# 基于似物性和空时协方差特征的行人检测算法

刘春阳 吴泽民 胡 磊 刘 熹
(解放军理工大学通信工程学院 南京 210007)

**摘 要** 针对行人检测算法中缺少空时信息融合、检测区域过大等问题,提出了一种联合似物性检测和基于通道协方差信息的改进算法。该算法首先对图像进行二进制梯度归一化的似物性检测,并形成行人检测候选区域,缩小检测区域;然后提取待测目标的空间和时间特征;最后基于协方差信息构造一种融合空时特征的检测器,以提高检测精度。在公开的数据集 INRIA 和 Caltech 上的实验结果表明:该算法的性能优于目前主流的行人检测算法。

**关键词** 计算机视觉,行人检测,似物性,协方差特征
**中图法分类号** TP391 **文献标识码** A

## Pedestrian Detection Based on Objectness and Sapce-Time Covariance Features

LIU Chun-yang WU Ze-min HU Lei LIU Xi
(College of Communications Engineering, PLAUST, Nanjing 210007, China)

**Abstract** In order to solve the fusion of space-time information and excessive detection area in pedestrian detection, a pedestrian detection method was proposed based on objectness and space-time covariance features. Firstly, binarized normed gradients algorithm is used for a test image to get objectness evaluations, and a pedestrian detection candidate area is formed. Secondly, the spatial and temporal features are extracted. Finally, a space-time detector based on covariance information was proposed to improve the accuracy. Experimental results on the INRIA and Caltech demonstrate that the proposed method outperforms the state-of-art pedestrian detectors in accuracy.

**Keywords** Computer vision, Pedestrian detection, Objectness, Covariance features

## 1 引言

计算机视觉是一门研究图像和视频的获取处理、分析和理解的科学,它的目标是使机器具有像人一样的视觉感知和理解能力[1]。行人检测的定义为判断输入图片(或视频帧)是否包含行人,如果包含则给出位置信息[2]。作为计算机视觉领域中多项研究的前提,行人检测与行人跟踪、人体再识别、姿态估计和行为分析等问题紧密相连,因此具有很高的研究价值[3],是目前研究的热点和难点。

目前行人检测方法主要有两大类:第一类是基于卷积神经网络的深度学习模型,该模型利用卷积神经网络得到具有较强分辨力的特征,然后用这个特征进行分类学习,其中最具代表性的模型有 R-CNN[4],SPP-net[5],Fast-R-CNN[6],Deep-R-CNN[7],YOLO[8],DeepCascade[9] 等。该方法减少了研究者手工设计行人特征和标注行人信息的时间,特征提取更加符合人类真实学习的过程,能获得描述能力较好的高维特征。但是该方法存在训练数据多、训练耗时长、硬件要求高等缺点。第二类是传统的图像处理方法,通过人工设计较好的样本特征,包括边缘、形状和变换特征等来描述行人。其中最具代表性的特征包括方向梯度直方图特征、Haar 小波特征、颜色的自相关特征和形状轮廓模板特征等[10]。该方法的检测性能依赖于提取的人工特征是否能够充分表征行人信息。其中最具代表性的模型有 ACF[11],InformedHaar[12],LDCF[13],SpatialPooling[14],其核心为通道特征,利用多种异质特征融合的方法从不同方面来描述图像,其参数少且对具体参数值不敏感,可使检测器空间定位更加准确[15]。

本文在通道特征的基础上,利用协方差矩阵构造了图像的空时协方差特征,联合似物性检测缩小了检测区域的面积,在提高检测精度的同时减少了检测时间。实验的仿真结果表明,相比传统的行人检测算法,本文算法取得了最好的检测效果。

## 2 基于似物性的检测区域快速提取

滑动窗口法在目标检测计算中被广泛采用,其定义为目标检测器滑动地评估检测区域的每个窗口图像。检测区域通常为整幅图像,并未区分前景目标和背景区域。但对行人检测而言,背景区域较多,使用滑动窗口法遍历整幅图像耗时较多且容易受到背景区域的干扰。因此采用基于似物性的检测方法来快速提取前景目标,减少行人检测器的搜索面积。

### 2.1 二进制梯度归一化的似物性检测

似物性检测通过通用目标检测器来产生一组候选对象窗口,对象状态表示一个图像窗口包含任意类型对象的概率值。不同于目标检测需要精确定位某类目标所在的位置,似物性

本文受国家自然科学基金(61501509)资助。

刘春阳(1993-),男,硕士,主要研究方向为计算机视觉、机器学习,E-mail:plaust_liu@163.com;吴泽民(1973-),男,教授,硕士生导师,主要研究方向为数据融合、视觉信息处理;胡 磊(1985-),男,讲师,主要研究方向为压缩感知、视频信息处理;刘 熹(1972-),男,副教授,主要研究方向为数据链信息处理。

检测仅需指出该区域包含前景目标的概率，与目标的具体类型无关[16]，能够有效减少检测的区域，因此似物性检测非常适合进行行人检测的预处理。

通过分析可以发现，前景目标通常是独立的且具有良好的封闭轮廓，同时将图像归一化到一个相同的尺度上，前景目标的封闭轮廓和梯度范数之间具有很强的相关性。因此，Cheng 等提出了一种基于二进制梯度归一化（Binarized Normed Gradients，BING）的似物性估计方法[17]。算法流程如图 1 所示。

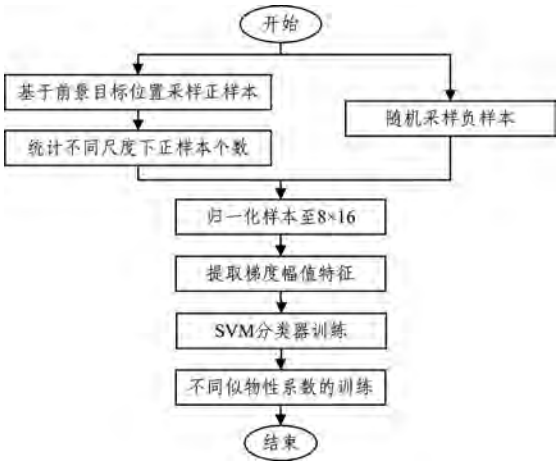
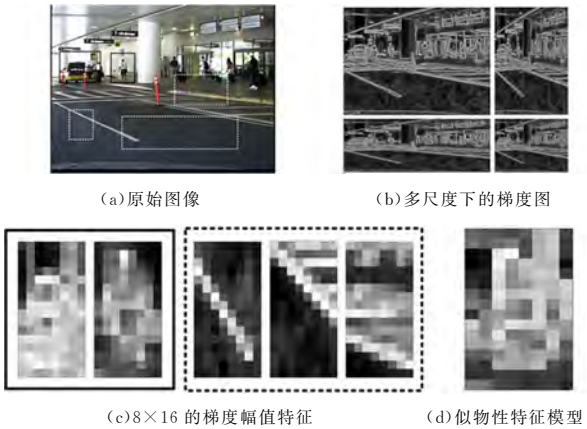



图 1　似物性检测算法流程

通过分析 BING 在 PASCAL VOC 数据集上的检测结果可以发现，该方法对待测图像仅生成 1000～2000 个前景预测位置，计算速度达到了每秒 300 幅图像且在仅使用 1000 个窗口（约占整个滑动窗口的 0.2%）的情况下检测精度为 96.2%[17]。其极快的检测速度为似物性作为行人检测的预处理过程提供了可行性，并且其极高的前景目标检测精度为准确性提供了保障。

### 2.2　梯度幅值模型

梯度幅值是一个密集且紧凑的似物性特征，具有以下优点：1）采用了固定尺寸的归一化特征信息，梯度幅值特征对于位置、尺度、纵横比不敏感，非常适用于通用目标的检测；2）梯度幅值特征的结构简单且紧凑，计算效率很高，耗时较少，适合作为行人检测的预处理过程。前景目标一般具有良好的封闭轮廓和中心位置。在重置目标窗口时，相当于将实际对象缩小到一个固定大小的区域，因为在封闭的轮廓中，图像梯度变化很小，所以梯度幅值是一个很好的可区分特征。



（a）原始图像　（b）多尺度下的梯度图　（c）8×16 的梯度幅值特征　（d）似物性特征模型

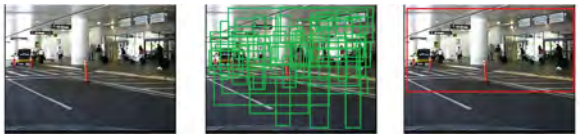
图 2　梯度幅值特征模型

如图 2 所示，黑色实线框是前景目标，白色虚线框是背景区域。图 2（a）中对于行人和车，不同类型的前景目标在颜色、形状、纹理、光照等方面都有很大的不同。将输入图像重置为不同尺度的空间，并如图 2（b）计算不同尺度下的梯度，结果如图 2（c）所示，由结果可以发现其前景目标的梯度幅值特征具有某些共性，而背景区域特征则比较杂乱。基于梯度幅值特征得到了一个行人检测模型用来删选前景目标，如图 2（d）所示。

为了更符合行人的目标特征，采用 8×16 大小的梯度幅值模型替代 BING 原始算法中 8×8 大小的模型，并采用行人检测的 Caltech 数据库替代 BING 原始算法中的 PASCAL VOC 数据库进行训练（见图 2（d）），生成一个 128 维的符合行人检测所需的梯度幅值特征模型。

### 2.3　行人检测区域的提取

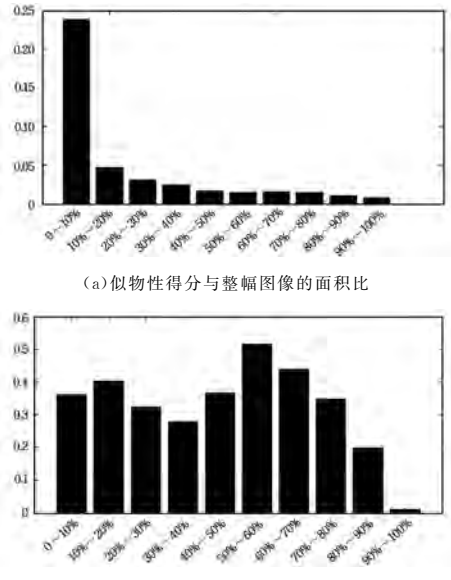
在 Caltech 数据库中，检测图像大多如图 3（a）所示，具有很多背景冗余信息。滑动窗口法需要遍历所有检测区域，因此快速地提取行人检测区域能够有效地提高检测速度和检测精度。如图 3（b）所示，本文提出对待测图像采用 BING 方法进行似物性检测，然后对候选区域进行筛选与合并，得到如图 3（c）所示的行人检测区域[18]。



（a）样本图像　（b）似物性检测框　（c）行人检测区域

图 3　基于似物性的快速搜索

如图 4 所示，对似物性的得分进行排序，可以发现两个现象：1）随着得分的降低，其似物性区域相对于整幅图像所占的比例相应减小；2）似物性区域与检测对象的重合面积比例不规则地变化，在 10%～20%和 50%～60%之间出现了两个较为显著的峰值。



（a）似物性得分与整幅图像的面积比

（b）似物性得分与目标重叠的比例

注：横坐标表示似物性得分，纵坐标表示面积比

图 4　似物性统计分析

我们对 Caltech 图像库中的图像进行具体分析。得分在前 10%的似物性窗口虽然可以较好地覆盖检测对象，但是其

尺寸较大，包含了较多的背景信息，不利于有效地提取行人检测区域。得分在 10%～20%和 50%～60%的似物性窗口尺寸较小，能够有效覆盖行人检测区域并且包含较少的背景冗余。为了准确提取行人检测的区域并减少计算的复杂度，本文提出了如图 5 所示的行人检测区域提取的流程，首先选择得分在 10%～20%和 50%～60%的似物性窗口，然后通过对这两类得分检测窗口的归并得到最终的行人检测区域。

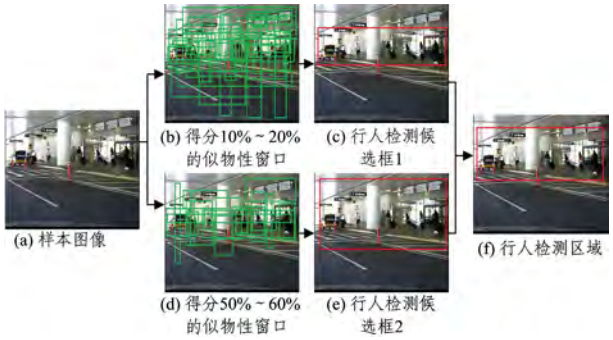



图 5 行人检测区域提取的流程

根据在 Caltech 行人检测数据库训练所得的 8×16 的梯度幅值模型对待测图像进行基于 BING 算法的似物性检测，得到 $p$ 个似物位置，并根据其得分从大到小进行排序。这里以提取候选框 1 为例，即选取得分在 10%～20%之间的似物性窗口形成高精度的行人检测候选区域，其具体流程如下。

设待测图像的分辨率大小为 $m\times n$，$R_k$ 为第 $k$ 个似物位置，$F_k(i,j)$为像素的似物性函数。

$$F_k(i,j)=\begin{cases}1, & \text{if } I(i,j)\in R_k\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

每个像素 $I(i,j)$在候选区域中出现的频数为 $F(i,j)$：

$$F(i,j)=\sum_{k=p_1}^{p_2}F_k(i,j) \tag{2}$$

其中，$p_1=\lfloor 0.1\times p\rfloor$，$p_2=\lceil 0.2\times p\rceil$。

然后对 $F(i,j)$设置阈值 $s$ 进行二进制归一化：

$$A(i,j)=Binary[F(i,j)]=\begin{cases}1, & \text{if } F(i,j)\geqslant s\\ 0, & \text{otherwise}\end{cases} \tag{3}$$

对 $A(i,j)$进行逐行累加，形成 $m$ 维列向量 $V$：

$$V_i=\sum_{j=1}^{n}A(i,j) \tag{4}$$

将向量 $V$ 的 $m$ 个元素按从小到大的顺序进行排序，设第 $\delta m$ 个元素的值为 $t$。对行累加向量 $V$ 设置阈值 $t$ 进行二进制归一化：

$$V=Bin(V_i)=\begin{cases}1, & \text{if } v_i\geqslant t\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

同理，对 $A(i,j)$的每列求和并二值化，得到 $n$ 维行向量 $W$。通过式(6)可以得到 0-1 矩阵 $M_1$：

$$M_1=V\times W \tag{6}$$

则 $M_1$ 中包含所有元素 1 的最小矩形区域为 $N_1$，即行人检测区域候选框 1，如图 5(c)所示。同理，使用得分在 50%～60%的似物性窗口形成行人检测区域候选框 2，即矩形区域 $N_2$，如图 5(e)所示。

候选框 1 为高精度的似物性窗口，候选框 2 为低精度的似物性窗口，为了进一步去除背景区域，减少搜索面积，本文根据式(7)将 $N_1$ 和 $N_2$ 相与，从而得到最终的行人检测区域。

$$N=N_1\bigcap N_2 \tag{7}$$

## 3 基于空时协方差通道的模型

在近年来的研究中，通道特征在图像和视频的行人检测中被广泛应用，其主要通过将手工提取的不同图像特征作为独立通道，对不同的特征通道进行融合，综合地表征行人特性来提高行人检测的精度。人工选取的特征通常包括：梯度特征、颜色特征和纹理特征等。但对视频来说，运动特征同样非常重要，针对运动场景仅依靠空间特征的检测，其性能还有很大的提升空间。因此，考虑融合空间和时间特征来构造行人空时特征的描述子。而通道协方差通过计算待测图像的不同通道的协方差矩阵来融合多种特征，同时结合光流法的时间域特征构建空时融合模型。

### 3.1 空时特征向量

对于一个连续视频序列，首先提取单帧图像的空间特征通道。对于行人检测，Dollar 教授等于 2014 年提出了聚合通道特征[11]，该通道特征联合了梯度直方图、梯度强度和 LUV 颜色空间，是目前最常用的行人检测通道。在空间域的特征通道的基础上加入时间域通道，采用 Brox 光流算法[19]提取连续帧的光流场。定义第 $t$ 帧的每个像素点的 12 维空时特征通道为：

$$z(x,y,t)=[L(x,y,t),U(x,y,t),V(x,y,t),G(x,y,t),\alpha_1,\alpha_2,\alpha_3,\alpha_4,\alpha_5,\alpha_6,V_x(x,y,t),V_y(x,y,t)]^{\mathrm{T}} \tag{8}$$

其中，$x,y$ 为空间的像素坐标；$L,U,V$ 为像素点在 LUV 颜色空间的颜色值；$G$ 为视频帧的梯度值；$\alpha_i$ 为图像的 6 个梯度方向量化值，由每个池化块内的梯度方向离散化到 6 个直方图区间所得；$V_x$ 为光流场的水平分量，$V_y$ 为光流场的垂直分量。同时，考虑多分辨率的池化域能够更全面地表征行人特征，设计了 8×8，16×16 和 32×32 像素 3 种不同大小的池化域来聚合人体信息。

### 3.2 空时特征协方差通道

协方差矩阵通常为正半定矩阵，体现了多组变量之间的关系。协方差矩阵的对角线上的值表示每个通道的方差，非对角线上的值表示不同通道之间的相关系数。方差体现了不同通道的特征值与平均值的偏差，并提供了特征分布的相关信息。而相关系数提供了池化域内不同通道特征之间的关系。

对于某个图像池化域 $R$，其协方差矩阵 $C_R$ 通过式(9)计算：

$$C_R=\frac{1}{n-1}\sum_{i=0}^{n}(z_i-\mu)(z_i-\mu)^{\mathrm{T}} \tag{9}$$

其中，$n$ 为池化域内的像素数量，$z_{i=1,\cdots,n}$为池化域内 $i$ 像素所提取的 $d$ 个特征通道，$\mu=\sum_{i=1}^{n}z_i/n$ 为池化域内通道特征的均值。如式(9)所示，通道数量为 12，一个空间特征通道的协方差可以表述为一个 12×12 的矩阵，由于矩阵的对称性，只存储上三角形部分，共计 78 个不同的值。因为 LUV 颜色空间通道相对独立，所以只选择图像的 7 个梯度直方图特征通道(即 1 个梯度强度和 6 个梯度方向特征通道)和 2 个光流场通道的方差和相关系数。这样在不同分辨率的池化域，都可以得到 9 个方差通道和 36 个相关系数通道，在 3 种不同分辨率下，一共提取了 165 个基于协方差的特征通道。通道共有 3(LUV 颜色空间通道)+3×9(梯度直方图特征通道和光流场通道)+3×9(方差通道)+3×36(相关系数通道)=165 个。

### 3.3　行人检测器的训练

行人检测器的流程包括读取数据集、提取空时协方差通道和训练检测器。针对空时协方差通道，采用 AdaBoost 分类器实现，共进行 3 轮训练。以 Caltech 图库为例，在第一轮训练过程中，首先将包含行人的正样本图片和随机截取的负样本图片投入分类器进行训练，得到包含空时特征的 165 个通道的初始检测器；然后利用初始检测器对负样本进行检测，生成的矩形框是分类错误的负样本即为“难例”，将“难例”作为负样本重新投入分类器进行训练；重复上述操作，进行新一轮的“难例”删选，得到包含 3 个不同尺度的行人检测器。

算法 1 给出了空时行人检测器的具体流程。

**算法 1**　空时行人检测器

输入：视频序列 $I=\{I_t\}$，帧数 N

输出：行人检测器{S}

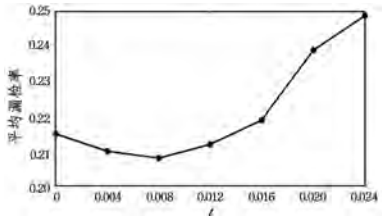
1. 初始化：行人检测器池化域大小为 $S^{8\times8}=8\times8$，$S^{16\times16}=16\times16$，$S^{32\times32}=32\times32$；
2. 按式(8)－式(9)计算第一帧 $I_1$ 的空时特征协方差通道 $S_1^{8\times8}$，$S_1^{16\times16}$，$S_1^{32\times32}$；
3. for t=1:N do
4. 按式(8)－式(9)计算第 i 帧 $I_i$ 的空时特征协方差通道 $S_i^{8\times8}$，$S_i^{16\times16}$，$S_i^{32\times32}$；
5. end
6. for $C=S_i^{8\times8}$，$S_i^{16\times16}$，$S_i^{32\times32}$；
7. 将 $S_i^{8\times8}$ 的正样本图像和负样本图像投入 AdaBoost 分类器，得到初始检测器；
8. for i=1:2 do
9. 对负样本进行检测，得到难例，同时将难例投入负样本中，重新进行 AdaBoost 分类，得到行人检测器；
10. end
11. end

## 4　实验结果及分析

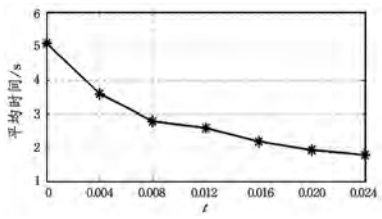
在 INRIA 和 Caltech 两个行人检测数据集上测试了不同的算法。INRIA 数据集分为训练集和测试集，训练集包含 614 张正样本图片和 1218 张负样本图片，测试集包含 288 张。检测图片数据集中的图像来自 GRAZ 01 数据集和网络图片，相互独立且不是连续视频帧。Caltech 数据集分为测试集和训练集，包含像素为 640×480、帧率为 30 Hz 的 10 段连续视频。视频来自于一辆在城市正常行驶的汽车的车载摄像头，标注了约 250000 帧(约 137 min)，350000 个矩形框，2300 个行人。两个数据集都已经做好了真值标注，这些图像和视频包含尺寸不同、目标遮挡和背景复杂等各种情况，具有非常大的挑战性。这里主要比较 2014 年以来比较典型的传统算法：ACF，InformedHaar，LDCF 和 SpaticalPooling。本文采用漏检率-平均误检率(miss rate-false positive per image)曲线作为检测精度的评价[20]，采用检测一幅图像的平均时间进行实时性的评价。实验的硬件环境为 2.40 GHz 双核的 Inter Core i3 处理器的电脑，仿真平台为 Matlab 2012b。

### 4.1　基于似物性的快速搜索中参数的讨论

在 2.3 节中，阈值 $t$ 被用来平衡提取的行人检测区域的大小，而行人检测区域的大小对算法的检测精度和检测速度有着重要的影响。本节主要讨论阈值的取值对行人检测性能的影响。实验中，$t$ 值从 0 变化至 0.024，每次调整的步进为 0.004。在 Caltech 数据集上进行检测，结果如图 6 所示。从图 6(a)可以看出，随着阈值 $t$ 的增加，平均误检率有一个先下降再上升的过程，当 $t$ 取 0.008 时，平均误检率最低。从图 6(b)可以看出，随着阈值 $t$ 的增加，平均耗时不断下降。如表 1 所列，随着阈值不断增大，检测区域的面积逐渐减小，大量的相关信息被去除掉，影响检测的精度。因此综合考虑不同阈值下的检测精度和检测耗时，将阈值 $t$ 设置为0.008。




(a)平均漏检率




(b)平均耗时

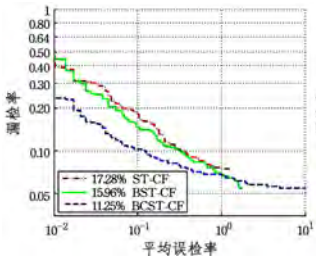
图 6　基于似物性的快速搜索的 $t$ 值的性能分析

表 1　基于似物性的快速检测下的阈值情况

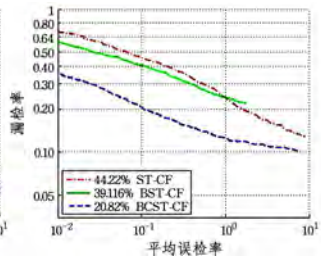
| 阈值 $t$ | 检测区域与原始图像的面积比/% | 平均误检率/% | 平均耗时/s |
|---|---|---|---|
| 0 | 87 | 21.5 | 5.1 |
| 0.004 | 64 | 21.0 | 3.6 |
| 0.008 | 50 | 20.8 | 2.8 |
| 0.012 | 41 | 21.2 | 2.6 |
| 0.016 | 35 | 21.9 | 2.2 |
| 0.020 | 26 | 23.9 | 1.9 |
| 0.024 | 20 | 24.9 | 1.8 |

### 4.2　所提算法的性能分析

本文算法主要分为两步：基于似物性的快速搜索(记为 BST-CF)的预处理和联合空时协方差通道特征(记为 BCST-CF)的行人检测。另外，为了验证算法的有效性，先不考虑似物性的预处理和构造空时协方差通道模型，仅使用原始的空间特征通道进行检测(ST-CF)。为了更加全面地分析算法的性能，在两个数据集上进行了测试。因为 INRIA 图库并不是连续视频帧，不能加入包含时间信息的光流场通道，所以对于 INRIA 图库，空时协方差通道模型仅包含空间特征，而 Caltech 包含完整的空时特征。如图 7 所示，比较了两个处理阶段的漏检率-误检率曲线的性能。从图中可以看出，经过基于似物性的预处理后，行人检测精度有明显提升，当进一步联合基于空时协方差的通道模型后，行人检测的精度获得了大幅提高。




(a)INRIA 数据集

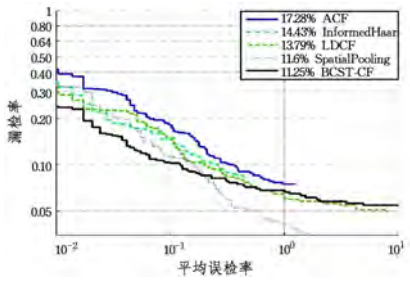



(b)Caltech 数据集

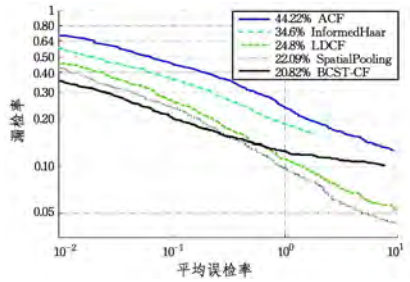
图 7　基于似物性的快速搜索和空时协方差通道模型优化的性能分析

### 4.3 与其他算法的比较

分别使用 ACF 算法、InformedHaar 算法、LDCF 算法、SpaticalPooling 算法及基于似物性和空时协方差通道特征的(BSCT-CF)算法在 INRIA 数据集和 Caltech 数据集上进行测试。图 8 给出了本文算法与其他算法的漏检率-误检率曲线的比较。从图中可以看出,算法在不同图库中的检测精度相比其他算法均有明显的提高。




(a)INRIA 数据集




(b)Caltech 数据集

图 8 两个数据集上不同算法的性能比较

在较大的平均误检率范围$[10^{-2},10^{0}]$内,本文方法的平均漏检率均低于目前的主流算法,反映出本算法在低误检率的情况下具有良好的性能,其精度很高。从图中可以得到,相对于其他 4 种行人检测算法,本文算法在较低误检率的情况下性能优异,但在较高误检率的区域其性能曲线逐渐变得平滑,在 INRIA 图库中 FPPI 为 0.5 左右,其漏检率逐渐比 SpaticalPooling 算法高;在 Caltech 图库中 FPPI 为 100 左右,漏检率逐渐比 LDCF 和 SpaticalPooling 高。经过分析,本文算法采用了基于似物性的快速搜索的预处理,在去除背景区域的同时将少部分行人也排除在行人检测区域之外,其行人的正确检测数将略少于未去除背景的其他算法,因此当平均误检率变大时,其漏检率曲线将逐渐平滑。考虑到去除背景区域在较低平均误检率时取得的巨大收益,其总体检测精度依然提升明显。该算法无论是在复杂的背景,还是目标运动以及目标出现遮挡和形变等情况下都取得了较好的检测效果,具有较强的鲁棒性。

### 4.4 运行时间分析

表 2 列出了本文算法和其他算法在每帧图像上的平均运算时间的比较。所有代码都在 Matlab 2012b 平台上运行,测试数据集为 Caltech。SpaticalPooling 算法和本文算法都需要光流场的计算,光流场的计算时间复杂度大(约为 12.5 秒/帧),因此其比无光流场的其他算法慢。该算法在去除光流场的计算后处理一帧的时间约为 3.5 s,与其他算法相比耗时相差不多。另外,LDCF 算法在检测之前需要对整个视频进行联合优化,得到相应的去相关通道模型,其预处理过程与帧数有很大的关系,当帧数过大时,计算耗时长,并且容易出现内存溢出的问题。

表 2 不同视频行人检测算法平均运算时间的比较

(单位:s)

| 算法 | ACF | Informed-Haar | LDCF | Spatical-Pooling | BCST-CF |
|---|---|---|---|---|---|
| 时间 | 0.1 | 2.1 | 1.4 | 18.2(5.8) | 15.4(3.5) |

注:括号内为去除光流后的运行时间

**结束语** 本文基于半正定的协方差矩阵,构造了空时协方差通道模型,将图像的行人检测扩展到视频的行人检测中,利于协方差矩阵的相关性融合了空时特征,大幅度改进了行人检测的精度。同时,提出一种基于似物性的快速搜索的预处理操作对前景目标进行优化,有效缩小了前景目标的检测范围,在不影响检测速度的前提下提高了检测精度。在两个数据集上的测试结果表明,本文算法在精度上的优势明显。同时通道协方差模型可以针对不同的视频进行扩展,例如对于运动剧烈的场景,可以加大光流场通道的权重;对于背景复杂的场景,可以加大方向梯度直方图的权重。另一方面,本文的运动矢量信息通过光流场来提取,时间复杂度较高,对算法的实时性提出了新的挑战。下一步将考虑利用视频的编码特性来更加快速地提取运动矢量信息。

## 参 考 文 献

[1] IKEUCHI K. Computer Vision:A Reference Guide[M]. Springer Publishing Company,Incorporated,2014.

[2] 苏松志,李绍滋,陈淑媛,等. 行人检测技术综述[J]. 电子学报,2012,40(4):814-820.

[3] CAO J,PANG Y,LI X. Pedestrian Detection Inspired by Appearance Constancy and Shape Symmetry[J]. IEEE Transactions on Image Processing,2016,25(12):5538-5551.

[4] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]// Computer Vision and Pattern Recognition. IEEE,2014:580-587.

[5] HE K,ZHANG X,REN S,et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,37(9):1904-1916.

[6] GIRSHICK R. Fast R-CNN[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago,Chile:IEEE,2015:1440-1448.

[7] HE K,ZHANG X,REN S,et al. Deep Residual Learning for Image Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2016:770-778.

[8] REDMON J,DIVVALA S,GIRSHICK R,et al. You Only Look Once:Unified,Real-Time Object Detection[C]// Computer Vision and Pattern Recognition. IEEE,2016:779-788.

[9] ANGELOVA A,KRIZHEVSKY A,VANHOUCKE V,et al. Real-Time Pedestrian Detection with Deep Network Cascades[C]// British Machine Vision Conference. 2015:1-32.

[10] ZHANG S,BAUCKHAGE C,CREMERS A B. Informed Haar-Like Features Improve Pedestrian Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. 2014:947-954.

(下转第 246 页)

绕 $x$,$y$ 及 $z$ 轴的顺序进行)分解出平移向量及 3 个欧拉角,作为对应仿真图像的位姿真值记录。这里在仿真图像的生成过程中只在以飞机模型为中心的上半球面上进行视点的采样,图 3 给出了球面上 5 个采样点下对应的仿真图像。在实际的应用中,可根据飞机的运动轨迹及与相机的相对位置关系等约束进一步缩小采样的范围,如图 3 中深色框选择区域,这样可以进一步减少样本数量,或者采用更加密集的视点采样,以提高后续参数估计的精度。

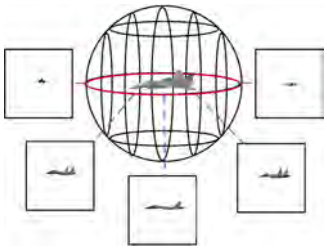


图 3 对三维模型进行二维投影的示意图

在实际仿真过程中,通过控制俯仰角 $\psi$、偏航角 $\theta$ 及半径 $r$,即可实现视点的设置,其中 $\psi\in[0,\pi/2]$,$\theta\in[0,2\pi)$,如图 4 所示。

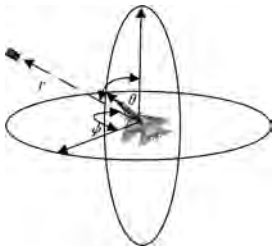


图 4 仿真图像生成中视点设置示意图

采样密度的增加有利于提高参数估计的精度,但样本数量的增加会增加计算负担,因此应综合考虑精度及处理时间的要求,设置合理的采样密度,完成样本数据库的制备工作。本着算法研究的目的,本文将 $\psi$ 及 $\theta$ 的采样步长均设为 1°,共生成 32760 张仿真图,部分仿真图如图 5 所示。

图 5 部分仿真图

**结束语** 本文以基于 CAD 模型的飞机目标位姿估计问题中的难点为研究目标,通过 OpenGL,保持相机不动,旋转飞机目标来离线建立具有精确位姿的 32760 张飞机图像库,后续工作将把本文算法应用到飞机位姿精确估计中,以解决三维位姿估计问题。

## 参 考 文 献


[1] 丁险峰,吴洪,张宏江,等. 形状匹配综述[J]. 自动化学报,2001,27(5):678-694.

[2] 高腾飞,曾朝阳,郝刚涛,等. 基于 OpenGL 的空间目标图像序列仿真[J]. 计算机工程与设计,2014,35(11):3946-3950.

[3] HEIKKILA J,SILVEN O. A four-step camera calibration procedure with implicit image correction[C]// CVPR. 1997.

[4] MONTI M M. Statistical Analysis of fMRI Time-Series:A Critical Review of the GLM Approach[J]. Frontiers in Human Neuroscience,2011,5(5):28.

[5] ROST R J,KESSENICH J M,LICHTENBELT B. OpenGL shading language[J]. WWW,2006,26(4):853-854.


(上接第 214 页)


[11] DOLLAR P,APPEL R,BELONGIE S,et al. Fast Feature Pyramids for Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,36(8):1532-1545.

[12] ZHANG S,BAUCKHAGE C,CREMERS A B. Informed Haar-Like Features Improve Pedestrian Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society,2014:947-954.

[13] NAM W,DOLLAR P,HAN J H. Local Decorrelation For Improved Pedestrian Detection[J]. Advances in Neural Information Processing Systems,2014,1:424-432.

[14] PAISITKRIANGKRAI S,SHEN C,HENGEL A V D. Pedestrian Detection with Spatially Pooled Features and Structured Ensemble Learning[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(6):1243.

[15] ZHANG S,BENENSON R,OMRAN M,et al. How Far are We from Solving Pedestrian Detection? [C]// IEEE Conference on Computer Vision & Pattern Recognition. 2016:1259-1267.

[16] ZHANG H,XU M,ZHUO L,et al. A novel optimization framework for salient object detection[J]. The Visual Computer,2016,32(1):31-41.

[17] CHENG M M,ZHANG Z,LIN W Y,et al. BING:Binarized Normed Gradients for Objectness Estimation at 300fps[C]// Computer Vision and Pattern Recognition. IEEE,2014:3286-3293.

[18] 刘涛,吴泽民,姜青竹,等. 基于似物性的快速视觉目标识别算法[J]. 计算机科学,2016,43(7):73-76.

[19] BROX T,BREGLER C,MALIK J. Large displacement optical flow[C]// IEEE Conference on Computer Vision and Pattern Recognition,2009. (CVPR 2009). IEEE,2009:41-48.

[20] DOLLAR P,WOJEK C,SCHIELE B,et al. Pedestrian detection:A benchmark[C]// IEEE Conference on Computer Vision and Pattern Recognition,2009(CVPR 2009). IEEE,2009:304-311.