

# 基于少量样本学习的多目标检测跟踪方法

罗大鹏<sup>1</sup>, 杜国庆<sup>1</sup>, 曾志鹏<sup>1</sup>, 魏龙生<sup>2</sup>, 高常鑫<sup>3</sup>, 陈应<sup>4</sup>, 肖菲<sup>4</sup>, 罗琛<sup>5</sup>

(1. 中国地质大学(武汉)机械与电子信息学院, 湖北武汉 430074; 2. 中国地质大学(武汉)自动化学院, 湖北武汉 430074;  
3. 华中科技大学自动化学院, 湖北武汉 430074; 4. 中建三局智能技术有限公司, 湖北武汉 430070;  
5. 北京师范大学惠州附属学校, 广东惠州 516002)

**摘要:** 视频目标检测跟踪算法一直是计算机视觉领域的研究热点, 目前大部分方法均需人工采集样本训练检测模型, 搭建目标检测跟踪系统. 当目标成像条件发生变化时, 需重新采集样本, 训练模型, 调试整个检测跟踪系统, 耗费大量人力、物力. 本文提出一种基于少量样本学习的多目标检测跟踪算法, 只需在监控视频第一帧指定待检测目标, 即可自主生成混合分类模型, 进行目标检测. 采用在线渐进学习算法学习目标姿态变化, 更新该模型. 结合基于颜色的目标跟踪算法, 自动构建高精度目标检测跟踪系统. 整个过程无需手工采集、标注训练样本. 因此, 易于扩展到其它监控场景, 通过自主学习形成该场景专用的检测跟踪系统, 实现不同监控环境下, 不同成像条件下都有较好的检测跟踪效果. 实验表明, 本方法能自主学习多种监控场景下的目标姿态, 无需手工标注训练样本, 在基于在线学习的算法中有最佳的检测精度, 此外也取得了和离线目标检测跟踪系统相似的性能.

**关键词:** 少量样本学习; 多目标检测; 多目标跟踪; 在线学习

**中图分类号:** TP391      **文献标识码:** A      **文章编号:** 0372-2112 (2021)01-0183-09  
**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.12263/DZXB.20180045

## Multi-Object Detection and Tracking Based on Few-Shot Learning

LUO Da-peng<sup>1</sup>, DU Guo-qing<sup>1</sup>, ZENG Zhi-peng<sup>1</sup>, WEI Long-sheng<sup>2</sup>, GAO Chang-xin<sup>3</sup>,  
CHENG Ying<sup>4</sup>, XIAO Fei<sup>4</sup>, LUO Chen<sup>5</sup>

(1. School of Electronic Information and Mechanics, China University of Geosciences, Wuhan, Hubei 430074, China;  
2. School of Automation, China University of Geosciences, Wuhan, Hubei 430074, China;  
3. School of Automation, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China;  
4. Intelligent Technology Co., Ltd. of Chinese Construction Third Engineering Bureau, Wuhan, Hubei 430070, China;  
5. Huizhou School Affiliated to Beijing Normal University, Huizhou, Guangdong 516002, China)

**Abstract:** Video object detection and tracking algorithms have become the research focus in the field of computer vision. Traditional methods need to manually collect samples to train detection models, and build object detection and tracking systems. When the imaging conditions change, it is necessary to re-collect samples to train the detection model and re-adjust the entire detection and tracking system, which requires tedious human efforts. In this paper, a multi-object detection and tracking algorithm is proposed based on few-shot learning. With this approach, a hybrid classifier that models one object class can be generated by simply marking several bounding boxes around the object in the first video frame. An online gradual learning algorithm is proposed to learn the object pose changes and update the model. Combined with the color-based object tracking algorithm, our method automatically builds high-precision object detection and tracking systems without manual collection and labeling training samples. This approach can be conveniently replicated many times in different surveillance scenes and produce scene-specific detectors under various camera viewpoints. Experimental results on several video datasets show our approach achieves comparable performance to robust supervised methods, and outperforms the state-of-the-art online learning methods in varying imaging conditions.

**Key words:** few-shot learning; multi-object detection; multi-object tracking; online learning

## 1 引言

视频目标检测跟踪是计算机视觉等领域的研究热点,在视频监控、虚拟现实、人机交互、自主导航等应用中具有重要的理论研究意义和实际价值.目前,大部分视频目标检测跟踪系统包括了目标定位、目标数据关联跟踪及轨迹生成三个模块.其中数据关联跟踪算法已随着基于检测的跟踪方法的深入研究取得了巨大的突破<sup>[1]</sup>.但是,目标定位模块仍大量采用离线学习算法训练分类器实现<sup>[2,3]</sup>.当面对不同的监控场景时,由于成像条件、光照、目标形变均不同,需重新采集样本训练不同场景的目标检测器,耗费大量人力、物力.

近年来,在线学习算法已被广泛用于提高检测模型的自适应能力.但是,基于以下两个问题,这种在线学习算法很难用于多场景,多成像条件下的视频目标检测跟踪系统.

### (1) 需人工采集、标注部分训练样本

在线学习目标检测系统的初始目标模型需要通过人工采集、标注的少量样本训练,再通过在线学习逐渐提高该模型的检测精度.相对于离线训练模型,降低了人工标注样本个数.但是,当监控场景众多时,人工标注每个场景的初始训练样本仍需耗费大量人力物力.

近年来,半监督学习<sup>[4]</sup>,迁移学习<sup>[5]</sup>以及弱监督学习<sup>[6-8]</sup>已被用于降低人工标注工作量.但如何实现无需人工标注的自主学习模型,仍是当前的研究热点.

### (2) 在线学习样本的自主获取、标注方法

通常要求可在线学习的目标检测系统自主获取在线学习的新样本,并对其类别进行自动标注,否则无法实现检测系统的智能化.但是如何对样本进行自动的、正确的标注仍然没有很好的解决办法.而训练样本标注的正确与否决定了整个训练过程的有效性.这是在线学习的难点.

目前,样本自动获取、标注方法分为以下几种:(1)基于协同训练的方法;(2)基于背景建模的方法;(3)基于产生式模型的方法;(4)基于跟踪的方法.在协同训练的方法中<sup>[9]</sup>,同时初始化两个分类器,共同训练,互相为对方标注训练样本;基于背景建模的方法通过背景模型标注在线学习样本<sup>[10]</sup>;基于产生式模型的方法通过重构误差验证检测结果<sup>[11]</sup>,从而实现在线学习样本的获取和标注;基于跟踪的方法通过获取目标移动轨迹识别漏检及虚警<sup>[12]</sup>,并分别作为在线学习的正负样本训练目标检测模型.但是,以上方法对分类面附近的样本(即困难样本)标注错误率较高,其分类正确率难以和手工标注的离线学习分类器比较.

本文提出一种基于少量样本学习的视频目标检测跟踪算法,只需在视频第一帧指定待检测目标,即可启

动针对该类目标的自主学习,并自动构建视频目标检测跟踪系统.整个过程无需手工采集标注样本,因此,可方便的扩展到其它监控场景,使不同成像条件的监控视频,都有自己专用的目标检测跟踪系统.此外,提出在线渐进学习算法对系统进行自主学习,逐渐提高目标检测模型的精度,这样,只需采用简单的数据关联算法就可以在保持算法实时性的基础上达到与离线学习算法近似的多目标跟踪精度.

(1)实现了一个自主学习框架,可以针对某一类目标自主学习检测模型,无需先验知识和人工标注训练样本.

(2)对目标检测模型设置了正负决策边界,将检测结果分为正样本、负样本及不确定样本(困难样本),提出在线渐进学习算法逐渐缩小正负决策边界间距,从而逐渐减小这种分类不确定性.

(3)在目标检测模块中融入跟踪算法构建目标检测跟踪系统,利用跟踪收集样本训练分类器,进一步提升了系统的鲁棒性.

## 2 算法整体框架

本文算法整体框架如图1所示,该框架包含混合分类器的初始化,视频目标检测跟踪,以及混合分类器在线渐进学习三个部分.首先通过待检测目标初始化在线选择随机森林分类器(Online Selector Fern, OSF)及迭代支持向量机分类器(Iterative SVM, ISVM)组成的混合分类器,然后将该分类器用于视频中的目标检测跟踪.通过检测跟踪过程收集困难样本,用于该混合分类器的在线渐进学习,从而逐步提高分类器的目标检测精度.具体如下:

首先,在视频第一帧指定目标进行仿射变换形成训练样本集初始化由 OSF 和 ISVM 组成的混合分类器,进行视频目标检测.显然,初始训练样本集并不完备,初始检测器的检测性能较差.

对此,本文提出在线渐进学习算法迭代训练该混合分类器<sup>[13]</sup>.在 OSF 分类器中定义正负决策边界,将高于正决策边界的检测结果标注为正样本、低于负决策边界的为负样本,而位于正负决策边界之间的为困难样本.为了保证有较高的标注正确率,初始的正负决策边界间距较大.之后,通过 ISVM 分类器对困难样本进一步分类,并对正确的检测结果进行跟踪,通过整个检测跟踪过程收集困难样本.最后,将收集的正负样本更新 OSF 分类器,同时根据主动学习理论<sup>[14]</sup>逐渐减小 OSF 分类器中正负决策边界间距.当满足收敛条件时,停止迭代.

最后结合多目标数据关联跟踪构建全自主的视频目标检测跟踪系统.

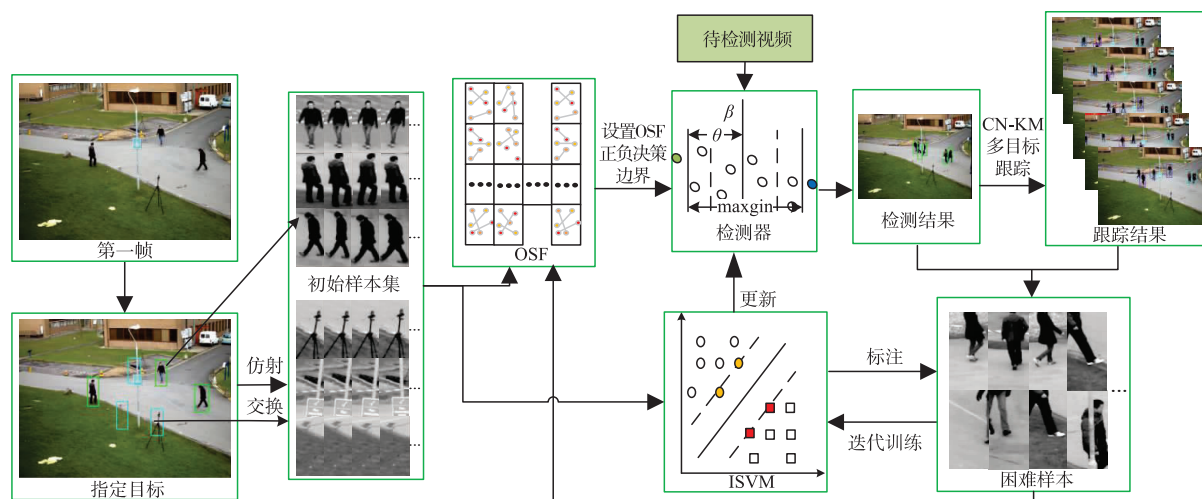


图1 基于少量样本学习的多目标检测跟踪方法流程图

下面分别介绍 OSF 及 ISVM 分类器,然后介绍在线渐进学习算法.最后介绍跟踪算法及如何通过跟踪持续提升检测器模型精,构建基于少量样本学习的视频目标检测跟踪系统.

### 3 少量样本学习的视频目标检测算法

本视频目标检测模块由 OSF 分类器及 ISVM 分类器组成,该模块能在保持实时计算速度的情况下有较高检测精度.此外,该模块可以通过在线渐进学习算法进行自主学习.

#### 3.1 OSF 分类器

随机蕨 (fern)<sup>[15]</sup> 由一系列随机二值特征组成,成功应用于不同视角下目标检测并在很大程度上提升了检测性能<sup>[16]</sup>,在检测效率和目标分类上有不俗的表现.本文在随机蕨分类器的基础上融入在线选择算

子,构造 OSF 分类器用于目标检测.训练样本集  $S^m = (f^m, c^m)$ ,  $m = 1, 2, \dots, M$ .  $M$  为样本个数,  $C$  为样本标签  $C = \{c_+, c_-\}$ ,  $c_+$  为正样本,  $c_-$  为负样本,  $f$  为样本  $N$  维特征向量空间,  $f = (f_1, f_2, \dots, f_N)$ , 每个二值特征  $f_n$  取决于样本  $x(x \in S^m)$  中  $d_{n,1}$  与  $d_{n,2}$  两点像素强度大小:

$$f_n = \begin{cases} 1, & x(d_{n,1} > d_{n,2}) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

如图 2(a) 所示,在样本  $x$  中随机选取  $s$  对像素点,即用  $s$  位二进制编码表示该样本,其对应的十进制数有  $2^s$  种可能的数值,则每个样本可对应于随机蕨中的一个数值.需要注意的是在提取特征时,一旦初始随机选取的  $s$  对像素点位置确定后,其它样本所选取像素点位置应与其保持一致.

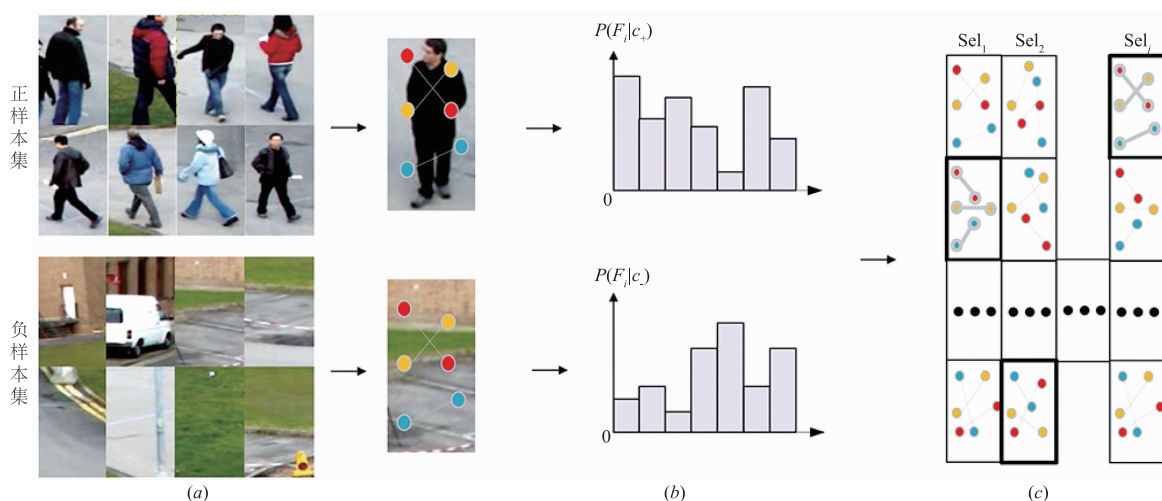


图2 OSF分类器结构

从特征集中随机抽取一组特征数量相同的子集, 每个子集称为一个随机蕨, 设每组子集包含  $s$  个特征:  $F_l = f_{l,1}, f_{l,2}, \dots, f_{l,s}$ . 因此第  $l$  个 fern 可用  $s$  个特征表示为

$$F_l = f_{l,1}, f_{l,2}, \dots, f_{l,s}, \quad l = 1, 2, \dots, L \quad (2)$$

分别统计正负样本在  $F_l$  中后验概率分布直方图、 $P(F_l C_+)$ 、 $P(F_l C_-)$  如图 2(b) 所示. 则  $F_l$  可视为一个弱分类器, 其分类结果为:

$$h_{F_l} = \frac{P(F_l(x) = k | c_+)}{P(F_l(x) = k | c_+) + P(F_l(x) = k | c_-)} \quad (3)$$

$$k \in \{2^0, 2^1, \dots, 2^s\}$$

对于多个弱分类器  $h_{F_l}^{set} = h_{F_l}(x)$ ,  $h_{F_M}(x)$  计算每个弱分类器上正负样本对应后验概率分布直方图间的巴氏距离

$$B_j = \sum_{k=2^s} \sqrt{(P(F_l(x) = k | c_+)P(F_l(x) = k | c_-))} \quad (4)$$

如图 2(c) 所示. 随机初始化  $J$  组随机蕨 (selector),  $Sel_1, \dots, Sel_j$  每组 selector 由  $L$  个 fern 组成, 选取每个 selector 中巴氏距离最小的 fern 组合, 形成最终强分类器:

$$H_{fern}(x) = \frac{1}{J} \sum_j Sel_j^l(x) \quad (5)$$

其中  $Sel_j^l(x)$  表示从第  $j$  组 selector 中挑选出的第  $l$  个 fern.

### 3.2 ISVM 分类器

SVM 是一种常见的有监督分类器, 但在本文提出的混合分类器系统中, 需对无标注的困难样本进行进一步分类, 因此采用一种半监督迭代训练 SVM 分类器的方法, 具体步骤如下:

(1) 根据仿射变换产生的初始样本集, 提取 HOG 特征初始化 SVM 分类器  $H_{svm}^0(x)$ .

(2) 对 OSF 分类产生的困难样本集  $U = \{x_{n+1}, x_{n+2}, \dots, x_{n+r}\}$  分类得到样本标签  $L^0 = \{y_{n+1}^0, y_{n+2}^0, \dots, y_{n+r}^0\}$  及其分数  $P_{svm}^0 = \{p_{n+1}^0, p_{n+2}^0, \dots, p_{n+r}^0\}$ .

(3) 采取保守方式, 分别选取  $P_{svm}^0 < -T_h$  与  $P_{svm}^0 > +T_h$  的困难样本作为正负样本更新 SVM 分类器, 并用更新后的 SVM 分类器  $H_{svm}^1(x)$  再次对  $U$  分类, 得到新的样本标签  $L^1$  与分数  $P_{svm}^1$ , 若前后两次样本标签一致, 则停止迭代. 否则, 重复上述步骤, 直至经过  $k$  次迭代后样本标签不再发生变化, 得到最终的分类器  $H_{svm}^k(x)$ . 本算法采用困难样本集迭代训练 SVM 分类器, 实验中, 迭代 2~5 次均可收敛.

### 3.3 在线渐进学习算法

由第一节可知, 混合分类器的初始样本集来自视频第一帧框指定目标的仿射变换. 因此完备性不足, 训练的初始目标检测模型性能较差. 本文提出在线渐进学习算法, 同时针对 OSF 分类器和 ISVM 分类器进行全

自主学习.

首先, 在 OSF 分类器中设定正负决策边界:

$$x_i \in \begin{cases} S_{pos}, & \text{if } H_{fern}(x_i) > \beta + \theta \\ S_{neg}, & \text{if } H_{fern}(x_i) < \beta - \theta \\ S_{hard}, & \text{otherwise} \end{cases} \quad (6)$$

其中  $\beta$  为决策超平面,  $\beta \pm \theta$  为正负决策边界,  $\theta$  决定了正负决策边界的距离. 根据 OSF 分类器计算的分数, 将检测跟踪结果  $x_i$  标注为正样本 ( $S_{pos}$ )、负样本 ( $S_{neg}$ ) 和困难样本 ( $S_{hard}$ ), 再由 ISVM 分类器对困难样本进行学习. 根据主动学习理论, 该决策边界距离可通过提高分类器性能而逐渐缩短. 其中  $\theta$  为:

$$\theta = 1 - \xi \lambda_k \quad (7)$$

$\xi$  为敏感系数,  $\lambda_k$  衡量 OSF 分类器性能, 可以通过

$$\lambda_k = \frac{M_r}{M_{tol}} \quad (8)$$

计算得到.  $M_{tol}$  表示需要标注的所有困难样本个数,  $M_r$  为 OSF 和 ISVM 将困难样本分为同一类的个数.

主动学习需要人为标注困难样本更新分类器模型. 与该过程类似, 在本框架中利用 ISVM 分类器代替人对困难样本进行标注, 通过在线渐进学习算法, 逐步减少正负决策边界间距, 进而逐步提高分类器性能. 算法具体流程如下.

输入: 视频第一帧中指定几个待检测目标及背景仿射变换后得到初始样本集  $I_0 = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ , 初始化困难样本集

$$Set_{hard} \in \varphi, \beta = 0.5, \theta = 0.5, \xi = 0.85$$

(1) 利用  $I_0$  初始化 SVM 分类器  $H_{svm}^0$  和 OSF 分类器  $H_{fern}^0$

(2)  $t = 0$ .

(3) 若  $\theta > 0.2$

(4)  $H_{fern}^t$  进行目标检测, 并跟踪检测结果.

(5) 收集困难样本并添加至  $Set_{hard}$ , 统计  $Set_{hard}$  中 hard sample 个数  $N_{hard}$ .

(6) 当  $N_{hard} > 100$  时, 使用  $Set_{hard}$  半监督迭代训练  $H_{svm}^t$  并标注  $Set_{hard}$ .

(7) 将标注后的  $Set_{hard}$  训练更新  $H_{fern}^t$ .

(8) 更新  $\theta$ ,  $Set_{hard} \in \varphi, N_{hard} = 0$ .

(9)  $t = t + 1$

(10) 跳转至步骤 3.

输出: 训练好的分类器  $H_{fern}(x)$  和  $H_{svm}(x)$ .

## 4 CN-KM 多目标跟踪框架

Danelljan 等人通过对样本的全局学习和特征降维技术将色彩属性 (CN) 引入 CSK 跟踪框架, 表现出卓越的跟踪性能与极快的跟踪速度, 能满足现实场景中实

时跟踪的应用。

#### 4.1 CN 跟踪算法

CN 跟踪器主要包括分类器训练模型更新和自适应颜色选取两部分。训练模型由学习目标表观模型  $\hat{x}^p$  和分类器参数  $A^p$  两部分组成, 设  $x$  为  $M \times N$  的包含目标区域, 通过循环采样得稠密样本  $x_{m,n}, m \in \{0, \dots, M-1\}, n \in \{0, \dots, N-1\}$ , 将样本用高斯函数标记得到带标签的样本  $(x_{m,n}, y_{m,n})$  训练分类器。训练过程中, 通过不断调整  $\omega$  使代价函数最小:

$$\varepsilon = \sum_{j=1}^p \beta_j \left( \sum_{m,n} \left| \langle \varphi(x_{m,n}^j), \omega^j \rangle - y^j(m,n) \right|^2 + \lambda \langle \omega^j, \omega^j \rangle \right) \quad (9)$$

$$\omega^j = \sum_{k,l} a(k,l) \varphi(x_{k,l}^j) \quad (10)$$

其中,  $p$  为当前帧索引,  $\beta_j$  为第  $j$  帧图像所占权重,  $\varphi(\cdot)$  是通过内积核  $\langle \varphi(f), \varphi(g) \rangle = k(f,g)$  定义的映射到 Hilbert 空间的函数。常数  $\lambda$  为正则化参数。  $y$  为以目标中心点为中心的二维高斯函数。当  $A^p$  满足下式时代价函数最小。

$$A^p = \frac{\sum_{j=1}^p \beta_j Y^j U_x^j}{\sum_{j=1}^p \beta_j U_x^j (U_x^j + \lambda)} \quad (11)$$

其中  $U_x^j = \text{Fourier}(u_x)$ ,  $u_x = k(x_{m,n}^j, x^j)$ ,  $Y = \text{Fourier}(y)$ ,  $A = \text{Fourier}(a)$   $A^p = \frac{A_N^p}{A_D^p}$  训练模型更新方式如下:

$$A_N^p = (1 - \gamma) A_N^{p-1} + \gamma Y^p U_x^p \quad (12)$$

$$A_D^p = (1 - \gamma) A_D^{p-1} + \gamma U_x^p (U_x^p + \lambda) \quad (13)$$

$$\hat{x}^p = (1 - \gamma) \hat{x}^{p-1} + \gamma x^p \quad (14)$$

其中,  $\gamma$  为学习速率。

CN 跟踪器将 RGB 投影到黑、蓝、棕、灰、绿、橙、粉、紫、红、白、黄 11 维特征空间得到颜色特征。为了提高算法运算速度, 利用 PCA 主成分分析思想进行降维处理。降维后样本  $\tilde{x}$  可表示为  $\tilde{x} = \mathbf{B}_p^T \mathbf{x}^p$ 。用重构最小代价函数作为决策函数, 得到降维映射矩阵  $\mathbf{B}_p$ :

$$\min_{\mathbf{B}_p} \left[ a_p \frac{1}{MN} \sum_{m,n} \left\| \hat{x}^p(m,n) - \mathbf{B}_p \mathbf{B}_p^T \hat{x}^p(m,n) \right\|^2 + \sum_{j=1}^p \sum_{k} a_j \lambda_j \left\| \mathbf{b}_j^{(k)} - \mathbf{B}_p \mathbf{B}_p^T \mathbf{b}_j^{(k)} \right\|^2 \right] \quad (15)$$

其中,  $\mathbf{b}_j$  为投影矩阵  $\mathbf{B}_p$  的列向量。通过  $\tilde{x}$  更新训练模型计算响应分数  $\hat{y}: \hat{y} = \text{Fourier}^{-1}(A U_z)$  其  $U_z = \text{Fourier}(u_z)$ ,  $u_z(m,n) = k(z_{m,n}, \hat{x})$ 。响应分数最大的位置即为新目标的中心位置。

#### 4.2 KM 匹配算法

Kuhn-Munkres(KM) 算法<sup>[17]</sup> 作为框架的核心算法, 通过使用欧式距离, 外观模型度量值作为前后两组帧检

测结果的权重, 从而构成了一幅具有  $(U, V)$  的加权完全二分图, 其中  $U$  是上一帧的检测结果,  $V$  是当前帧的检测结果。然后再通过 KM 算法来获得其的最佳匹配, 从而获得每一个跟踪目标在这前后两帧之间的运动轨迹。

#### 4.3 CN-KM 多目标跟踪框架

为了克服单独使用 CN 算法出现跟踪失败问题, 以及减少单独使用 Kuhn-Munkres (KM) 算法进行耦合而产生的对检测结果的依耐性, 本文将两算法结合, 提出 CN-KM 多目标跟踪框架。在目标检测模块中加入 CN 跟踪算法与 KM 匹配算法跟踪检测结果, 并自主选取合适跟踪结果更新分类器, 构建基于少量样本学习的多目标检测跟踪系统, 流程图如图 3 所示。

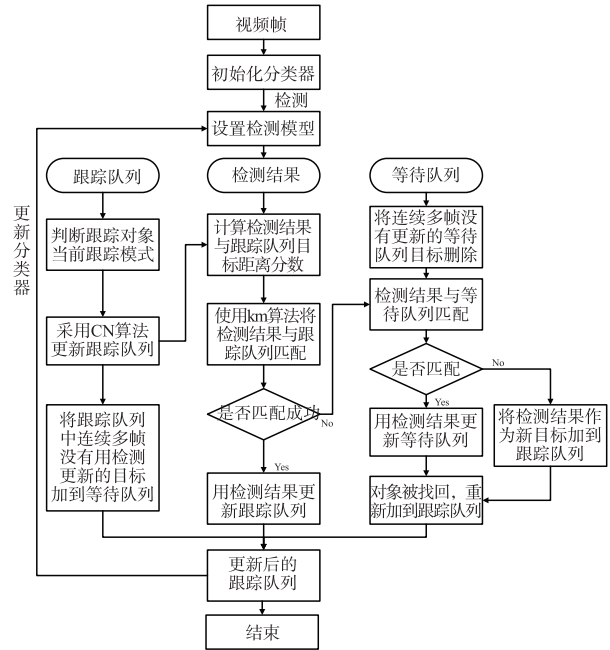


图3 CN-KM 多目标跟踪框架

整个多目标跟踪流程主要包括检测目标与跟踪目标的匹配、处理未匹配上的跟踪目标、处理未匹配上的检测目标以及确定新目标等四个步骤。

首先, 对当前帧检测到的目标, 先和跟踪队列进行匹配, 计算检测目标和跟踪目标两两间距 (为匹配距离即为 CN 算法和 KM 算法切换的阈值), 距离越小, 表明检测和跟踪的目标为同一目标的概率越大, 采用 KM 匹配算法将检测跟踪目标进行匹配, 对成功匹配上的目标采用检测结果更新跟踪的目标的位置。

对没有与检测队列目标匹配上的跟踪目标, 采用 CN 算法更新目标位置。

对没有与跟踪队列目标匹配的检测结果, 再采用 KM 算法与等待队列目标进行匹配, 若有目标匹配成功, 则表明该目标消失一段时间后, 被再次检测出, 将该目标重新加入到跟踪队列进行跟踪。

## 5 实验结果及分析

### 5.1 参数设置

实验参数设置如下:初始化 10 组 selector ( $J = 10$ ) 每组 selector 由 10 个 fern 构成 ( $L = 10$ ), 每个 fern 包含 6 个特征 ( $s = 6$ ). 提 HOG 特征时 block 为  $16 \times 16$  像素, cell 为  $8 \times 8$  像素. 实验环境为: Intel Core i5 3.2 GHz 处理器, 8GB RAM, 64 位 Windows 操作系统.

### 5.2 评价指标

本文采用下列指标对本算法进行评价.

召回率 ( $Rell$ ): 检测算法得到的正确分类数占真实目标的比例,  $Rell = TP/GT$ .

精确率 ( $Prcsn$ ): 体现检测算法的正确率,  $Prcsn = TP/(TP + FP)$

F 值 ( $F$ -measure) 度量检测算法性能的综合指标, 反映了检测算法的灵敏性和有效性,  $F$ -measure =  $2 * Rell * Prcsn / (Rell + Prcsn)$

多目标跟踪准确率 ( $MOTA$ ):  $MOTA = 1 - (FP + FN + IDS)/GT$

多目标跟踪精确率 ( $MOTP$ ): 跟踪结果与对应目标的平均偏差, 值越大表明跟踪的轨迹越接近目标实际的运动轨迹.

其中,  $TP$  ( $True\ positive$ ) 表示正确检测的目标个数,  $GT$  ( $Ground\ truth$ ) 为实际目标个数,  $FP$  ( $False\ positive$ ) 表示虚警个数,  $FN$  ( $False\ negative$ ) 为漏检个数,  $IDS$  ( $Identity\ switches$ ) 为跟踪过程中目标身份转换数.

### 5.3 实验结果及分析

#### 5.3.1 视频目标检测结果分析

实验采用 CAVIAR<sup>[18]</sup> 数据集中 WalkByShop1-front 序列以及自己准备的车辆视频数据 Hx 测试目标检测算法, 将本算法分别与在线学习算法和离线学习算法比较验证本算法有效性, 为后续目标跟踪提供可靠的初始信息.

分别采用本算法和其他几种经典在线学习方法对 WalkByShop1front 序列进行目标检测, 部分检测效果图如图 4 第一行所示, 比较结果如表 1 所示, 相应 PR 曲线图如图 5(a) 所示. 从表中明显看到本算法检测性能远高于其他几种在线学习方法.

Hx 共 6415 帧, 分辨率为  $1224 \times 1024$ . 对 Hx 序列分别采取本算法和其它几种有监督学习算法进行测试, 其中: ACF<sup>[19]</sup>: 采用 200 正样本 100 个负样本训练得到. SVM: 采用与 ACF 同样本提取 HOG 特征训练得到. Fern: 人为标注样本在线训练 OSF 分类器至 848 帧, 共训练正样本 246 个, 负样本 245 个. FernSVM: 上述有监督训练的 SVM 与 OSF 组成的混和分类器. 部分检测效果图如图 4 第二行所示, 比较结果如表 1 所示, 相应 PR 曲线图如图 5(b) 所示, 虽然 ACF 检测性能最佳, 但耗费大量人

力筛选样本训练分类器. 而本算法在没有任何人为干预情况下获得了比其他几种离线方法更好的检测性能.



图4 检测结果图

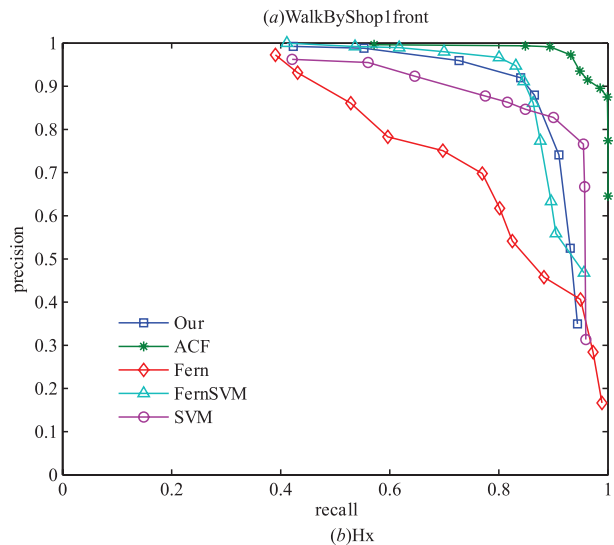
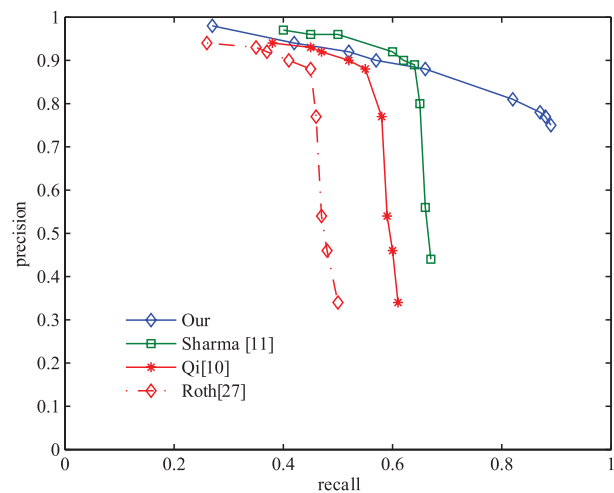


图5 检测算法比较

#### 5.3.2 视频目标检测跟踪结果分析

在检测算法上加入跟踪算法, 实现一个无监督视频目标检测跟踪系统, 并选用公共数据集 PETS2009-S2L1<sup>[20]</sup> 和 TUD-standemitte<sup>[21]</sup> 对本体系进行测试.

PETS2009-S2L1 视频序列共 795 帧, 分辨率为  $768 \times 576$ . 视频中行人运动轨迹复杂且遮挡严重, 很大程度上增加了目标检测跟踪难度. 部分测试效果图如图 6

(1)所示,其中(a)~(d)为检测效果图,由于目标漏检导致目标轨迹出现断裂情况.通过加入跟踪算法减少漏检修复了轨迹断裂问题,相应跟踪结果如图6(e)~

(h)所示.实验比较结果如表2所示.其中文献[14,16]通过读取 Ground truth 获取目标进行跟踪得到测试结果,而本文通过全自主学习即获得与之可比的性能.

表 1 检测算法比较

Online	WalkByShop1 front	Method	Roth <sup>[25]</sup>	Qi <sup>[9]</sup>	Sharma <sup>[12]</sup>	Our	-
		F-measure	0.5955	0.6769	0.7446	0.8225	-
Offline	Hx	Method	ACF	Fern	FernSVM	SVM	Our
		Background modeling	No	No	No	No	No
		F-measure	0.9518	0.7320	0.8771	0.8624	0.8779

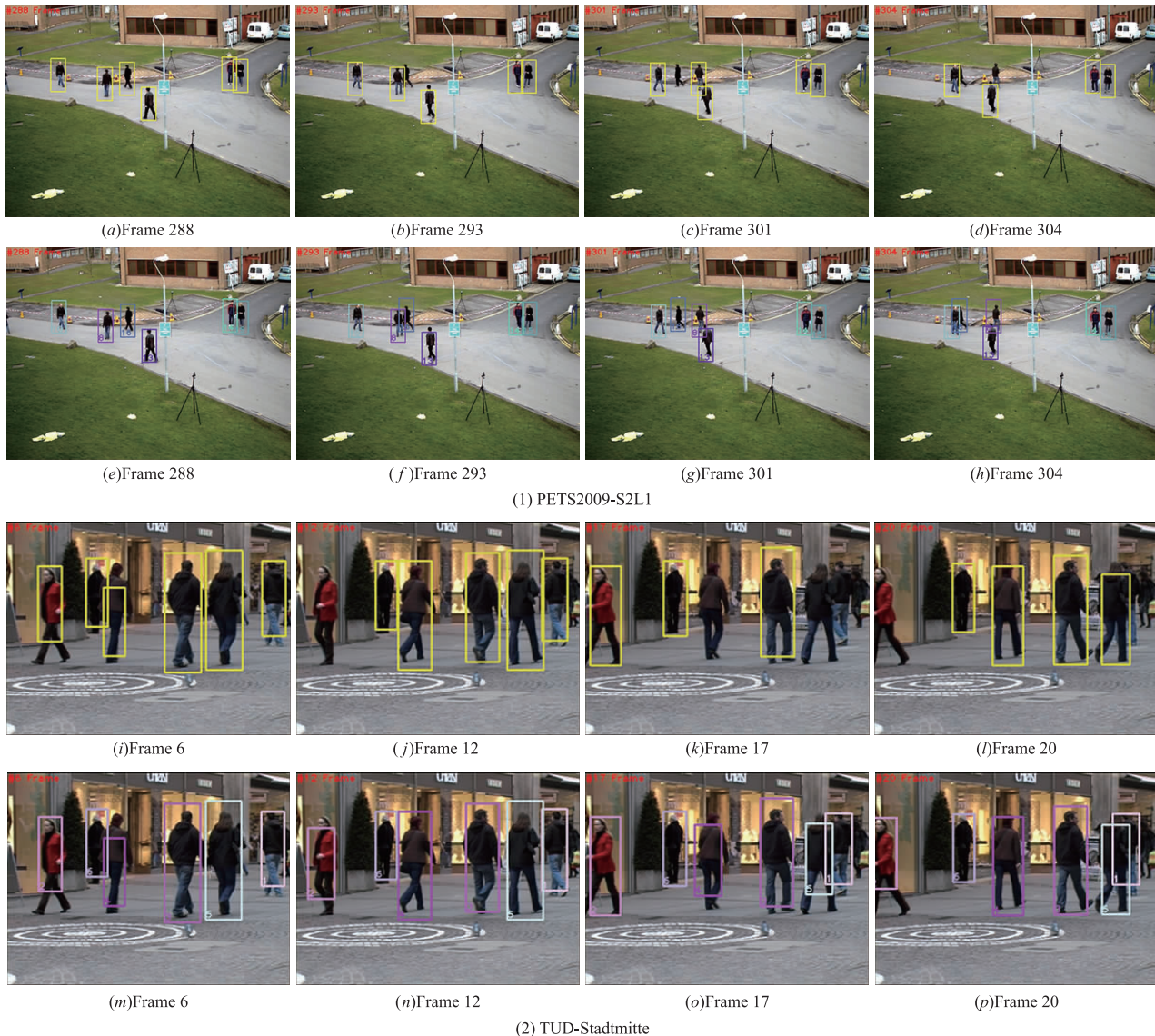


图 6 本算法测试效果图

TUD-standemitte 视频序列有 179 帧,分辨率为  $640 \times 480$ .比较结果如表 3 所示.从表中可以看到,本算法 *MOTA* 最佳,*Recall* 相对较低是由于视频中目标尺度相差较大,且目标遮挡严重,在检测过程中,部分尺度较小、遮挡严重的目标难以被检测出,如图 6(2)中(i)~(l)

所示.加入跟踪后相应帧检测跟踪结果如图 6(2)中(m)~(p)所示,通过加入跟踪算法,使漏检目标被成功跟踪,修复了轨迹断裂问题.

将本算法成功移植到 NVIDIA 嵌入式平台 Jetson TK1 中,如图 7(a)所示. TK1 预装了 Tegra 操作系统,

TK1 开发工具包为开发人员提供了全套开发工具,可以用于机器人学、计算机视觉、自动驾驶汽车等计算密集型系统. 运行帧率如表 4 所示.

表 2 PETS2009-S2L1 序列跟踪结果

Method	MOTA%	MOTP%	Rec%	Prcsn%
ECCV2012 <sup>[22]</sup>	53.4	70.9	72.6	80.3
CVPR2013 <sup>[23]</sup>	90.8	74.2	97.1	94.4
CVPR2015 <sup>[24]</sup>	87.9	64.5	98.6	90.8
IET2016 <sup>[23]</sup>	90.0	56.8	91.7	98.5
Our	88.4	66.7	91.1	97.9

表 3 TUD-Stadtmitte 序列跟踪结果

Method	MOTA%	MOTP%	Rec%	Prcsn%
ECCV2012 <sup>[22]</sup>	60.0	56.5	83.7	78.7
CVPR2013 <sup>[23]</sup>	53.5	72.4	79.1	82.0
CVPR2015 <sup>[24]</sup>	69.7	53.4	74.7	94.3
IET2016 <sup>[23]</sup>	72.4	52.6	75.1	99.8
Our	73.4	65.7	75.6	98.6

表 4 帧率

Datasets	System	Background modeling	Detection	Detection and tracking
S2L1	PC/windows	Yes	61.5	27.6
	TK1/ubuntu	Yes	17.7	9.2
Stadtmitte	PC/windows	No	11.6	6.5
	TK1/ubuntu	No	6.7	5.1



图 7 NVIDIA Jetson TK1 嵌入式平台

## 6 结语

本文提出了一种基于少量样本学习的多目标检测跟踪方法,在没有人干预的情况下通过少量样本学习完成分类器训练实现目标检测跟踪,从而很大程度上减少了人的工作量. 实验结果表明本算法在不同的监控场景下的多目标检测跟踪中均表现出良好性能.

后续,将采用更多的数据集对本算法进行测试,并在本算法框架下融入新特征训练分类器,进一步提升算法整体性能.

### 参考文献

[1] Gwak J. Multi-object tracking through learning relational

appearance features and motion patterns[J]. Computer Vision and Image Understanding, 2017, 162(9): 103 – 115

- [2] Felzenszwalb P F, Girshick R B, Mcallester D, et al. Object detection with discriminatively trained part based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627 – 1645.
- [3] Bourdev L, J Malik, Poselets. Body part detectors trained using 3d human pose annotations [A]. 2009 IEEE 12th International Conference on Computer Vision [C]. Kyoto, Japan: IEEE, 2010. 1365 – 1372.
- [4] Yang Y, Shu G, Shah M. Semi-supervised Learning of Feature Hierarchies for Object Detection in a Video [A]. 2013 IEEE Conference on Computer Vision and Pattern Recognition [C]. Portland, USA: IEEE, 2013. 1650 – 1657
- [5] Wang X, Li W, Wang M. Transferring a generic pedestrian detector towards specific scenes [A]. 2012 IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence, RI, USA: IEEE, 2012. 3274 – 3281.
- [6] Cinbis R G, Verbeek J, et al. Weakly supervised object localization with multi-fold multiple instance learning [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(1): 189 – 203.
- [7] Liu Y, Wang Y, et al. Soft hough forests; Generalized hough transform based object detection from soft-labelled training data [J]. Pattern Recognition, 2016, 60: 145 – 156.
- [8] Singh K K, Xiao F, Yong J L. Track and transfer; Watching videos to simulate strong human supervision for weakly-supervised object detection [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas, USA: IEEE, 2016. 3548 – 3556.
- [9] Qi Z, Xu Y, Wang L, Song Y. Online multiple instance boosting for object detection [J]. Neur-computing, 2011, 74(10): 1769 – 1775.
- [10] Nair Clark V, J J. An unsupervised online learning framework for moving object detection [A]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. Washington, DC, USA: IEEE, 2004. 317 – 325.
- [11] Wang X, Hua G, Han T X. Detection by detections; Non-parametric detector adaptation for a video [A]. Computer Vision and Pattern Recognition [C]. Providence, RI, USA: IEEE, 2012. 350 – 357.
- [12] Sharma P, Nevatia R. Ecient detector adaptation for object detection in a video [A]. Computer Vision and Pattern Recognition [C]. Kyoto, Japan: IEEE, 2013. 3254 – 3261.
- [13] Wozniak M, Grana M, et al. A survey of multiple classifier systems as hybrid systems [J]. Information Fusion, 2014, 16(1): 3 – 17.
- [14] Villamizar M, Andrade-Cetto J, Sanfeliu A, et al. Boosted

- random ferns for object detection [ J ]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2018, 40 ( 2 ): 272 – 288.
- [ 15 ] Ozuysal M, Calonder M, et al. Fast keypoint recognition using random ferns [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 ( 3 ): 448 – 461.
- [ 16 ] 罗大鹏, 罗琛, 魏龙生, 韩家宝, 王勇, 马丽. 基于在线随机蕨分类器的实时视觉感知系统 [ J ]. 电子学报, 2016, 44 ( 5 ): 1139 – 1148.  
LUO Da-peng, LUO Chen, WEI Long-sheng, HAN Jiabao, WANG Yong, MA Li. Real time visual perception system based on online fern classifier [ J ]. Acta Electronica Sinica, 2016, 44 ( 5 ): 1139 – 1148. ( in Chinese )
- [ 17 ] Kuhn H. W. Variants of the hungarian method for assignment problems [ J ]. Naval Research Logistics, 2010, 3 ( 4 ): 253 – 258.
- [ 18 ] CAVIAR DATA [ D/OL ]. <http://groups.inf.ed.ac.uk/vision/CAVIAR/>, 2018.
- [ 19 ] Dollár P, Appel R, et al. Fast feature pyramids for object detection [ J ]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36 ( 8 ): 1532 – 1545.
- [ 20 ] Ferryman J, Shahrokni A. PETS2009: Dataset and challenge [ A ]. Proceedings of 2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance [ C ]. Snowbird, USA; IEEE, 2009. 1 – 6.
- [ 21 ] Andriyenko A, Schindler K. Multi-target tracking by continuous energy minimization [ A ]. Computer Vision and Pattern Recognition [ C ]. Colorado Springs, USA; IEEE, 2011. 1265 – 1272.
- [ 22 ] Fragkiadaki K, Zhang W, et al. Two-granularity tracking: mediating trajectory and detection graphs for tracking under occlusions [ A ]. European Conference on Computer Vision [ C ]. Florence, Italy; Springer, 2012. 552 – 565.
- [ 23 ] Milan A, Schindler K, Roth S. Detection and trajectory level exclusion in multiple object tracking [ A ]. IEEE Conference on Computer Vision and Pattern Recognition [ C ]. Portland, USA; IEEE, 2013. 3682 – 3689.
- [ 24 ] Milan A, Leal-Taixé L, Schindler K, Reid I. Joint tracking and segmentation of multiple targets [ A ]. IEEE Conference on Computer Vision and Pattern Recognition [ C ]. Boston, USA; IEEE, 2015. 5397 – 5406.
- [ 25 ] Roth P M, Grabner H, et al. Online conservative learning for person detection [ A ]. International Conference on Computer Communications and Networks [ C ]. Beijing, China; IEEE, 2005. 223 – 230.

### 作者简介



曾志鹏 男, 1993 年生, 湖北仙桃人, 中国地质大学(武汉)机械与电子信息学院硕士研究生. 主要研究方向为计算机视觉、机器学习.  
E-mail: zengzhipeng@cug.edu.cn



杜国庆 男, 1993 年生, 湖北武汉人, 中国地质大学(武汉)机械与电子信息学院硕士研究生. 主要研究方向为发育神经网络、自主学习、目标跟踪与检测.  
E-mail: guoqingdu@foxmail.com



罗大鹏(通信作者) 男, 1978 年出生, 四川泸定县人, 于华中科技大学自动化学院获博士学位, 现为中国地质大学(武汉)副教授, 主要研究方向为计算机视觉、图像/视频处理、机器学习等.  
E-mail: luodapeng@cug.edu.cn