

# 基于部件模型及颜色信息的行人检测

赵金金,姚汉利,鲍文霞

(安徽大学 电子信息工程学院,安徽 合肥 230601)

**摘要:**行人识别是人工智能与模式识别领域内一个新兴的研究方向,具有极其广泛的应用前景。但是由于人体是一个非刚性的运动体,相对普通物体的检测增加了不少难度。可变形部件模型算法对行人检测有着不错的效果,在此基础上提出了一种对传统的部件模型的改进方法,弥补了颜色特征在行人检测时的丢失。其基本思想是:使用传统的 DPM 方法对待检测窗口进行检测,然后判断检测的得分是否属于可疑区间,如果属于则进一步使用基于颜色特征的分类器对可疑区域进行检测,判断结果由两次的决策值共同决定。在 INRIA 数据库的检测结果表明,基于多决策的行人检测方法能够在几乎不影响检测速度的同时提高检测准确率,为精准地对图片或视频中的行人做进一步的分析提供了有利的基础。

**关键词:**行人检测;色彩空间;可变形部件模型;可疑区间;多决策

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2017)11-0058-04

doi:10.3969/j.issn.1673-629X.2017.11.012

## Pedestrian Detection Based on Part Model and Color Information

ZHAO Jin-jin, YAO Han-li, BAO Wen-xia

(School of Electronics and Information Engineering, Anhui University, Hefei 230601, China)

**Abstract:** Pedestrian recognition is an emerging research in artificial intelligence and pattern recognition, and owns the extremely wide-spread application prospect. However, because the human body is a non-rigid body motion, it increases a lot of difficulty compared with ordinary objects detection. Deformable Part Model (DPM) algorithm has a good effect on pedestrian detection. On the basis of that, an improved algorithm for the traditional DPM is presented to makes up for the loss of color features in the pedestrian detection. Its thought is following: using the traditional DPM for detection of window, then judging whether the classification decision value belongs to the suspicious interval or not. If it does, the classifier based on RGB feature will make the further classification on characteristics, and the results are decided by the two decision values jointly. The experimental results in INRIA database show that the proposed algorithm can raise the detection accuracy without impact on detection speed, and provide the basis for further analysis of pedestrians in pictures or videos.

**Key words:** pedestrian detection; color space; deformable part model; suspicious interval; multiple decisions

### 1 概述

行人目标的识别与检测技术是对行人做视觉分析的基础,只有在此基础之上人们才可以进一步对人体行为理解、高级人机接口、人体运动分析、智能监控以及智能驾驶系统<sup>[1]</sup>等领域进行深入研究。因此,行人检测逐渐成为模式识别和计算机视觉领域中的研究热点。

行人检测的主要内容是利用计算机视觉技术,在无人干预的情况下,对视频图像进行分析和处理,并从中提取有用信息,实现对动态场景中行人的定位、跟踪和识别。但是由于人体是非刚性的,四肢和各个关节

可以呈现很多角度的转动及伸展,所以人体可以呈现的姿态千变万化,在检测时如果用单一固定的检测模板通常效果并不理想。目前行人检测的方法大多数是基于机器学习,其主要包含两个方面,一个是特征描述算子,另一个是学习方法。对于人体特征通常又分成三大类<sup>[2]</sup>:底层特征、组合特征和深度学习特征。底层特征是指边缘、纹理和颜色等图像的基本特征;组合特征是多种底层特征的组合,或是底层特征的高阶统计特征;深度学习特征是指通过深度学习从图像原始数据中学习到的特征。特征描述算子 HOG<sup>[3]</sup>是当前使用特别广泛的底层特征<sup>[4-6]</sup>,它刻画了图像局部梯度

收稿日期:2016-12-28

修回日期:2017-04-28

网络出版时间:2017-08-01

基金项目:国家自然科学基金青年基金(61401001)

作者简介:赵金金(1993-),男,硕士研究生,CCF 会员(70493G),研究方向为图像处理和计算机视觉;鲍文霞,副教授,通讯作者,研究方向为计算机图形学和计算机视觉。

网络出版地址: <http://cnki.net/kcms/detail/61.1450.TP.20170801.1559.082.html>

值和方向。为了能够较好地让模型去适应物体的变化,可变形部件模型<sup>[7]</sup>是在 HOG 的基础上添加部件模型的一种改进算法。由于部件模型算法在进行行人识别时对颜色空间信息的丢失,提出了利用颜色空间信息 RGB 与 DPM 算法相结合的行人检测算法,增加 RGB 检测可以在一定程度上提高检测的准确性。

具体的算法流程如图 1 所示。其中,  $S_0$  为部件模型检测得分;  $T$  为阈值;  $sp$  为可疑区间;  $S_1$  为颜色空间检测得分。

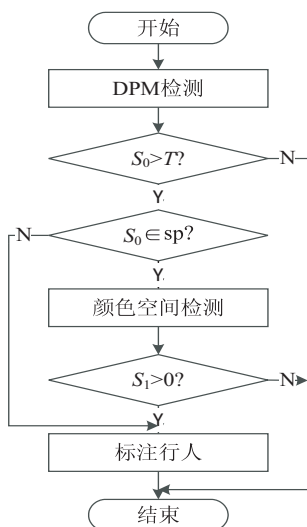


图 1 算法流程

## 2 基于 DPM 模型的目标识别

可变形部件模型(DPM)是 Felzenszwalb 等提出的一种非常成功的目标检测算法,曾在 PASCAL VOC 中连续获得 07,08,09 年的检测冠军。选择 DPM 是由于人体姿态的多样性以及个体差异,传统固定的单一模型对复杂人体的匹配有较大的不足,而 DPM 在 HOG 行人检测框架<sup>[8]</sup>的基础上引入了可变形部件,这样就能够让模型去适应物体的变化。

特征描述器具体的训练过程简述如下:

(1)将图像分成小的连通区域,叫做细胞单元(cell);

(2)采集细胞单元中各像素点的梯度或边缘的方向直方图<sup>[9]</sup>;

(3)在多个细胞单元内进行对比度归一化,然后组合这些直方图就可以构成特征描述器。

DPM 模型包括一个覆盖整个目标的粗糙的根滤波器(见图 2(a))和若干个表示目标各个部件的高分辨率的部件滤波器(见图 2(b))。根过滤器能够在较粗糙的分辨率上近似覆盖整个目标物体,部件过滤器是在更高分辨率上覆盖物体中较小的部件。如建立人脸的模型,根滤波器捕捉的是整个人脸边界这些粗糙信息,部件滤波器可以捕捉眼睛、鼻子、嘴这些细节

信息。

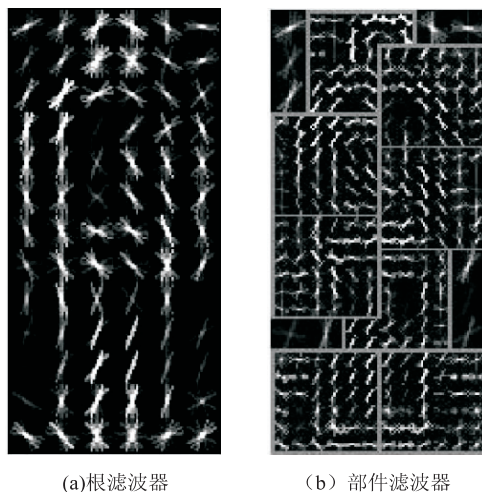


图 2 DPM 行人模型

DPM 模型使用了多尺度特征金字塔(multi-scale feature pyramid)在不同的伸缩尺度中提取特征,这样训练出的模版也是多尺度的,因此能够适应待检测图片的大小在一定尺度上的伸缩变化。

目标的检测主要根据待检测目标和人体模型的匹配得分来判断。目标假设的得分等于每个滤波器在各自位置的得分减去此位置相对于根位置的变形花费(deformation cost)再加上偏差值,其中位置的变形花费衡量了部件偏离其理想位置的程度:

$$\text{score}(p_0, p_1, \dots, p_n) = \sum_{i=0}^n F_i \cdot \varphi(H, p_i) - \sum_{i=1}^n d_i \cdot \Phi_d(dx_i, dy_i) + b \quad (1)$$

其中,  $F_i$  为第  $i$  个部件的滤波器;  $H$  为特征金字塔;  $p_i = (x_i, y_i, l_i)$  为第  $i$  个滤波器所在的层和位置坐标;  $d_i$  为部件的每个可能位置相对于锚点位置的变形花费;  $(dx_i, dy_i)$  给出了第  $i$  个部件相对于其锚点位置的位移:

$$(dx_i, dy_i) = (x_i, y_i) - (t \cdot (x_0, y_0) + v_i) \quad (2)$$

其中,  $(x_0, y_0)$  是根滤波器在其所在层的坐标;部件滤波器的分辨率是根滤波器的  $t$  倍,所以为了统一到部件滤波器所在层需乘以  $t$ ;  $v_i$  为部件  $i$  的锚点相对于根的坐标偏移。

使用非极大值抑制方法对每个检测框进行整合后,得到一个得分图,每个分数代表对象位于这个位置的可能性。

## 3 颜色空间

根据不同的应用可以把颜色空间分为以下三大类:TV 系统颜色空间、计算机图像颜色空间、国际照明委员会颜色空间(International Commission on Illumination, CIE)。TV 颜色包括 YUV、YIQ 和 YCbCr 等,计算

机图像颜色包括 HSV、HIS 和 RGB 等,Cffi 颜色包括 LUV、XYZ 和 Lab 等。Wang 等<sup>[10]</sup> 在比较各种颜色空间用于行人检测之后,得出结论:OE-Lab、CIE-LUV、HSV、HSI 对于行人检测有很好的效果。文中分别采用 LUV 和 RGB,根据文献[11]中提出的色彩空间的特征提取方法进行检测实验。

3.1 cell 直方图计算

首先将训练样本(或待检测窗口)分割成无重叠的单个 cell。分别计算每个 cell 的累计直方图,其中直方图由这个 cell 中所有颜色通道的连接而得。然后对 cell 中的直方图进行归一化,最终得到的一幅包含整个样本的 Cell-histogram 图。具体操作时,将图片分割成 12 \* 5 个 cell。

3.2 方向滤波

根据文献[11]提出的方法,定义了 4 种方向滤波,在每种滤波器下分别计算临近 cell 的直方图不相似度,具体计算如下:

$$d(i,j,k) = \text{dissim}(\text{cell}_{i,j}, \text{cell}_{i+j_k}) \tag{3}$$

其中,  $k(k \in \{1,2,3,4\})$  表示滤波器的种类,如图 3 所示;  $\text{cell}_{i,j}$  表示图像中位于  $i$  行  $j$  列的 cell 的颜色累计直方图;  $(i_k, j_k)$  表示针对不同滤波器相对于  $\text{cell}_{i,j}$  的偏移位置。

$\text{dismiss}(\cdot)$  选用的是直方图相交的方式进行匹配,具体计算如下:

$$D(\text{cell}_{i,j}, \text{cell}_{i+j_k}) = \frac{\sum_{n=1}^S \min(\text{cell}_{i,j}(n), \text{cell}_{i+j_k}(n))}{\sum_{n=1}^S \text{cell}_{i,j}(n)} \tag{4}$$

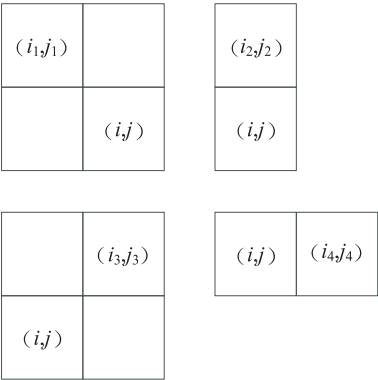


图 3 4 种方向滤波

3.3 池化

实验中使用最大池去筛选所有方向滤波器下的最大不相似度。这是因为神经元之间存在 CRF<sup>[12]</sup> 的抑制效果,所以通过最大池去增加特征的空间不变性。具体的池化操作如下:

$$H(i,j,k) = \max(d(m,n,k)) \tag{5}$$

其中,  $k \in \{1,2,3,4\}$ 。  $I, J$  分别为由  $i$  和  $j$  扩展所

得针对  $k$  的不同取值,对应的  $I, J$  有下列规定:

$$k = \begin{cases} 1, & I \in \{i, i+1\}, J \in \{j, j+1\} \\ 2, & I \in \{i, i+1\}, J \in \{j-1, j, j+1\} \\ 3, & I \in \{i, i+1\}, J \in \{j-1, j\} \\ 4, & I \in \{i-1, i, i+1\}, J \in \{j, j+1\} \end{cases} \tag{6}$$

$H$  作为最终的颜色特征用作训练分类器以及在待检测窗口进行检测。

4 定义可疑区间

当利用可变形部件模型检测到的分数  $S_0$  大于阈值时,判断  $S_0$  是否属于所定义的可疑区间,若不属于,结束判断,若属于,就重新对 DPM 判定为行人的区域进行 RGB 或者 LUV 颜色空间的检测并得到检测得分  $S_1$ 。若  $S_1$  大于 0,则判定检测窗口是行人;若  $S_1$  小于 0,则判断非行人。具体流程见图 1。实验选取了 4 个可疑区间(sp)进行对比。其中 sp 的取值范围如下:

$$\text{sp} \leq P(P = -0.4, 0, 0.4, 0.8) \tag{7}$$

5 实验结果与分析

文中训练数据采用的是 INRIA 数据集,选取正样本为 INRIA 训练集中所有 2 416 个行人,负样本为 1 218 个;测试集为 INRIA 测试集中所有图片,其中正样本 568,负样本 462 个。

5.1 判断标准

实验使用重叠率作为行人是否准确定位的标准。重叠率的定义为检测结果与真实值之间重叠区域的大小占它们之和的百分比:

$$r_{\text{overlap}} = \frac{\text{Area}(R_d \cap R_g)}{\text{Area}(R_d \cup R_g)} \tag{8}$$

其中,  $R_d$  表示对同一目标使用 NMS<sup>[13]</sup> 合并之后的有效框;  $R_g$  表示数据库标记的行人区域。

重叠率的阈值固定为 0.5,当检测到的  $r_{\text{overlap}}$  大于阈值时,则判定该有效框属于 True Positive,否则判定为 False Positive。

5.2 颜色空间的对比

在 RGB 颜色空间中,使用 LIBSVM<sup>[14]</sup> 对颜色空间进行训练,参数设置为 ‘-s 1 -t 0 -c 0.1 -g 0.8’。实验时将检测窗口固定为每 9 \* 9 为一个 cell,分别对不同窗口大小对准确度的影响进行实验。如表 1 所示,将窗口分割成 12 \* 5 个 cell 时准确率较高。

表 1 不同的窗口大小对准确度的影响

窗口大小	准确率/%
108 * 36	86.01
108 * 45	87.004
108 * 54	86.45

使用同样的训练参数以及  $12 * 5$  个 cell 作为窗口大小,实验中分别就颜色空间的 LUV 和 RGB 特征单独在 INRIA 数据库上进行检测,LUV 特征的准确率为 81.89%,而 RGB 的准确率达到 87.004%,结果表明 RGB 特征在 INRIA 数据库中的检测效果优于 LUV 特征。

5.3 整体的实验效果

在 DPM 检测部分使用作者训练好的 INRIA 模型,阈值  $T$  的取值范围  $T \in (-1, -0.8, \dots, 0.8)$ 。然后对不同的可疑区间进行对比。实验表明,在 RGB 空间中,发现在可疑区间  $p = 0$  时,效果最好,而在 LUV 空间中,  $p = -0.4$  效果最好。

实验分别对比了  $p = -0.4$  的 LUV 空间和  $p = 0$  的 RGB 空间在 INRIA 数据库下的检测效果。实验结果表明,在 LUV 空间中,DPM-LUV 的效果和 DPM 效果相差不是特别大,而在 RGB 空间中,DPM-RGB 的效果相对传统的 DPM 来说有了一定的进步,如图 4 和图 5 所示。

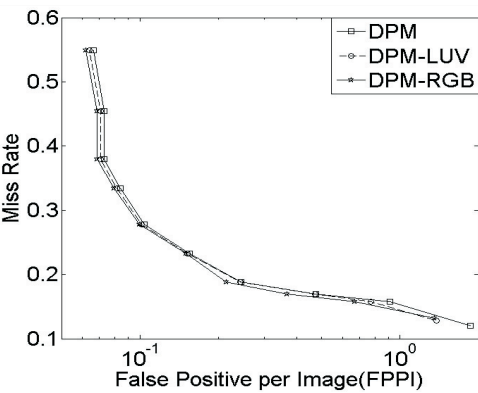


图 4 整体的实验效果对比

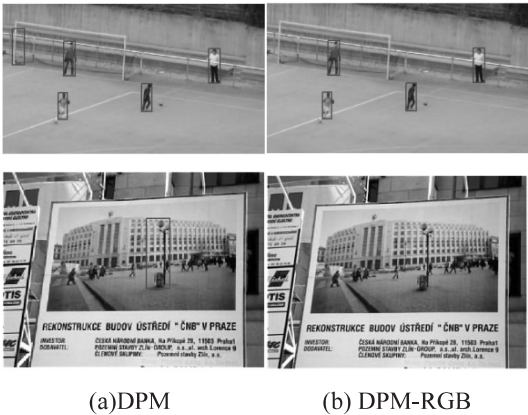


图 5 DPM 和 DPM-RGB 在 INRIA 测试集上的实验效果

6 结束语

文中分析了行人检测中由于行人非刚性所造成的检测难点,结合实验效果较好的 DPM 算法和颜色空间的算法进行行人检测,分别在 RGB 空间和 LUV 空间

设定不同的可疑区间进行对比。实验结果表明,在 INRIA 数据库中,DPM-RGB 的检测效果取得了一定的提升,有效减少了 False Positive,但是会增加少量的 Miss。下一步将尝试在 RGB 空间中使用不同的滤波进行检测以及对多决策条件进行优化。

参考文献:

[1] 贾慧星,章毓晋. 车辆辅助驾驶系统中基于计算机视觉的行人检测研究综述[J]. 自动化学报,2007,33(1):84-90.

[2] 万雪飞. 基于双人体可变形部件模型的深层人体检测[D]. 大连:大连海事大学,2016.

[3] Dalal N,Triggs B. Histograms of oriented gradients for human detection[C]//IEEE computer society conference on computer vision and pattern recognition. Washington DC: IEEE, 2005:886-893.

[4] 孙 昀,刘富强,李志鹏. 基于空间梯度直方图的行人检测算法[J]. 中国图象图形学报,2008,13(10):1825-1828.

[5] 黄 茜,顾杰峰,杨文亮. 基于梯度向量直方图的行人检测[J]. 科学技术与工程,2009,9(13):3646-3651.

[6] 李梦涵,熊淑华,熊 文,等. 多尺度级联行人检测算法的研究与实现[J]. 计算机技术与发展,2014,24(8):10-13.

[7] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part-based models[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2010,32(9):1627-1645.

[8] Jiménez P G, Bascón S M, Moreno H G, et al. Traffic sign shape classification and localization based on the normalized FFT of the signature of blobs and 2D homographies[J]. Signal Processing,2008,88(12):2943-2955.

[9] Dalal N. Finding people in images and videos[D]. France: Institut National Polytechnique de Grenoble-INPG,2006.

[10] Wang Q, Pang J, Liu G, et al. Color maximal-dissimilarity pattern for pedestrian detection[C]//Proceedings of the 21st international conference on pattern recognition. [ s. l. ]: [ s. n. ],2012:1952-1955.

[11] Wang Q, Pang J, Qin L, et al. Justifying the importance of color cues in object detection;a case study on pedestrian[M]. [ s. l. ]:[ s. n. ],2013:387-397.

[12] Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29 ( 3 ) : 411 - 426.

[13] Wojek C, Dollar P, Schiele B, et al. Pedestrian detection; an evaluation of the state of the art[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34 ( 4 ) : 743 - 761.

[14] Chang C C, Lin C J. LIBSVM:a library for support vector machines [ J ]. ACM Transactions on Intelligent Systems and Technology,2011,2(3):1-27.