项目(No. 61072135)

1 引 言

人体检测是目标检测中的一个典型问题,在车载行人保护系统、安防领域有着广泛应用。由于人体姿态多变且易受遮挡,故这种非刚体检测存在较大的困难。近年来,基于部位的建模算法能较好地克服遮挡和形变的问题,因此,成为主流的检测方法。

传统的整体模型法中的经典算法是 Dalal 等人提出的梯度方向直方图(Histograms of Oriented Gradient, HOG)^[1]特征结合支持向量机的方法,即将人体检测看作是将人体和背景进行分类的问题,但该方法对遮挡、形变的检测性能不佳。部位法是在标注出人体各部位的基础上,建立部位模型(part-based model)。部位法采用“分而治之”的策略,克服了整体模型无法准确描述多种人体姿态的不足。如 Mikolajczyk^[2]等人将人体分为 7 个部位,对每一部位用类尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)特征和 AdaBoost 分类器建立起单独的检测器,将各部位间的位置关系看作是一个概率图模型;Lin 和 Davis^[3]将人体的头部、上身和腿组成树模型,将这些部位的位置、尺寸看成树模型的参数,通过求取部位匹配函数的最大似然来确定人体的位置;Andriluka 等将描述外形的判别模型和描述人体结构的产生式模型结合起来,同时进行人体检测和姿态判断^[4]。以上方法都是基于人体的分解结构,受限于固定的部位划分。随后,Dollar 等将人体随机地划分成部位集合,用多实例学习(Multiple Instance Learning, MIL)^[5]的方法来求解这种集合分类问题,该方法不需对部位进行标注,因此,它不能反映人体部位的语义。

近年来,许多算法都是以 Felzenszwalb 提出的形变部位模型(Deformable Part-based Model, DPM)^[6]为基础,通过计算检测窗口图像的低分辨率整体模型、高分辨率的部位模型间的匹配度及部位的位置偏移损失来判断窗口中是否包含人体。但其模型训练过程复杂,且自动寻找到的部位的位置常出现偏移。

Bourdev 和 Malik^[7]提出了一种穷举姿态的字典‘Poselets’,该算法在 PASCAL VOC^[8]比赛中取得了领先的成绩,其存在的问题在于:需同时

检测 150 种部位,从而存在大量的时间开销;另外,150 种部位里包含着重复或检测率低的部位。基于此,本文采用了一种启发式的、有监督的部位筛选方法组成模型,大大降低了模型的训练复杂度和算法时间复杂度,而检测率依然很高。

2 部位定义与样本获取

本文采用的训练样本来自 H3D^[9]人体库,该数据库有着详细的人体关节标注(20 个主要关节),如图 1 所示,关节点的位置分布代表了人体不同的姿态。



图 1 本文采用的训练样本

Fig. 1 Training sample

部位的定义不限于生理学结构,图 2(a)来自图 1 中的黑框裁剪区域。需要说明的是:某种姿态的人体部位即称为一种 Poselet;图 2(b)显示的是以图 2(a)作为种子,在样本集中按式(1)的方法寻找到的结构相似样本,这一过程仅用到了关节点坐标,与图片的灰度信息无关。

$$D(P_1, P_2) = D_{\text{proc}}(P_1, P_2) + \lambda D_{\text{vis}}(P_1, P_2), \quad (1)$$

$$D_{\text{proc}}^2 = \sum_{j=1}^n [(x_{j1} - x_{j2})^2 + (y_{j1} - y_{j2})^2], \quad (2)$$

$$D_{\text{vis}} = \frac{\sum srckeypt \cdot dstkeypt}{\sum srckeypt / dstkeypt}. \quad (3)$$

式(2)中, D_{proc}^E 是种子和候选样本关节点间的 Procrustes 距离,即关节点通过仿射变换后的差值,也是形变距离; n 表示关节点数,下标 1、2 表示两幅图。式(3)中 srckeypt、dstkeypt 数组分别

表示种子样本、候选样本上关节点的可见性 D_{vis} 是视觉距离,表示两种可见性的重合率。式(1)中, λ 是比例常数,起到调整两种距离权重的作用,影响着样本相似度的排序。当 λ 取 0.1 时,图 2(b)与种子的相似度高于图 2(c);当 λ 取 0.5 时,则相反。原因是图 2(b)在姿态空间与种子更接近,即 D_{proc} 更小,但部分关节点不可见,即 D_{vis} 更大。

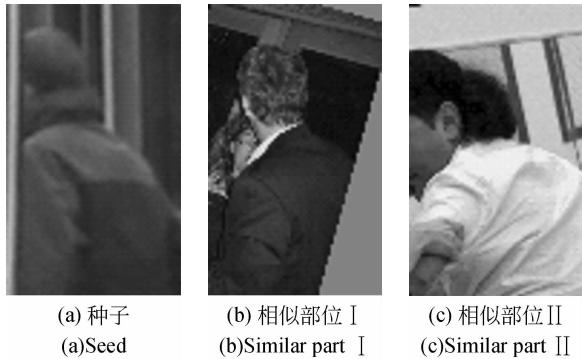


图 2 寻找相似部位

Fig. 2 Looking for similar parts

如图 2 所示, Poselet 会保留原始图片(图 1 方框中)所有关节点、外接矩形归一化后的坐标信息。对同种 Poselet 的所有原始图片关节点分别建立高斯模型,即每个关节点有一个平均位置和一个分布区域,外接矩形也类似。这些统计信息为后续的检测过程服务。

3 部位选取

Poselets 是由多种人体部位组成的混合模型,部位的数量越多,对人体各种姿态的覆盖就越全面,但难免形成冗余。如何选取 Poselets 的数量是一个典型的集合覆盖问题,而集合覆盖是一个 NP(Non-deterministic Polynomial time)困难的问题,通常采用贪婪解法。文献[10]从 2 000 种随机选取的部位中挑选出 150 种符合要求的部位组成模型。而本文从已有的先验知识出发,结合 HOG 特征的特点,采取了有监督的选取方式,大大简化了这一过程。本文采取的准则如下:

1. 每种 Poselet 有较高的检测率

这一准则要求选取的部位要合适,且该部位要有足够多的样本来保证分类器的训练效果,并使其对普遍样本的检测率高。

2. Poselet 之间互相补充,尽可能的对验证集形成覆盖,且种类数要少

这一准则要求模型中要有对准则 1 中未检出的样本检测率高的 Poselet,即覆盖特定的样本。

4 检测过程

Poselets 集合对图片进行检测时,其中的每一种 Poselet 都会产生单独的响应,而如何判定这些响应是否发生在同一个人身上成了检测过程的关键。

本文算法的检测过程如图 3 所示。图 3(a)显示了两种 Poselets 在原图不同金字塔层次上的响应(按比例还原到原图显示),红色表示膝盖以上正面人体 Poselet 的响应,绿色表示正面腿部 Poselet 的响应(彩图见期刊电子版);图 3(b)显示的是经过非最大抑制后的响应,可以看到图 3(a)在多层金字塔图片上产生的响应融合成了集群。图 3(c)显示了这两种 Poselet 关节点的平均位置,这一位置来源于第 2 节中的统计值。

每种 Poselet 的响应实际是一种对关节点分布位置的估计, Kullback-Leibler 距离正是对两种概率分布相似性的判断。KL 距离的表达式如下:

$$D_{KL}(p_1, p_2) = \int p_1(x) \cdot \log \frac{p_1(x)}{p_2(x)} dx, \quad (4)$$

$$D_{SKL}(N_i^k, N_j^k) = D_{KL}(N_i^k \parallel N_j^k) + D_{KL}(N_j^k \parallel N_i^k), \quad (5)$$

$$d_{i,j} = \frac{1}{K} \sum_k D_{SKL}(N_i^k, N_j^k). \quad (6)$$

式(4)表示 P_1, P_2 两种概率分布间的 KL 距离,式(5)是将 KL 距离的不对称性加入到距离测度中,式(5)中的 N 代表高斯分布, i, j 表示两种 Poselets 的响应, k 表示关节点数。式(6)表示两种 Poselets 中 K 个关节点间的平均距离,即两种 Poselets 间的 KL 距离^[10]。

图 3(d)显示了图 3(c)中两种 Poselets 发生在同一个人身上的响应。这一过程在采用多种 Poselets 对人群检测时极为重要,防止了将多个人判断为一个人的情形,后文中对人群检测结果也表明了这点。图 3(e)显示了人体的外接矩形,这一矩形框位置同样来自第 2 节末尾所述的统计量。图 3(f)显示了利用均值漂移算法(Mean Shift, MS)对同一人进行外接矩形融合的结果。



(a) 金字塔结构上的多层响应
(a) Multilayer response on pyramid structure



(f) 均值漂移算法对同一人的融合
(f) Fusing to the same person by mean shift algorithm



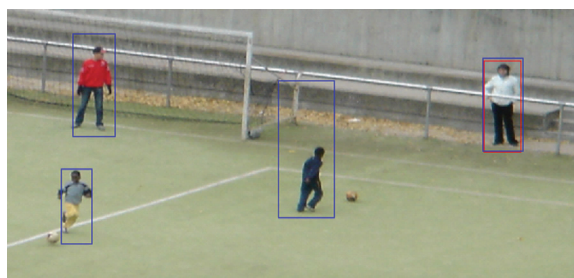
(b) 非最大化抑制
(b) Non-maximized suppression



(c) 每种 poselets 的关节点分布
(c) Joint point distribution of each poselets



(d) 两种 poselet 关节点分布间的 KL 距离
(d) KL distance between two kinds of poselets joints



(e) Poselet 响应对应的人体矩形框
(e) Body rectangular frame corresponding to poselets response

图 3 检测过程

Fig. 3 Detecting procedure

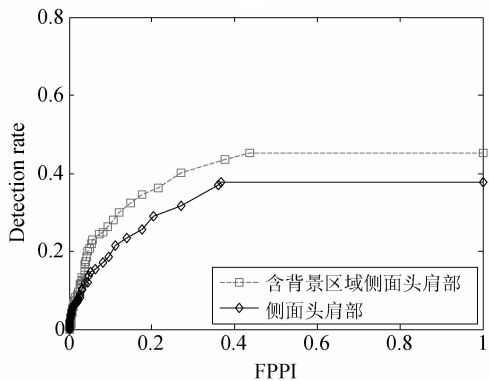
5 实验过程及结果分析

5.1 样本裁剪

本文的训练样本来自 H3D 库,共 2 480 个标注,在这些图片中裁剪出所需的各种 poselet 部位。相比原始算法的随机裁剪,本文更有针对性地利用了 HOG 的特点,即为人体周围预留背景区域,以便利用人体与背景过渡区域的梯度信息,图 4(a)与图 4(b)相比,就缺少这种信息。文献 [11]中提到后一种裁剪方式的检测率提升了约 4%,图 4(c)比较了使用这 2 种方式裁剪同一部位时的检测性能。



(a) 裁剪示意 I
(a) Cropping illustration I
(b) 裁剪示意 II
(b) Cropping illustration II



(c) 各部位在 INRIA Person 训练库上的检测结果
(c) Detection results of each part on INRIA Person training library

图 4 2 种裁剪方式对性能的影响

Fig. 4 Effect of cropping methods on detection performance

5.2 部位训练

采用 HOG 特征描述样本,该特征先计算每个像素点的梯度幅值和角度($0\sim 180^\circ$),将角度划分为 9 个盒子,以 8×8 的子块为单元统计梯度方向直方图。头肩部 Poselet 的归一化尺寸为 64×96 ,上半身人体的尺寸为 96×64 ,对应的特征向量的长度均为 2 772;完整的行人归一化尺寸为 128×64 ,特征向量长度为 3 780。

在不含人体的图片中随机裁剪出上述尺寸的负样本,其数量将远大于正样本,故发掘困难的负样本(易造成误判的样本)显得尤为重要,本文采取和文献[12]一致的 Bootstrapping 过程:首先用随机负样本和正样本训练出初始模型,再用该模型发掘出困难的负样本,如图 5 所示,困难的负样本多是树木、建筑等与人体较为类似的物体,最后,将困难负样本加入到随机负样本中,再次训练得到新的模型。



图 5 困难的负样本

Fig. 5 Hard negative examples

5.3 部位选择

由于 INRIA 训练库没有关节标注,不能作为本文的训练样本,故本文将作为部位选择的验证集。首先验证了脸部、头肩、上半身、膝盖以上、腿部以及全身在 INRIA 训练库上的检测效果,如图 6 所示,从图 6 可以看出,正面样本(90° 左右)占多数,为评价不同部位在同一角度下的检测性能,全部选取正面样本进行检测,检测结果如图 7 所示。

由图 7 可以看出,膝盖以上和整个人 2 个正面样本的检测性能最好,头肩部和上身 2 个正面样本的性能次之。与 Bourdev 等人的结论不同的是,腿部在检测中的结果要好于脸部的检测结果,

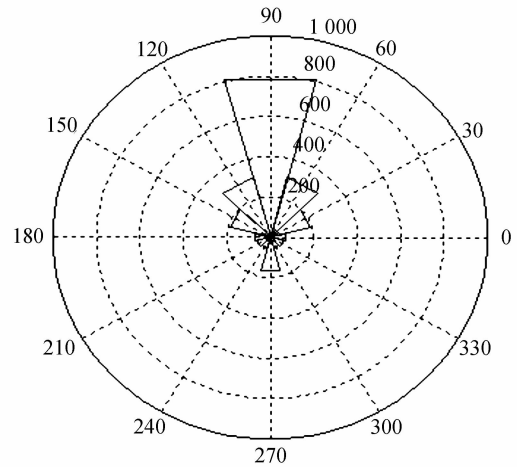


图 6 对样本库视角的统计

Fig. 6 Statistical of perspective for sample library

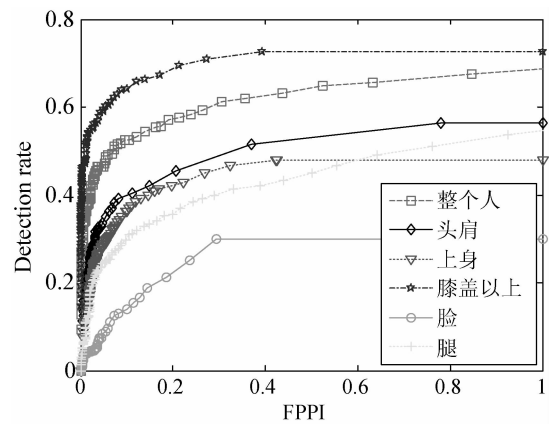


图 7 各部位的检测结果

Fig. 7 Detection results of various parts

脸部检测效果最差。原因是验证集中多是行人,大多是站立或正在行走的样本,大多为保持近似直立的姿势,而 PASCAL 数据库中,较多的是坐着的姿势,它们有着明显的区别,且 PASCAL 数据库中的人体的分辨率较高,脸部区域较大,而本文采用的验证集图片大小均为 128×64 ,脸部区域模糊。至今为止,未有文献表明单独使用 HOG 算子可以在人脸检测中取得良好效果,故本文推断,在小分辨率图片中,基于 HOG 特征的人脸检测不适用于人体检测。

为满足第 3 节中的第 1 准则,本文选取上步中检测率排序靠前的几种 Poselets,如膝盖以上、全身、上半身等;再选取以上部位的背面 Poselet,及以上部位旋转角度在 30° 以上的 Poselets,这一角度值的选取基于 HOG 特征抗角度旋转的参数。由以上选取的相应部位的不同角度的 Pose-

lets 组成了本文的高检测率部位模型。

对上文中在验证集上的检测结果进行统计,筛选出从没被覆盖的样本集,称为困难集。这一过程通过建立覆盖表来实现, C_{ij} 表示类型 i 的 Poselet 在验证集样本 j 上的无正确检测结果;若 C_{ij} 等于验证集样本数目,则表明从未被覆盖。图 8 显示了筛选出的从未被覆盖的样本,多是四肢形变较大或几乎完全侧立的人体。对于前者,本文通过裁剪出不包含手臂的上半身 Poselet 来解决这一问题;对于后者,由于侧立人体关节左右对称,大量重复,通过关节分布相似度的样本聚类很难找到足够多的类似样本。从图 6 对样本库视角的统计中可以看出,完全侧立的人体 (0° 、 180°) 样本数目很少,本文采取只截取头肩部的方法。由于样本的尺寸越小、得到的特征向量也越短,训练时要求的样本数目也就降低了。图 4 中的侧面头肩部的检测性能低于图 7 中的正面头肩部的检测性能,但其提高了高检测率模型的覆盖率。几种方法以及最终模型与高检测率部位模型的对比结果如图 9 所示。由图 9(b) 可以看到,最终模型较高检测率模型的性能有所提升。本算法将一个对大数据集的覆盖问题转换为一个对部分数据集的覆盖问题,减小了贪婪式搜索的复杂度。



图 8 筛选出的未覆盖的样本

Fig. 8 Uncovered samples

5.4 检测性能分析

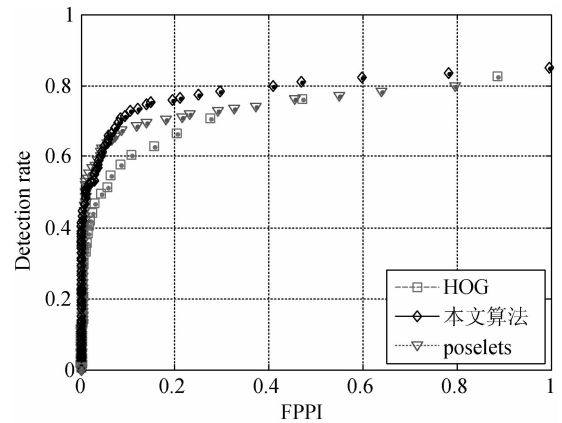
测试库包含 288 张图片,检测过程如前所述。采用 DET(Detection Error Tradeoff) 曲线来评价检测器的性能,即统计不同的阈值设定在测试样本上的检测率 (Detection Rate) 和误检率 FPPI (False Positive Per Image)。检测结果的正确性由 PASCAL VOC 准则^[8] 决定,即检测框 BB_{dt} 与真实框 BB_{gt} 的重合率 a 大于 0.5 才算正确,如下式所示:

$$a = \frac{area(BB_{dt} \cap BB_{gt})}{area(BB_{dt} \cup BB_{gt})} > 0.5, \quad (7)$$

$$\text{检测率} = \frac{\text{正确检出的正样本数}}{\text{所有正样本数}} \quad (8)$$

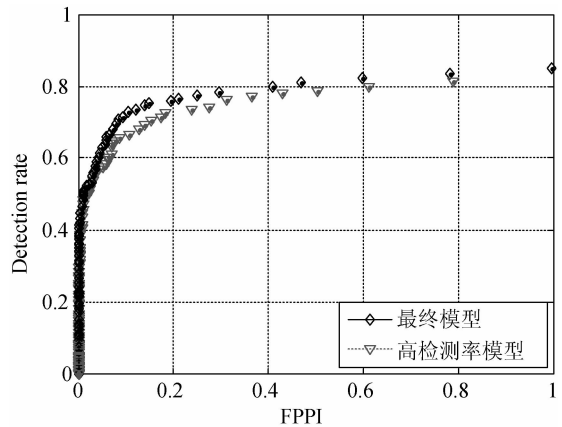
$$\text{误检率} = \frac{\text{错误的正检测结果}}{\text{所有图片数}} \quad (9)$$

实验结果曲线如图 9(a) 所示, HOG 算法作为基准,在低误检率时, Poselets 算法和本文算法显著优于 HOG 算法。原因在于本文和 Poselets 均采用了混合模型,比单一模型覆盖面更广,即使误检率较低时也可检测出各种姿态的人体。随着误检率升高, HOG 算法性能提升较快,而本文算法和 Poselets 的差异开始显现。这是由于本文采用的特征和分类器与 HOG 算法一致,混合模型中的单个元素性能与 HOG 也无本质区别,而误检率正比于元素个数,随着误检率的升高,混合模型带来的优势被削弱了。在 FPPI=0.5 时,本文算法较 Poselets 算法的检测率提升约 4%,原因在于本文算法剔除了部分检测率较低的部位,如



(a) 3种算法检测结果的对比

(a) Comparison of detection results for 3 algorithms



(b) 部位选择两步间的对比

(b) Comparison of two stages

图 9 实验结果 (DET 曲线)

Fig. 9 Experimental results (DET curves)

脸部、腿部。另外,其裁剪的部位居中,更符合 HOG 特征的要求。算法在文献[10]提供的源码基础上进行修改,在 Matlab 2010 环境下进行,计算机为 2.1 G 双核、内存为 4 G。对 60 万像素的图片检测时间为 3.9 s,聚类时间为 0.06 s;而原始 Poselets 算法的检测时间为 7.3 s,聚类时间为 0.6 s。图 10 依次展示了本文算法对人群、形变人体、不同视角人体、侧立人体的检测结果。



图 10 部分检测结果

Fig. 10 Some detection results

参考文献:

- [1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. *International Conference on Computer Vision & Pattern Recognition*, 2005, 1: 886-893.
- [2] MIKOLAJCZYK K, SCHMID C, ZISSERMAN A. Human detection based on a probabilistic assembly of robust part detectors [C]. *European Conference on Computer Vision (ECCV04)*, 2004, 3021: 69-82.
- [3] LIN Z, DAVIS L S. A pose-invariant descriptor for human detection and segmentation [C]. *European Conference on Computer Vision*, 2008, 5305: 423-436.
- [4] ANDRILUKA M, ROTH S SCHIELE B. Discriminative appearance models for pictorial structures [J]. *International Journal of Computer Vision*, 2012, 99(3): 259-280.
- [5] DOLLAR P, BABENKO B, BELONGIE S, et al.. Multiple component learning for object detection [C]. *Computer Vision-ECCV 2008*, 2008, 5303: 211-224.

6 结 论

Poselets 通过对样本进行详细标注和用先对样本在姿态空间聚类,再进行训练的方法,保证了同类样本的一致性。检测过程中依据概率模型而不是重叠面积的聚类,很好地处理了模型中多种 Poselets 对同一目标响应不同的融合问题。本文首先指出了 Poselets 算法中的冗余和存在疑问之处——小分辨率图片下由 HOG 特征描述的人脸 Poselet 不适合于人体检测;接着,采取了一种启发式有监督的选择 Poselets 集合方法,仅用原算法 1/12 的模型就达到了更佳的效果。通过对 60 万像素的图片的检测结果显示,本文算法的检测时间较原始 Poselets 算法的检测时间提高了近一倍,而聚类时间仅为原始 Poselets 算法的 1/10;而在 Poselets 算法中加入上下文信息如多种 Poselet 之间的位置关系,可以减小误检,这将是本文去进一步探索的问题。

- [6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D. Cascade object detection with deformable part models [C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010: 2241-2248.
- [7] BOURDEV L, MAJI S, BROX T, et al.. Detecting people using mutually consistent poselet activations [C]. *Computer Vision-ECCV 2010*, 2010, 6316: 168-181.
- [8] EVERINGHAM M, VAN G L, WILLIAMS, et al.. The PASCAL Visual Object Classes (VOC) Challenge [J]. *International Journal of Computer Vision*, 2010, 88(2), 303-338.
- [9] BOURDEV L, MALIK J. Poselets: Body part detectors trained using 3D human pose annotations [C]. *2009 IEEE 12th International Conference on Computer Vision*, 2009: 1365-1372.
- [10] BOURDEV L. *Poselets and their applications in high-level computer vision* [D]. California: Computer Science Graduate Division of the University of California Berkeley, 2011.
- [11] DALAL N. *Finding people in images and videos*

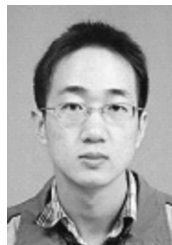
[D]. France: the French National Institute for Research in Computer Science and Control, 2006.

- [12] 黄炎,范赐恩,朱秋平,等. 联合梯度直方图和局部二值模式特征的人体检测[J]. 光学精密工程,

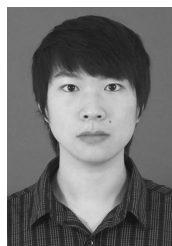
2013, 21(4):1047-1053.

HUANG Y, FAN C E, ZHU Q P, *et al.*. HOG-LBP pedestrian detection [J]. *Opt. Precision Eng.*, 2013, 21(4): 1047-1053. (in Chinese)

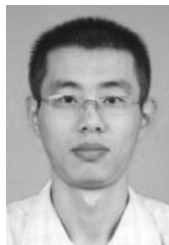
作者简介:



黄 炎(1988—),男,湖北荆州人,博士研究生,2008年于武汉大学获得学士学位,主要从事机器视觉、模式识别等方面的研究。E-mail: up2mail@163.com



颜 佳(1983—),男,汉族,湖北天门人,博士后,2010年于武汉大学获得博士学位,主要从事目标跟踪、机器视觉等方面的研究。E-mail: yanjiaapple@tom.com



张 虎(1986—),男,天津人,博士研究生,2009年于武汉大学获得学士学位,主要从事图像处理、机器视觉等方面的研究。E-mail: zhanghu0722@126.com

导师简介:



邓德祥(1961—),男,湖北荆州人,教授,博士生导师,1982年、1985年于武汉测绘科技大学分别获得学士、硕士学位,主要从事空间图像处理,片上系统的研究工作。E-mail: ddx@eis.whu.edu.cn

(版权所有 未经许可 不得转载)