

## 基于优化协同训练理论的自适应融合跟踪

郑超<sup>1,2,3</sup>, 陈杰<sup>4</sup>, 杨星<sup>1,2,3</sup>, 殷松峰<sup>1,2,3</sup>, 冯云松<sup>1,2,3</sup>

(1. 电子工程学院, 安徽, 合肥 230037;

2. 红外与低温等离子体安徽省重点实验室, 安徽, 合肥 230037;

3. 脉冲功率激光技术国家重点实验室, 安徽, 合肥 230037;

4. 安徽建筑大学电子与信息工程学院, 安徽, 合肥 230037)

**摘要:**针对基于可见光和红外图像的分析型融合跟踪算法在复杂环境下的鲁棒性不高, 提出一种新颖的基于优化协同训练理论的自适应分析型融合跟踪算法. 首先, 通过加权多示例学习 boosting 技术分别从基于可见光和红外图像的弱分类器池中实现判别能力最好的弱分类器挑选, 减弱引入的误差样本对联合分类器判别能力的影响; 然后, 在自适应先验知识引入机制辅助下, 完成分类器样本包的协同训练更新, 减小相互引入误差样本的概率; 最后, 通过误差模型完成算法有效性分析. 多组序列跟踪的对比实验结果展示了该算法各部分对提高跟踪鲁棒性的贡献, 验证了该算法相比于基于单源图像或其它融合机制的跟踪算法更好的鲁棒性.

**关键词:**分析型融合; 协同训练; 加权多示例学习 boosting; 先验知识

**中图分类号:** TP391.41 **文献标识码:** A

## Adaptive fusion tracking based on optimized co-training framework

ZHENG Chao<sup>1,2,3</sup>, CHEN Jie<sup>4</sup>, YANG Xing<sup>1,2,3</sup>, YIN Song-Feng<sup>1,2,3</sup>, FENG Yun-Song<sup>1,2,3</sup>

(1. Electronic Engineering Institute, Hefei 230037, China;

2. Key Laboratory of Infrared and Low Temperature Plasma of Anhui Province, Hefei 230037, China;

3. State Key Laboratory of Pulsed Power Laser Technology, Hefei 230037, China;

4. Electronics and Information Engineering Institute, Anhui Jianzhu University, Hefei 230037, China)

**Abstract:** As analytical fusion tracking algorithms based on visible and infrared images always have low robustness in complex environment, a novel adaptive analytical fusion tracking algorithm based on optimized co-training framework was proposed. Firstly, selecting the most discriminative weak classifiers from weak classifier pools based on infrared and visible images respectively are achieved by weighted multiple instance learning boosting technology, which relieving classifiers' discriminative capacity decreasing owing to the added error positive samples. Then, classifiers' sample bags are updated by co-training criterion under the help of adaptive prior knowledge import strategy. Lastly, efficiency analysis of the proposed algorithm was achieved based on error model. Comparative experiments on multiple sequences tracking show the contributions for improving tracking robustness from different parts of the proposed algorithm, and then, demonstrate that it outperforms state-of-the-art tracking algorithms based on single source image or other fusion schemes on robustness.

**Key words:** analytical fusion, co-training, weighted multiple instance learning boosting, prior knowledge

**PACS:** 07.05.Pj

收稿日期: 2015-08-01, 修回日期: 2016-04-01

Received date: 2015-08-01, revised date: 2016-04-01

基金项目: 安徽高校自然科学基金重大项目资助 (KJ2015ZD14), 国家自然科学基金 (61405248, 61503394), 安徽省自然科学基金 (1408085QF131, 1508085QF121)

Foundation items: Higher Education Institutes Nature Science Research Project of Anhui Province in China (KJ2015ZD14), NSFC (61405248, 61503394), NSF of Anhui province (1408085QF131, 1508085QF121)

作者简介 (Biography): 郑超 (1987-), 男, 博士研究生, 主要从事光信息处理方面的研究. E-mail: zhengchao\_eei@163.com

## 引言

红外传感器获取的自发辐射信息与可见光传感器获取的反射信息具有很好的互补性,因此在多源传感器计算机视觉系统中广泛应用<sup>[1]</sup>. 本文研究如何通过红外和可见光图像的分析型融合,提高复杂环境下目标跟踪鲁棒性问题.

不同于以提高图像质量为目的的表征型融合方式,基于视频的分析型融合面向具体的定位、跟踪和识别等问题,由于研究起步较晚,目前还没有形成完整的理论框架<sup>[2]</sup>. 文献[3]、[4]虽然同时利用了红外和可见光图像提高了目标定位或跟踪的精度,但采用的方法基于两种图像信息的简单传递或级联,并没有实现真正意义上信息的融合;文献[5]、[6]分别讨论了基于粒子滤波和均值漂移两种经典跟踪理论的红外与可见光图像融合跟踪问题,由于都严重依赖于基于背景模型的特征判别性度量,当存在背景聚类、近似目标等干扰条件时,鲁棒性不好.

文献[7]最早提出了基于协同训练的半监督分类器,由于能同时利用两类互补特征训练分类器,因此在解决目标分类问题上取得了令人满意的结果. 为了解决判别式跟踪中“自学习”导致的跟踪漂移问题,研究者很快将协同训练理论与常用的贝叶斯分类器结合,解决基于可见光图像的目标跟踪问题<sup>[8-10]</sup>,但上述算法忽略了存在复杂干扰条件时,可见光图像的特征很难满足协同训练工作条件——特征的独立性和充分性,导致在训练更新时容易相互引入误差样本,使分类器性能迅速下降.

考虑到基于不同成像机理的红外与可见光图像特征不但能始终保持相互独立性,而且具有很强的互补性,本文将协同训练思想作为一种新的方法用于解决红外与可见光图像的自适应融合跟踪问题. 主要完成的工作包括:首先,提出了一种基于优化协同训练理论的自适应融合跟踪算法,流程图如图1所示,该算法不但通过加权多示例(Weighted Multiple Instance Learning, WMIL) boosting 技术,在误差样本干扰的情况下,也能够相对准确地挑选出判别能力较好的弱分类器,因此减弱了引入的误差样本对联合分类器判别性能的影响,而且提出的自适应先验知识引入机制可以从根本上减小协同训练更新样本包过程中相互引入误差样本的概率;然后,对本文算法的有效性进行了理论分析;最后,通过仿真实验验证了本文算法各部分对整体鲁棒性提高的贡献以及相比于基于单源图像或其它融合机制的跟踪算

法更好的鲁棒性.

## 1 基于贝叶斯分类器的判别式跟踪基本理论

判别式跟踪理论将视觉跟踪问题视作目标与局部背景的二分类问题,因此将二值分类器作为目标的外观模型,通过分类器的在线更新适应目标外观的变化. 目标在上一帧的位置记作  $l_{t-1}^*$ ,则在当前帧图像中以  $l_{t-1}^*$  为中心,  $R$  为半径的领域内提取若干图像块构成候选样本集  $X^R = \{x \mid R > \|l_t(x) - l_{t-1}^*\| \}$ ,  $l_t(x) \in R^2$  代表图像块  $x$  在当前帧的坐标位置. 首先基于贝叶斯理论计算候选样本的后验概率,然后按照贪婪策略搜索后验概率最大值对应的图像块,即为当前帧的目标位置,如式(1)所示,

$$\begin{aligned} l_t^* &= \operatorname{argmax}_{x \in X^R} p(y=1|x) \\ &= \operatorname{argmax}_{x \in X^R} \frac{p(x|y=1)p(y=1)}{p(x|y=1)p(y=1) + p(x|y=0)p(y=0)} \\ &= \operatorname{argmax}_{x \in X^R} \sigma(H_M(x)) \end{aligned} \quad (1)$$

其中  $\sigma(x) = 1/(1 + e^{-x})$  为 Sigmoid 函数,  $H_M(x) = \ln((p(x|y=1) \cdot p(y=1))/(p(x|y=0) \cdot p(y=0)))$ ,  $y \in \{0,1\}$  为二值变量,对应样本的标签. 假设  $x$  代表的图像块可以基于  $f(x) = \{f_1(x), f_2(x) \dots f_M(x)\}$  特征空间表征且各子特征  $f_m(x)$ ,  $m = 1, 2 \dots M$  之间相互独立,先验概率  $p(y=0) = p(y=1)$ ,则式(1)中的判别函数  $H_M(x)$  可以写成

$$\begin{aligned} H_M(x) &= \ln\left(\frac{p(f(x)|y=1)}{p(f(x)|y=0)}\right) \\ &= \sum_{m=1}^M \ln\left(\frac{p(f_m(x)|y=1)}{p(f_m(x)|y=0)}\right) \\ &= \sum_{m=1}^M h_m(x) \end{aligned} \quad (2)$$

代入式(2)后,式(1)可以改写成

$$l_t^* = \operatorname{argmax}_{x \in X^R} H_M(x) = \operatorname{argmax}_{x \in X^R} \sum_{m=1}^M h_m(x) \quad (3)$$

其中判别函数  $H_M(x)$  称为强分类器,由弱分类器  $h_m(x)$  线性关联组成. 对于许多数据组成的集合,文献[11]论证了高维向量的随机映射是近似高斯的,因此可以采用高斯分布对式(2)弱分类器  $h_m(x) = \ln\left(\frac{p(f_m(x)|y=1)}{p(f_m(x)|y=0)}\right)$  中的条件概率建模,则该高斯模型包含 4 个参数,即:  $p(f_m(x)|y=1) \sim N(\mu_m^1, \delta_m^1)$ ;  $p(f_m(x)|y=0) \sim N(\mu_m^0, \delta_m^0)$  参数  $(\mu_m^1, \delta_m^1)$  和  $(\mu_m^0, \delta_m^0)$  的更新准则相同,以  $(\mu_m^1, \delta_m^1)$  为例,如式(4)、(5)所示,

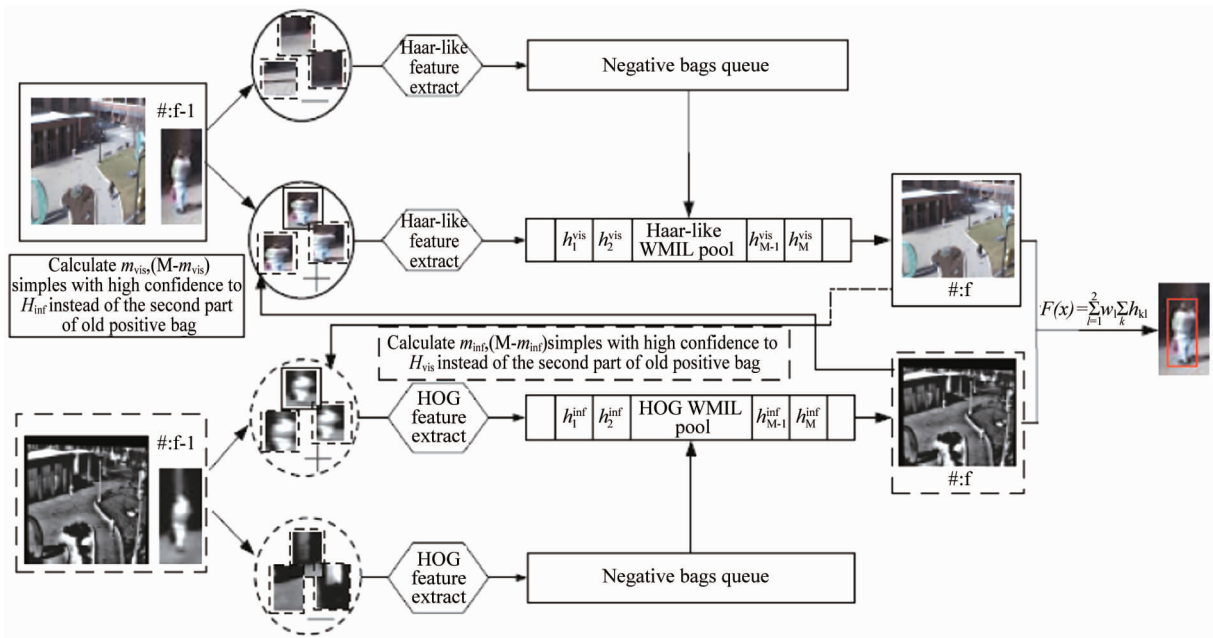


图1 本文提出的跟踪器流程图(番号 f-1 表征上一帧图像, f 表征当前帧图像;“+”表征正样本包,“-”表征负样本包,带实线矩形框的样本代表权重最高的正样本)

Fig. 1 Diagram of proposed tracker (“f-1” and “f” indicate the previous and current image respectively; “+” and “-” indicate positive and negative instance bag; positive sample with solid rectangular box has the max weight)

$$\mu_m^1 = \lambda \mu_m^1 + (1 - \lambda) \bar{\mu} \quad , \quad (4)$$

$$\delta_m^1 = \sqrt{\lambda(\delta_m^1)^2 + (1 - \lambda) \frac{1}{M} \sum_{m=1}^M (f_m(x) - \bar{\mu})^2 + \lambda(1 - \lambda)(\mu_m^1 - \bar{\mu})^2} \quad , \quad (5)$$

其中,  $\bar{\mu} = \frac{1}{M} \sum_{m=1}^M f_m(x)$ ,  $\lambda$  是学习率, 调节更新的速度. 综上, 判别式跟踪的基本过程如下: 首先以上一帧目标位置为中心, 在以  $r$  为半径的圆形区域内均匀提取若干图像块作为正样本集  $X^r = \{x \mid r > \|l_{t-1}(x) - l_{t-1}^*\| \}$ , 在  $r < a < b$  的环形区域内随机提取若干图像块作为负样本集  $X^{a,b} = \{x \mid a < \|l_{t-1}(x) - l_{t-1}^*\| < b\}$ ; 然后利用正、负样本集更新强分类器  $H_M(x)$ ; 最后根据式(1), 在当前帧的候选样本中找到后验概率最大值图像块对应的位置为目标位置  $l_t^*$ , 重复上述步骤完成目标的持续跟踪.

## 2 本文算法

### 2.1 分别构建基于红外和可见光图像特征的弱分类器池

从特征互补性的角度考虑, 本文依照文献[12]中的做法, 利用积分图法<sup>[13]</sup>提取可见光图像的 Haar-like 特征构建弱分类器池, 每个特征由个矩形框组成, 每个矩形框都有各自的权重, Haar-like 特征

本质上是矩形框内像素值的加权和, 作为经典的局部特征, 针对目标局部遮挡和形变有较好的鲁棒性; 提取红外图像的梯度方向直方图 (Histogram Oriented Gradient, HOG) 构建弱分类器池, HOG 特征作为一种有效纹理特征, 被广泛应用于目标探测与跟踪<sup>[14]</sup>.

### 2.2 基于 WMIL boosting 的在线弱分类器挑选

由章节 1 可知, 贝叶斯分类器可以写成弱分类器线性关联组成的强分类器形式, 文献[15]指出由 boosting 技术挑选出若干个判别能力最好的弱分类器组成强分类器, 可以有效提高分类器的整体性能. 文献[8]提出基于多示例 boosting, 虽然相比文献[9]的 AdaBoost 方法能一定程度上缓解相互引入的误差训练样本对弱分类器挑选精度的影响, 但因为在计算正样本包的概率时没有考虑不同正样本的重要性信息, 导致距离目标位置较远的正样本(可能包含部分背景信息)对包概率值依然有很大的影响, 必然影响以包似然函数作为目标函数的 boosting 迭代的精度. 因此, 本文采用文献[16]提出的示例重要性加权联合表征包的概率, 如式(6)、(7)所示:

$$p(y = 1 | X^+) = \sum_{j=1}^N w_j p(y = 1 | x_{1j}) \quad , \quad (6)$$

$$p(y = 0 | X^-) = w \sum_{j=N+1}^{N+L} (1 - p(y = 1 | x_{0j})) \quad . \quad (7)$$

其中,对于正包,  $w_{ij} \propto \exp\{-\|l_i(x_{ij}) - l_{i-1}^*\|\}$  是一个单调递减的权重函数,反映正包的  $N$  个样本中,距离目标位置越近的样本,作为正样本的概率就越大,对正包概率的贡献就应该越大,因此赋予高权重,反之亦然.这种定义方式减小了潜在误差样本的干扰,因为误差样本通常出现在离目标较远的位置,将被赋予较小的权重.由于负包中的  $L$  个样本都远离目标位置,因此赋予相同的权重  $w$ .将式(6)、(7)代入到包的对数似然函数表达式  $\ell = \sum (y_i \log p_i + (1 - y_i) \log(l - p_i))$  中,  $\ell$  可以改写成

$$\begin{aligned} \ell &= \sum_{s=0}^1 \left( y_s \log \left( \sum_{j=1}^N w_{ij} p(y=1 | x_{ij}) \right) + (1 - y_s) \log \left( \sum_{j=N+1}^{N+L} (1 - p(y=1 | x_{0j})) \right) \right) \\ &= \sum_{s=0}^1 \left( y_s \log \left( \sum_{j=1}^N w_{ij} \sigma(H(x_{ij})) \right) + (1 - y_s) \log \left( \sum_{j=N+1}^{N+L} (1 - \sigma(H(x_{0j}))) \right) \right) \end{aligned} \quad (9)$$

然后利用序贯前向选择的方法,从 2.1 节构建的弱分类器池  $\Phi = \{h_1, h_2, \dots, h_M\}$  中贪婪地挑选  $K$  个性能最好的弱分类器组成强分类器,如式(8)所示,

$$h_k = \arg \max_{h \in \Phi} \ell(H_{k-1} + h) \quad (8)$$

其中,  $H_{k-1} = \sum_{m=1}^{k-1} h_m$  由挑选出的  $k-1$  个弱分类器组成的强分类器.采用 AnyBoost 算法<sup>[10]</sup>建立效率更高的弱分类器挑选准则.首先利用一阶泰勒展开形式近似最大化的包对数似然函数,  $\ell(H_{k-1} + h) \approx \ell(H_{k-1}) + \langle h, \nabla \ell(H) \rangle |_{H=H_{k-1}}$ , 其中  $\langle h, \nabla \ell(H) \rangle |_{H=H_{k-1}} = \frac{1}{N+L} \sum_{j=1}^{N+L} h(x_{ij}) \nabla \ell(H)(x_{ij})$  是一个内积表达式,因此式(8)可以改写成

$$h_k = \arg \max_{h \in \Phi} \langle h, \nabla \ell(H) \rangle |_{H=H_{k-1}} \quad (10)$$

其中  $\nabla \ell(H)(x_{ij})$  为内积空间的泛函梯度,其表达式如下:

$$\begin{aligned} \nabla \ell(H)(x_{ij}) &= \left. \frac{\partial \ell(H + \delta 1_{x_{ij}})}{\partial \delta} \right|_{\delta=0} \\ &= y_i \frac{w_{ij} \sigma(H(x_{ij})) (1 - \sigma(H(x_{ij})))}{\sum_{m=1}^N w_{im} \sigma(H(x_{im}))} - (1 - y_i) \frac{\sigma(H(x_{ij})) (1 - \sigma(H(x_{ij})))}{\sum_{m=1}^L \sigma(H(x_{im}))} \end{aligned} \quad (11)$$

综上,基于 WMIL boosting 的在线弱分类器挑选方法的伪码如 Algorithm1 所示.

### Algorithm1: 基于 WMIL boosting 的在线弱分类器挑选

输入:训练集  $\{X^+, X^-\}$ , 正样本包  $X^+ = \{x_{ij}, y_j=1, j=1, 2, \dots, N\}$

负样本包  $X^- = \{x_{0j}, y_j=0, j=N+1, N+2, \dots, N+L\}$

1 基于训练集更新分类器池中的  $M$  个弱分类器

2 初始化  $H_0(x_{ij}) = 0 \quad i \in \{0, 1\}, j$

3 for  $k=1$  to  $K$  do

4 运用公式(11)计算  $\nabla \ell(H)(x_{ij}) | H=H_{k-1}$

5 设  $\ell_m = 0, m=1, \dots, M$

6 for  $m=1$  to  $M$  do

7 for  $i=0$  to 1 do

8 for  $j=1$  to  $N+L$  do

9  $\ell_m \leftarrow \ell_m + h_m(x_{ij}) \nabla \ell(H)(x_{ij}) | H=H_{k-1}$

10 end for

11 end for

12 end for

13  $m^* = \arg \max_m (\ell_m)$

14  $h_k(x_{ij}) \leftarrow h_{m^*}(x_{ij})$

15  $H_k(x_{ij}) \leftarrow H_{k-1}(x_{ij}) + h_k(x_{ij})$

16 end for

输出:分类器  $H_K(x_{ij}) = \sum_{k=1}^K h_k(x_{ij}); p(y=1 | \cdot) = \sigma(H_K(\cdot))$

### 2.3 分类器输出信任度的自适应融合

通过 Algorithm1,将基于红外图像 HOG 特征训练得到的子分类器定义为  $H_{\text{inf}}$ ,基于可见光图像 Haar-like 特征训练得到的子分类器定义为  $H_{\text{vis}}$ ;在当前帧,经  $H_{\text{inf}}$  和  $H_{\text{vis}}$  得到两个测试样本集的信任度映射,经加权融合后,得到联合强分类器的信任度值,融合机制如式(12)所示,

$$H = \max \left( \frac{H_{\text{inf}}}{H_{\text{vis}} + H_{\text{inf}}} \cdot H_{\text{inf}} + \frac{H_{\text{vis}}}{H_{\text{vis}} + H_{\text{inf}}} \cdot H_{\text{vis}} \right) \quad (12)$$

则  $H$  输出值对应的最大值就为当前帧目标所在的位置.采用上述融合输出机制使跟踪算法兼顾了自适应性和稳定性,当环境改变导致某一分类器的性能下降时,对应的权重降低,最终的跟踪结果就更依赖另一分类器<sup>[8]</sup>.

### 2.4 先验知识辅助下的协同训练准则

受在线 Semi-boosting<sup>[17]</sup>算法的启发,在第一帧手动确定目标位置后,在领域内提取  $M$  个正样本构成初始正样本包,即分类器的先验知识,以后每次样本包更新时,先从先验知识中提取  $m$  个样本作为新正样本包的第一部分,目的是避免跟踪结果的剧烈偏离;再从候选样本集中以协同训练的方式引入信任度最高的  $(M-m)$  个样本构成正样本包的第二部分,目的是提高分类器针对目标和背景变化的自适应性.  $m$  的取值对算法的鲁棒性影响很大,具体分析见第 4 节.本文依据当前帧的跟踪结果与第一帧手动确定的目标样本的差异性确定  $m$  的值,当差异较小时,选择较大的  $m$  值也不影响模型的自适应性;

当差异较大时,说明目标自身发生了改变,如果继续使用较多的先验知识训练,必然会引入误差.通过样本梯度方向直方图之间的巴氏距离  $d_B$  对这种差异性进行量化,下面以确定可见光图像的  $m_{\text{vis}}$  为例进行说明:

(1)第一帧手动选择的目标样本与第二帧所有测试样本的巴氏距离可以写成:  $d_B^i = \sqrt{1 - \sum_{b=1}^B \sqrt{p_b^i \cdot q_b}} (i = 1, 2, \dots)$ , 其中  $q_b$  和  $p_b^i$  分别代表第一帧目标样本和第二帧第  $i$  个测试样本的梯度方向直方图. 假设正样本包中固定的样本总数为  $M$ , 以  $k = (\max(d_B^i) - \min(d_B^i)) / J$  为间隔, 可以将  $d_B^i$  的取值范围划分为  $J$  个区间, 那么第  $j$  个区间可以写成  $(\min(d_B^i) + (j-1) \cdot k, \min(d_B^i) + j \cdot k)$ , 记作  $\Phi_j (j = 1, 2, \dots, J)$ ;

(2)从第三帧开始,跟踪结果的梯度方向直方图记作  $p_b$ , 计算得到它与第一帧手动选择的目标样本的巴氏距离为  $d_B$ , 然后按照表 1 所示的机制确定  $m_{\text{vis}}$ , 保证  $m_{\text{vis}}$  取值与  $d_B$  线性负相关.

表 1 计算  $m_{\text{vis}}$  值的准则

Condition	Criteria
$d_B \gg \max(d_B^i)$	stop updating
$d_B \geq \max(d_B^i)$	$m_{\text{vis}} = 0$
$d_B \leq \max(d_B^i)$	$m_{\text{vis}} = M$
$d_B \in \Phi_j$	$m_{\text{vis}} = \binom{J-j}{J} \cdot M$

求红外图像下的  $m_{\text{inf}}$  与上述方法一致. 由此, 基于本文提出的优化协同训练准则更新训练集正、负样本包的具体过程是: 首先, 确定分类器  $H_{\text{inf}}$  测试集中  $(M - m_{\text{vis}})$  个高信任度的图像块位置, 将可见光图像中对应位置的图像块提取出来, 作为新的正样本集替换  $H_{\text{vis}}$  正样本包的第二部分, 同理, 确定  $H_{\text{vis}}$  测试集中  $(M - m_{\text{inf}})$  个高信任度的图像块位置后, 将红外图像中对应位置的图像块提取出来, 作为新的正样本集替换  $H_{\text{inf}}$  正样本包的第二部分. 最后, 分别在红外和可见光图像相同的远离目标的环形区域内均匀、随机采样若干图像块作为负样本添加到分类器  $H_{\text{inf}}$  和  $H_{\text{vis}}$  各自的训练集中.

文中提出的自适应融合跟踪算法, 在红外图像和可见光图像配准的前提下, 分别从提取红外图像 HOG 特征构建的弱分类器池中和提取可见光图像 Haar-like 特征构建的弱分类器池中, 基于 WMIL boosting 技术挑选出判别能力好的弱分类器组成两个独立的强分类器, 在自适应先验知识引入机制辅

助下经过协同训练完成分类器样本包的在线更新, 伪码如 Algorithm2 所示.

#### Algorithm2: 基于优化协同训练理论的自适应融合跟踪器

输入: 配准的红外和可见光图像序列

1 初始化: 分别在第一帧红外和可见光图像手动确定目标所在位置, 然后在  $X_q^a = \{x \mid r > \|l_i(x) - l_i^*\|$  的领域内均匀提取若干图像块作为先验知识, 即初始正样本包; 在  $X_q^b = \{x \mid a < \|l_i(x) - l_i^*\| < b\}$  领域内均匀、随机提取若干图像块作为负样本包; 然后分别提取红外图像样本的 HOG 特征和可见光图像样本的 Haar-like 特征, 计算分类器参数, 然后分别构建基于红外和可见光图像的弱分类器池;

2 从第 2 帧开始, 在红外和可见光图像  $X_q^a = \{x \mid r > \|l_i(x) - l_{i-1}^*\|$  领域内提取测试样本, 其中  $r < R, l_{i-1}^*$  是上一帧目标所在位置, 然后提取红外图像测试样本的 HOG 特征, 可见光图像测试样本的 Haar-like 特征;

3 利用 Algorithm1 分别从基于红外和可见光图像的弱分类器池中挑选  $K$  个判别性最好的弱分类器组成强分类器  $H_{\text{inf}}$  和  $H_{\text{vis}}$ , 然后测试样本集得到基于两个分类器的信任度值, 运用公式(12)和公式(1)得到目标在当前帧的位置  $l_i^*$ ;

4 基于 3.4 节先验知识辅助下的协同训练准则更新训练集中的正、负样本包, 提取对应样本的特征, 依据公式(4)(5)更新所有弱分类器参数;

输出: 目标位置  $l_i^*$ , 更新的分类器

### 3 基于误差模型算法有效性分析

文中算法本质上属于协同训练框架下 Semi-boosting 分类器算法, 基于文献[9]、[18]的思想, 当满足以下假设条件时: (1) 足够多的训练样本; (2) 各特征能相对独立的赋予样本标签; (3) 基于不同特征的分类结果总体上是一致的, 该类算法的联合分类器第  $q$  次更新的误差可以写成:

$$\frac{1}{n} \left| \{i: \text{sign}(H(x_i)) \neq y_i\} \right| = \frac{1}{2n} \{err_{\text{label}} + err_{\text{unlabel}}\} \quad (13)$$

$$err_{\text{label}} = \left( \sum_{i=1}^m (\exp(y_i H_1(x_i)) + \exp(y_i H_2(x_i))) \right) \cdot f(p, q) \quad (14)$$

$$err_{\text{unlabel}} = \sum_{i=m+1}^n \left( \exp\left(\sum_{k=1}^K -\text{sign}(h_{1,k}(x_i)) \cdot h_{2,k}(x_i)\right) + \exp\left(\sum_{k=1}^K -\text{sign}(h_{2,k}(x_i)) \cdot h_{1,k}(x_i)\right) \right) \quad (15)$$

其中  $n$  为总的样本数,  $m$  为带标签的样本数,  $H_1$  和  $H_2$  分别代表基于第一类特征和第二类特征的强分类器,  $h_{1,k}$  和  $h_{2,k}$  分别代表第  $k$  次迭代后挑选出的基于第一类特征和第二类特征的弱分类器,  $f(p, q) \in [0, 1]$ , 其中  $p$  代表当前帧真实目标的外观与第一帧目标外观的差异程度, 损失函数  $f(\cdot)$  取值与变量  $p$  和  $q$  正相关.

由式(13) ~ (15)可知, 当  $m$  取值逐渐从  $0 \rightarrow n$



增加时,算法总的误差也从依赖  $err_{\text{unlabel}}$  逐渐转变为依赖  $err_{\text{label}}$ ,其中  $err_{\text{unlabel}}$  大小主要由两种不同分类器分类结果的不一致性决定的,当完全一致时,即  $h_{1,k}(x_i)$  和  $h_{2,k}(x_i)$  符号相同, $err_{\text{unlabel}}$  取最小值,

$$\min (err_{\text{unlabel}}) = \sum_{i=m+1}^n \left( \exp\left(\sum_{k=1}^K - |h_{2,k}(x_i)|\right) + \exp\left(\sum_{k=1}^K - |h_{1,k}(x_i)|\right) \right), (16)$$

当完全不一致时,即  $h_{1,k}(x_i)$  和  $h_{2,k}(x_i)$  符号不相同, $err_{\text{unlabel}}$  取最大值,

$$\max (err_{\text{unlabel}}) = \sum_{i=m+1}^n \left( \exp\left(\sum_{k=1}^K |h_{2,k}(x_i)|\right) + \exp\left(\sum_{k=1}^K |h_{1,k}(x_i)|\right) \right). (17)$$

WMIL 可以解决样本置信度出现误差时的分类器更新问题,这是因为它利用包概率的对数似然函数代替在线自适应 boosting (online adaptive boosting, OAB) 算法迭代过程中的样本误差损失函数作为目标函数,而且包概率的计算考虑到了样本重要性信息,从而保证了挑选出的  $h_k(\cdot)$  的准确性,使得  $h_{1,k}(x_i)$  和  $h_{2,k}(x_i)$  符号不一致的概率减小,下面以遮挡条件下分类器更新举例说明,如图 2 所示.可见,当跟踪结果出现偏差时,正样本包中依然有很大的概率包含“最正确”样本,缓解了分类器在线更新过程中的误差积累.  $err_{\text{label}}$  大小主要由自适应性不足导致的损失函数  $f(p, q)$  决定. 综上,当  $f(p, q)$  较小时,取较大的  $m$  值,可以保证总误差维持在较小的水平;当  $f(p, q)$  较大时,应当取较小的  $m$  值,然后通过减小  $err_{\text{unlabel}}$  来降低总的误差,可见  $m$  的取值对训练误差十分重要,具体讨论见 4 节.

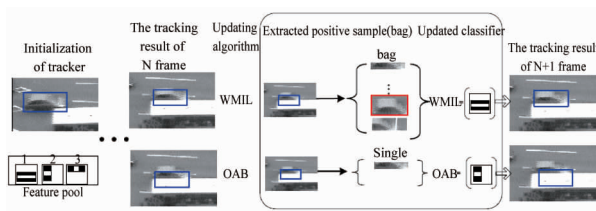


图 2 部分遮挡条件下 WMIL 提高分类器更新准确度的示意图

Fig. 2 An illustration of WMIL improving accuracy of classifier updating on the condition of partial occlusion

## 4 实验结果及分析

为了验证文中所提融合跟踪算法的有效性,选

用 OTCBVS 03/OSU Color and Thermal Database<sup>[19]</sup> 中 4 组人体目标序列和实际采集的 4 组车辆目标序列(场景示例如图 3 所示,测试图像序列的属性如表 2 所示)作为测试序列,比较算法的参数设置与原文献中的设置一致,文中所提的算法,正包的选择半径  $r=4$ ,经均匀采样共获取 45 个正样本;选择人体目标和车辆目标负包的环形区域内外半径分别为  $(8, 30)$  和  $(12, 34)$ ,经均匀随机采样 45 个样本作为负样本;测试样本的选择半径  $R=20$ ,共有 1 245 个图像块,学习率  $\lambda=0.85$ .



图 3 获取测试图像序列的 3 类场景

Fig. 3 Three scenes for acquiring testing image sequences

表 2 红外和可见光测试图像序列参数

Table 2 Reference of infrared and visible testing image sequences

Name	No.	Challenge	Name	No.	Challenge
S01	#533 ~ 699	SB, IC	AC1	#30 ~ 68	SB
S02	#365 ~ 550	IC, PO	AC2	#112 ~ 156	PO
ST1	#635 ~ 715	PO, SB	AC3	#136 ~ 190	PO, SB
ST2	#3 ~ 370	SB	AC4	#910 ~ 978	SB

Note: S0-Scene One; ST-Scene Two; AC-Actual Collection; PO-Partial Occlusion; SB-Similar Background; IC-Illumination Change

设计了两组实验,实验 1 中,分析了正样本包第一部分取自先验知识中的正样本数对算法鲁棒性的影响,然后将子分类器改为基于 OAB 算法构建,定义为 Fusion-CoOAB,将正样本包中第二部分样本集更新改为自学习更新,定义为 Fusion-Self + Semi,将上述两个跟踪器与本文跟踪器进行对比实验,验证算法各部分对整体鲁棒性的贡献;实验 2 中,将文献[16]提出的基于可见光图的 WMIL 跟踪器 (Vis-WMIL)、文献[20]提出的基于红外图像的粒子群优化后的粒子滤波跟踪器 (Inf-PSO + PF) 以及文献[5]提出的基于粒子滤波的多特征融合跟踪器 (Fusion-PF)、文献[6]提出的基于均值偏移的多特征融合跟踪器 (Fusion-MS) 与本文提出的跟踪器进行比较,同时采用定量和定性的分析方式,并对上述算法的实时性进行了说明.

### 4.1 实验 1: 主要工作对算法鲁棒性影响分析

为了分析协同训练更新时,提取的先验知识中

带正标签的样本数  $m$  对算法鲁棒性的影响,分别取  $m=0,15,30,45$ ,将上述固定  $m$  值条件下的跟踪算法与本文算法进行对比;基于所有测试序列得到的跟踪结果中心位置误差如表 3 所示,

表 3 不同标签样本数下的跟踪误差

Table 3 Tracking error based on different number of labeled samples

	$m=0$	$m=15$	$m=30$	$m=45$	Ours
SO1	29.1	3.6	31.5	32.9	2.2
SO2	11.5	8.2	4.5	13.5	1.9
ST1	4.2	1.8	6.8	16.5	2.2
ST2	12.6	12.3	8.5	132.5	4.0
AC1	22.8	20.1	2.5	6.8	3.8
AC2	35.2	36.1	8.6	16.8	10.3
AC3	80.7	9.4	16.5	110.8	5.4
AC4	4.4	3.9	5.5	4.0	3.2
average	25.0	11.9	10.6	41.7	4.1

$m=0$  代表完全去除先验信息, $m=45$  代表正样本集不更新,两种极端情况下取得了较差的跟踪结果;而基于本文方法实现了对  $m$  的自适应取值,对大多数序列取得了最优或次优的跟踪结果,验证了该方法的有效性.

为了验证算法各部分对整体鲁棒性的贡献,将 Fusion-CoOAB、Fusion-Self + Semi 和本文算法进行对比实验,得到三种场景下的部分跟踪结果如图 4 所示. SO2 序列中,当目标靠近灯柱时,无论是可见光图像还是红外图像,都出现了目标与局部背景的对比如度下降,Fusion-CoOAB、Fusion-Self + Semi 逐渐偏离了目标,到 134 帧时,只有本文算法依然能有效地跟踪目标;AC1 序列中目标受到路灯、交通牌的遮挡,Fusion-CoOAB 和 Fusion-Self + Semi 在第 32 帧附近,相继丢失了目标,遮挡结束后,只有本文算法能重新精确地跟踪目标;ST2 序列中,当目标经过垃圾桶时,由于可见光图像中目标与背景十分近似,Fu-

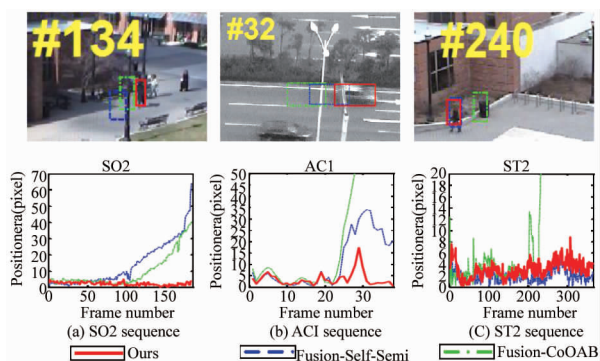


图 4 比较算法的部分跟踪结果

Fig. 4 partial tracking results of comparable algorithms

sion-CoOAB 丢失了目标. 在对应的中心位置误差 (center location error, CLE) 曲线图中,相对于比较算法,本文算法的 CLE 曲线也是最低和平滑的,验证了先验知识辅助下的协同训练和 WMIL 对提高算法鲁棒性的重要意义.

## 4.2 实验 2: 本文算法与主流算法性能对比

### 4.2.1 定量分析

四种主流算法与本文算法进行对比实验,得到的跟踪结果 CLE 如表 4 所示,对于绝大多数序列,本文算法都取得了最小的 CLE 均值和标准差,其余也是次优的结果,说明了本文算法不但跟踪的精度最高,且稳定性最好. 图 5 是 8 组测试序列的 CLE 曲线,相对于比较算法,本文算法的误差曲线在测试序列中都较低和平滑,所反映的结果与表 4 一致. 上述定量分析的结果说明本文算法有效地利用了两种图像互补信息,获得了比当前基于单源图像或其它融合策略的跟踪算法更好的鲁棒性.

表 4 跟踪结果的中心位置误差统计分析

Table 4 CLE statistical analysis of tracking results

Name	Ours		Vis-WMIL		Inf-PSO + PF		Fusion-PF		Fusion-MS	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
SO1	2.2	1.0	33.8	11.8	35.9	24.9	36.8	25.6	3.0	3.4
SO2	1.9	0.9	28.1	23.2	8.2	10.1	8.0	9.6	7.2	8.1
ST1	2.2	1.1	13.3	8.2	17.5	4.4	19.1	4.1	21.1	6.4
ST2	4.0	2.0	121.5	78.3	3.6	1.7	3.0	1.2	93.4	72.1
AC1	3.8	3.2	18.4	26.8	15.8	4.4	15.7	5.3	6.8	4.4
AC2	10.3	6.5	9.0	8.8	21.7	4.2	31.0	5.5	17.6	2.8
AC3	5.4	4.2	76.6	64.4	11.3	8.7	25.0	24.9	7.6	5.5
AC4	3.2	1.3	6.2	2.1	6.3	4.2	13.6	10.4	6.1	3.1

在 3.60 GHz CPU 的 PC 机上基于 MATLAB 2012,上述 5 种算法的运行速度如表 5 所示,由于要分别训练基于红外图像和可见光图像的分类器,使本文算法平均每秒完成的图像跟踪帧数只有 Vis-WMIL 和 Inf-PSO + PF 的一半左右,协同训练时的自适应先验知识引入机制和 WMIL 机制也增加了本文算法的复杂度,导致本文算法略落后于 Fusion-MS 和 Fusion-PF,但也达到了 8.4FPS 的跟踪速度,满足跟踪算法的实时性要求.

表 5 对比跟踪算法的运行速度

Table 5 Speeds of comparable tracking algorithms (FPS)

	Vis-WMIL	Inf-PSO + PF	Fusion-MS	Fusion-PF	Ours
Speeds	18.2	17.5	11.4	9.6	8.4

### 4.2.2 定性分析

为了更直观地反映本文算法在光照变化、遮挡、



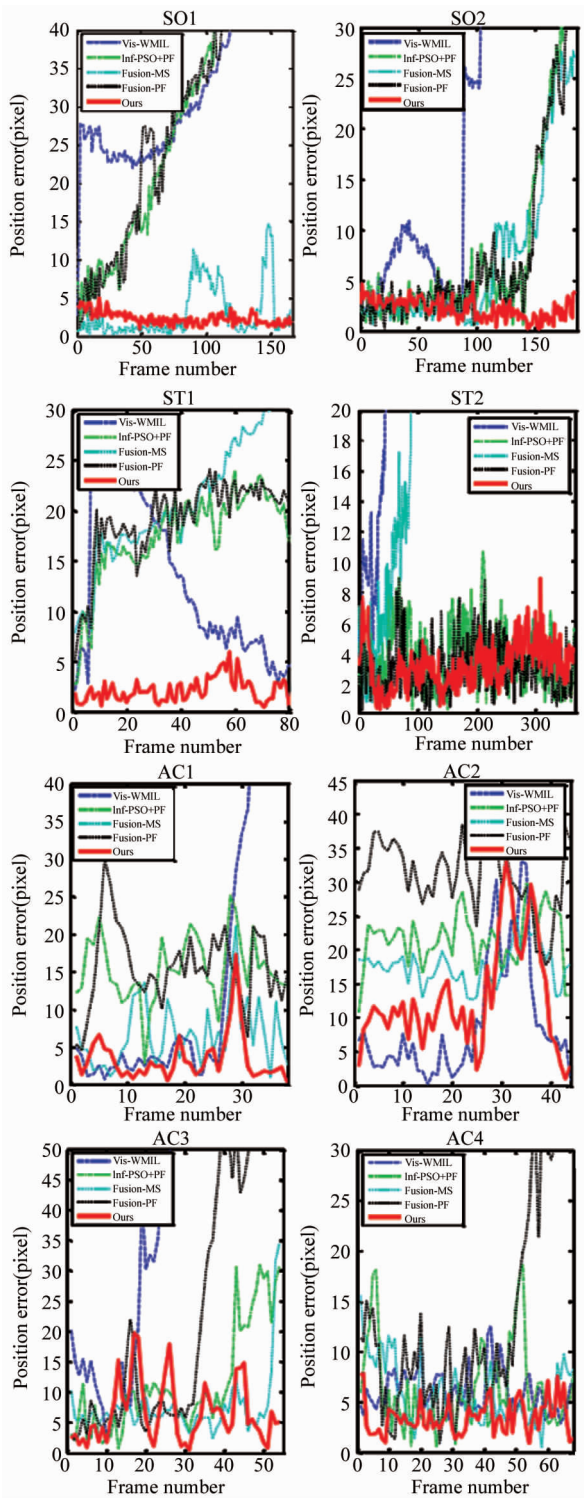


图5 比较算法对所有测试序列的跟踪误差曲线  
 Fig. 5 CLE curves of comparable algorithms for all testing sequences

近似背景等复杂环境下的鲁棒性,利用部分测试序列中关键帧的跟踪结果示例进行详细分析,如图6所示。

光照变化:序列 SO2、SO1 都经历了不同程度的

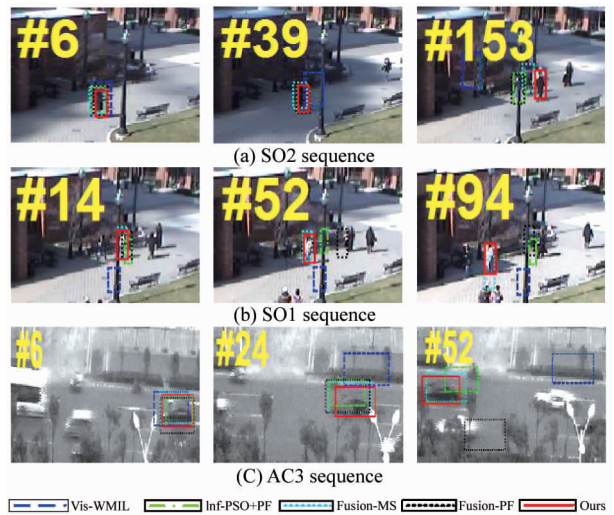


图6 比较算法的部分跟踪结果  
 Fig. 6 partial tracking results of comparable algorithms

光照变化,序列 SO2 的开始阶段,所有跟踪算法都能较好地跟踪目标,到了第 39 帧附近,由于附近建筑物阴影的影响,导致光照条件发生明显变化,Vis-WMIL 完全丢失了目标,Fusion-MS、Fusion-PF 也出现了偏离;序列 SO1 的 52 帧前,本文算法和 Fusion-MS 都能精度地跟踪目标,由于目标此时进入了阴影区域,光照发生改变,到第 94 帧时,Fusion-MS 已出现较严重的误差.本文算法通过协同训练进行信息交互时,由于红外图像受光照变化的影响较小,能校正可见光图像特征受光照变化出现的误差,因此融合跟踪的结果保持了较好的鲁棒性。

遮挡:3 组序列都出现了遮挡的干扰,序列 SO2 和序列 SO1,当目标经过灯柱时,被部分遮挡,且此时在红外图像中目标与灯柱的亮度值十分接近,导致单独基于红外图像的 Inf-PSO + PF 和融合跟踪算法 Fusion-PF 都丢掉了目标,如序列 SO2 第 153 帧、序列 SO1 第 52 帧所示.序列 AC3 到第 24 帧时,由于路灯对目标的部分遮挡,导致除本文算法外的其它跟踪算法都出现了不同程度的偏离.在本文提出的先验知识使用机制的帮助下,本文算法能在遮挡时阻止使用被严重遮挡的样本训练,避免了分类器判别能力的退化;遮挡结束后,在先验知识的帮助下能重新跟踪目标。

近似背景:序列 SO1 第 153 帧附近,背景中出现了其他迎面走来的行人目标,作为干扰目标,它们与真实目标在红外图像中拥有近似的亮度特征或是在可见光图像中拥有近似的纹理特征;导致原本精度较高的 Fusion-MS 最后也丢失了目标;序列 AC3



第 52 帧附近,搜索区域中出现了摩托车、汽车等近似干扰目标,导致 Fusion-PF 和 Inf-PSO + PF 分别错误地偏向汽车和摩托车,Fusion-MS 也出现较大的误差.当出现近似背景时,所有的跟踪算法都会出现一定的误差,受益于 WMIL 使得在定位不准确时,最正确的样本依然有很大的概率出现在正样本包中,因此本文算法受到相对最小的近似背景干扰.

## 5 结论

基于优化协同训练理论,实现了可见光和红外图像信息的自适应融合跟踪;通过基于 WMIL boosting 技术和自适应先验知识引入机制辅助下的协同训练准则,降低了两类分类器相互提供误差正样本训练所导致的性能下降风险,实验结果验证了在复杂环境下,本文提出的融合跟踪算法相对于单源或其它融合机制的跟踪算法更好的鲁棒性,且很容易推广到其它类型传感器的信息融合.下一步的研究重点包括:(1)研究复杂环境下判别性、稳定性好的图像特征,进一步满足协同训练对特征充分性的要求;(2)进一步优化协同训练方式,降低引入误差样本的风险,更好地实现分类器稳定性和自适应性的平衡;(3)研究通过并行计算和代码的优化处理提高跟踪器的实时性.

## References

- [1] Yuan J, Chen H, Sun F, *et al.* Multisensor information fusion for people tracking with a mobile robot: A particle filtering approach[J]. *IEEE transactions on Instrumentation and Measurement*, 2015, **4**(63):2427-2442.
- [2] Kumar P, Mittal A., Kumar P. Study of robust and intelligent surveillance in visible and multi-modal framework[J]. *Informatica*, 2007, **31**(4):63-77.
- [3] Kumar P, Mittal A, Kumar P. Fusion of thermal infrared and visible spectrum video for robust surveillance[C]. In CVGIP, 2006.
- [4] Wu Y, Blasch E, Chen G S, *et al.* Multiple source data fusion via sparse representation for robust visual tracking[C]. In FUSION 2011.
- [5] Tahha M, Stolk R. Particle filter tracking of camouflaged targets by adaptive fusion of thermal visible spectra camera data[J]. *Sensors Journal, IEEE*, 2014, **14**(1):159-166.
- [6] Stolk R, Rees D, Talha M, *et al.* Bayesian fusion of thermal and visible spectra camera data mean shift tracking with rapid background adaptation[C]. In *Sensors, IEEE*, 2012.
- [7] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training[C]. In *Proceedings of ACM 11th Annual Conference on Computational Learning Theory*, 1998:92-100.
- [8] Lu H, Zhou Q, Wang D, *et al.* A co-training framework for visual tracking with multiple instance learning [C]. In *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, 2011:539-544.
- [9] Liu R, Cheng J, Lu H. A robust boosting tracker with minimum error bound in a co-training framework[C]. In *CVPR*, 2009:1459-1466.
- [10] Zhu J, Ma Y, Qin Q, *et al.* Adaptive weighted real-time compressive tracking[J]. *IET Computer Vision*, 2014, **8**(6):740-752.
- [11] Diaconis P, Freedman D. Asymptotics of graphical projection pursuit[J]. *The Annals of Statistics*, 1984: **12**(3):793-815.
- [12] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]. In *CVPR*, 2009.
- [13] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]. In *CVPR*, 2001.
- [14] Zhang S, Yu X, Sui Y, *et al.* Object tracking with multi-view support vector machines[J]. *IEEE Transactions on Multimedia*, 2015, **17**(3):265-278.
- [15] Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting[J] *BMVC*. 2006, **1**(5):6.
- [16] Zhang K, Song, H. Real-time visual tracking via online weighted multiple instance learning[J]. *Pattern Recognition*, 2013, **46**(1):397-411.
- [17] Grabner H, Leistner C, Bischof H. Semi-supervised online boosting for robust tracking[C]. In *ECCV*, 2008.
- [18] Yu S, Krishnapuram B, Rosales R, *et al.* Bayesian co-training[J]. *The Journal of Machine Learning Research*, 2011, **12**:2649-2680.
- [19] OTCBVS Benchmark Dataset Collection. <http://www.cse.ohiostate.edu/otcbvs-bench/>.
- [20] Zhang M, Xin M, Yang J. Adaptive multi-cue based particle swarm optimization guided particle filter tracking in infrared videos[J]. *Neurocomputing*, 2013, **122**:163-171.