

基于多模态背景模型和霍夫森林的 红外目标跟踪

印剑飞¹, 杨尧², 张巍巍¹, 杨开²

(1. 上海航天控制技术研究所, 上海 201109; 2. 西北工业大学 航天学院, 西安 710072)

摘要: 针对红外目标跟踪过程中目标和背景耦合性强和目标形变带来的表现特征变化的问题, 提出了基于多模态背景模型和霍夫森林的红外目标跟踪算法, 将目标和背景的特征信息整体融入跟踪框架, 并引入霍夫森林对目标部件表征信息进行训练更新; 通过特征描述和模式列表建立多模态背景模型, 在观测序列中记录匹配的背景模式列表, 根据匹配结果进行权重和模态参数更新; 以背景建模获得的区域信息为采样基础, 将前景区域和背景区域分别作为正负样本集代入决策树进行训练; 根据叶节点存储的图像块信息对目标位置进行投票决策, 从图像特征空间映射到霍夫参数空间, 生成目标区域概率分布图, 获取高置信度区作为目标跟踪区。在红外公开测试集上实验结果表明, 在背景杂波和目标形变对跟踪造成干扰的情况下, 所提算法仍能保持跟踪的稳定性。

关键词: 多模态; 背景建模; 霍夫森林; 红外目标跟踪

Infrared Target Tracking Based on Multimodal Background Model and Hough Forest

Yin Jianfei¹, Yang Yao², Zhang Weiwei¹, Yang Kai²

(1. Shanghai Institute of Spaceflight Control Technology, Shanghai 201109, China;

2. College of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Aiming at the problems of strong coupling between the target and the background and the change of the apparent features caused by the target deformation in infrared target tracking, an algorithm based on multi-modal background model and Hough forest was proposed. Both the target and background feature information were integrated into the tracking framework, and Hough forest was utilized for training and updating the target component characterization information. The multi-modal background model was built upon the feature descriptions and the pattern lists. Along with the matching background pattern lists recording in the observed image sequence, the modal parameters were updated according to the matching results. After obtaining the regional information from the background modeling, the foreground and the background were respectively taken as the positive and negative sample sets for the training in the decision trees. The decision of target's position was carried out through Hough voting operation according to the image blocks' information stored by the leaf nodes. The images' feature space was mapped to the Hough parameter space to generate the probability distribution map of the target, and the high confidence area was determined as the tracking position. Experimental results on the public infrared datasets show that the proposed algorithm can keep track the target stably under the circumstance that background clutter and target deformation interfere with the tracking process.

Keywords: multimodal; background modeling; Hough forest; infrared target tracking

0 引言

目标跟踪是计算机视觉研究领域的一个重要分支, 广泛应用于视频监控、行为和事件分析和精确制导等方面, 跟踪过程中光照变化、外观形变和背景遮挡等因素对其造成了极大的影响。近年来, 基于判别式模型的算法在目标跟踪领域获得了广泛的应用^[1], 通过训练分类器将跟踪问题转化为目标和背景的分类过程, 以初始跟踪矩形框内的目标样本作为正样本, 矩形框外的区域作为负样本, 通过特征提取和分类求解预测目标在下一帧的位置。由于目标形状的多变性, 以矩形框作为跟踪区域不可避免的会引入大量的背景信息, 使模型更新的过程中误

差逐步积累, 造成跟踪框的漂移。相较于可见光目标跟踪而言, 红外目标分辨率和信噪比较低^[2], 更易受背景杂波影响, 因而迫切需要一种对目标模型更为精确的表征。部件级的目标表现模型在应对形变和部分遮挡带来的目标外观变化时仍能保持一定的鲁棒性, Felzenszwalb^[2]等提出的可形变部件模型(Deformable Part Model, DPM)在物体检测领域取得了巨大的成功, 但DPM模型需要待检测物体类别的数据集进行离线训练, 较难应用于在线通用目标的跟踪。Kwon and Lee^[4]等提出的基于目标部件拓扑结构的跟踪模型在目标发生大量形变时取得了较好的效果, 但引入的蒙特卡洛采样使算法的实时性较差。Gall J^[5]等对图像进行分解, 构建霍夫森林将图像块映射成对目标重心位置的投票实现目标检测。Godec M^[6]等将霍夫森林进一步扩展至目标跟踪领域, 将grab-cut粗分割和目标区域采样进行结合, 避免背景区域引入的杂波信息对目标样本造成的干扰。由Gall J^[5]和Godec M^[6]等提出的跟踪算法在目标和背景区分度比较明显的场景中取得了较好的跟踪效果, 由

收稿日期: 2018-02-24; 修回日期: 2018-03-20。

基金项目: 上海航天科技创新基金(SAST2016073)。

作者简介: 印剑飞(1979-), 男, 江苏丹阳人, 高级工程师, 主要从事控制总体技术方向的研究。

于红外场景中目标边缘模糊, 纹理信息缺乏, 与背景区域形成了一定的耦合性, 直接将上述算法应用于红外目标跟踪时, 难以对目标进行精确定位。

本文在上述算法的基础上, 引入背景建模的思想, 将目标和背景的特征信息整体融入跟踪框架, 对霍夫森林跟踪算法进行改进。基于局部二值模式 (Local Binary Pattern) 描述符来表示背景纹理, 同时引入光度不变性灰度特征, 应对整体光照的变化, 使得背景模型对局部或整体亮度的扰动均有一定的鲁棒性。背景建模提供区域信息作为精确采样的基础, 前景区域和背景区域分别作为正负样本集代入霍夫森林进行样本训练, 获取高置信度的样本区作为目标跟踪区域, 保证了模型更新的准确性和跟踪的稳定性。

1 特征描述

1.1 LBP 局部纹理特征

LBP (local binary patterns, 局部二值模式) 是度量图像局部纹理特征的算子, 在计算机视觉的诸多领域得到了广泛的应用^[7], 它的计算如下式所示, 其中 (x_c, y_c) 为中心像素的坐标, i_p 为第 p 个邻域像素的灰度值, i_c 为中心像素的灰度值, $s(x)$ 为符号函数。

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} 2^p s(i_p - i_c) \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

以图 1 为例, 在中心像素 65 的 3×3 邻域内, 相邻的 8 个像素分别与其进行比较, 若大于中心像素则相应位置值为 1, 反之则为 0, 最后形成的 8 位二进制序列即为中心像素的 LBP 值。

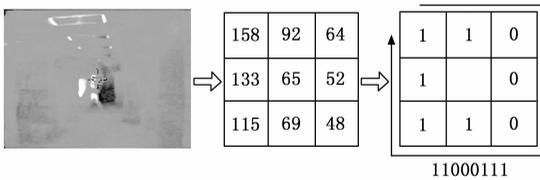


图 1 LBP 局部纹理特征

1.2 光度不变性灰度特征

Kyungnam Kim^[8]等通过改变光照条件, 观测由于亮度改变而发生的像素值变化, 主要表现为观测像素值与颜色模型中原点间连线方向的变化, 故提出将像素值在颜色空间中的夹角作为前景和背景的判断依据,

$$\tilde{I}^t = I^t + I_n, \|I_n\| \leq n_c \quad (3)$$

$$\theta_n = \arcsin(n_c / \|I^t\|) \quad (4)$$

$$\hat{I}_l^t = \min(\mu I^t, \tilde{I}^t), \mu \in [0.4, 0.7] \quad (5)$$

$$\hat{I}_h^t = \max(v I^t, \tilde{I}^t), v \in [1.1, 1.2] \quad (6)$$

设图像序列为 \tilde{I}^t , \tilde{I}^t 为当前时刻观测到的 I^t 值, I_n 为噪声信号, n_c 为噪声上限, θ_n 代表观测向量 \tilde{I}^t 与 I^t 之间夹角的最大值, \hat{I}_h^t 和 \hat{I}_l^t 分别代表灰度变化范围的上限和下限, u 和 v 为亮度系数, 通过 $[\hat{I}_l^t, \hat{I}_h^t]$ 定义了因光照变化而产生的像素值变化区间, 即亮度的允许波动范围。

1.3 特征距离测度

通过 LBP 局部纹理特征和光度不变性灰度特征对图像进

行描述后, 进一步地需要对图像不同区域间的相似性进行度量, 定义 LBP 局部纹理特征如式 (7) 所示, 其中 $D_{\text{ext}}(LBP_a, LBP_b)$ 代表区域 a 和区域 b 之间的局部纹理特征距离。

$$D_{\text{ext}}(LBP_a, LBP_b) = \frac{1}{P} \sum_{p=0}^{P-1} D_{011}(LBP_a^{(p)}, LBP_b^{(p)}) \quad (7)$$

$$D_{011}(x, y) = \begin{cases} 1, & |x - y| \leq T_D \\ 0, & \text{其它} \end{cases} \quad (8)$$

定义光度特征相似度距离如式 (9) 所示, $D_g(I_k^{-1}(x), I^t(x))$ 为像素 x 在时刻 t 与上一帧 k 模式 $I_k^{-1}(x)$ 的距离, 从背景模式列表中寻找与当前像素距离最小的模式, 背景模式列表的建立过程在文章的第二部分进行阐述。从光度特征距离的定义中可以看出当像素值处于因光照变化而产生的像素值变化区间时, $I^t(x)$ 和 $I_k^{-1}(x)$ 的之间夹角 θ 越小, 相应的距离越小, 若像素值不属于亮度的允许波动范围时, 则赋予其较大的距离值。

$$D_g(I_k^{-1}(x), I^t(x)) = \begin{cases} 1 - e^{-\frac{|\theta - \theta_n|}{\delta}}, & I^t \in [\hat{I}_l^t, \hat{I}_h^t] \\ 1, & I^t \notin [\hat{I}_l^t, \hat{I}_h^t] \end{cases} \quad (9)$$

基于纹理特征距离和光度特征距离建立特征距离测度如式 (10) 所示, 分别计算像素 x 在时刻 t 与上一帧 k 模式的纹理特征距离和光度特征距离, 其中 λ 为距离融合系数。

$$\text{Dist}(m_k^{-1}) = \lambda D_{\text{ext}}(LBP_k^{-1}(x), LBP^t(x)) + (1 - \lambda) D_g(I_k^{-1}(x), I^t(x)) \quad (10)$$

2 多模态背景建模

通过背景建模提取区域信息可以为后续目标区域定位提供良好的基础, Jian Yao^[9]等提出多层背景建模的思想, 通过建立模式列表和模式更新对背景模型进行建模, 结合上文所述的特征描述可对各个模态建立表征分量, 设图像序列为 I^t , 背景模型为 M^t , 设像素 x 处的背景模型为 $M^t(x)$, 令 $M^t(x) = \{\{m_k^t(x)\}_{k=1, \dots, K^t(x)}, K^t(x), B^t(x)\}$, $m_k^t(x)$ 代表当前图像观测序列下的背景模式列表。通过控制 $K^t(x)$ 的取值范围, 可以对背景模式列表的计算复杂度和模式列表的覆盖度进行均衡。在背景模式列表中每个模式都有对应的权重系数, 求解过程中通过对模式权重系数的降序排列实现对各模式在当前观测序列中的重要性排序, 进而根据重要性排序结果选取前 $B^t(x)$ 个模式代表当前背景模型, 通过 $B^t(x)$ 进一步对模式列表的复杂度和匹配精度进行控制。其中每个背景模式 m_k 有 6 个分量, $m_k = \{\bar{I}_k, \hat{I}_k^l, \hat{I}_k^h, LBP_k, \omega_k, \hat{\omega}_k\}$, $\bar{I}_k, \hat{I}_k^l, \hat{I}_k^h$ 分别代表当前模式下像素的平均值, 最小值和最大值, LBP_k 为 LBP 特征值, ω_k 为模式权重系数, $\hat{\omega}_k$ 则记录了该模式在历史图像序列中的最大权重系数。对于匹配的模式, 模态权重参数按照下式进行更新。

$$\omega_k^t = (1 - \alpha_w^i) \omega_k^{t-1} + \alpha_w^i \quad (11)$$

$$\alpha_w^i = \alpha_w (1 + \tau \omega_k^{t-1}) \quad (12)$$

$$\hat{\omega}_k^t = \max(\hat{\omega}_k^{t-1}, \omega_k^t) \quad (13)$$

对于非匹配模式, 按照式 (14) 进行更新, 其中 α_w^d 和 α_w^i 分别对应于权重增长系数和权重衰减系数, α_w 和 τ 为常数因子, 可以看到对应模式的历史最大权重 $\hat{\omega}_k^{t-1}$ 越大, 增长系数增加越快, 衰减系数减少越缓慢。当匹配模式在观测历史中有

较高的权重时, 即会进一步增加其可信度, 但如果有突然的扰动造成该模式不可靠时, 该权重不会立即下降, 而是根据历史权重制定衰减系数, 权重越大, 衰减越缓慢。当扰动消失时, 权重可逐步恢复, 这与实际观测结果也相一致。图 2 为经过多模态背景建模后提取的背景区域和前景区域, 从图中可以看出前景区域较好地保留了目标的形状信息, 将背景建模提取的区域信息作为后续霍夫森林的区域采样的输入, 通过细化采样提高霍夫森林训练的准确度, 进而提高目标跟踪精度。

$$\omega_k^d = (1 - \alpha_w^d) \omega_k^{d-1} \quad (14)$$

$$\alpha_w^d = \frac{\alpha_w}{1 + \tau \omega_k^{d-1}} \quad (15)$$



图 2 基于多模态背景建模提取的区域

3 霍夫森林

霍夫森林是随机森林和霍夫投票算法的结合^[5], 在随机森林的学习框架的基础上融合目标的位置信息通过广义霍夫变换进行投票, 构建决策树对图像块进行训练^[10]。

3.1 随机森林

设训练样本集 $S = \{P_i = \{I_i, c_i, d_i\}\}$, 其中 P_i 为图像块, I_i 为提取的图像块特征, c_i 为类别标签, d_i 为目标位置偏移信息, 通过二值检测对节点进行分割, 如式 (16) 所示, $I^f(p, q)$ 和 $I^f(u, v)$ 代表测试的图像块 I 在位置 (p, q) 和位置 (u, v) 的特征值, τ 为特征阈值, 生成二值测试集合 $\{t^k\}$ 。基于类别不确定度和位置偏差最小化原则对决策树进行训练, 其中类别不确定度通过信息熵进行衡量, 如式 (17) 所示, 其中 c 为 $\{c_i\}$ 的平均值, $|S|$ 为样本集的个数。位置偏差的计算如式 (18) 所示, \bar{d} 为偏移向量均值, 只对类别信息为目标时进行计算, 训练时随机选择一个最小化原则中对目标函数进行优化, 如式 (20) 所示, 使得最终构建完成的决策树类别和位置的偏差达到最小。

$$t_{f,p,q,u,v,\tau}(I) = \begin{cases} 0, & I^f(p, q) < I^f(u, v) + \tau \\ 1, & \text{其它} \end{cases} \quad (16)$$

$$U(S) = |S| \cdot \text{Entropy}\{c_i\} = |S| \cdot (-c \log c - (1-c) \log(1-c)) \quad (17)$$

$$D(S) = \sum_{i:c_i=1} (d_i - \bar{d})^2 \quad (18)$$

$$\text{Obj}(t^k) = \{U(S), D(S)\} \quad (19)$$

$$\text{argmin}_k (\text{Obj}(t^k | I_i = 0) + \text{Obj}(t^k | I_i = 1)) \quad (20)$$

3.2 霍夫投票

设中心位于 y 的图像块 $P(y) = (I(y), c(y), d(y))$, 由图像块的特征 $I(y)$ 可得目标中心位于位置 x 的概率为 $p(E(x) | I(y))$, 当图像块类别标签 $c(y) \neq 1$ 时, $p(E(x) | I(y)) = p(E(x))$, 即图像块的特征不影响目标中心位于位置 x 的随机

事件 $E(x)$, 当 $c(y) = 1$ 时计算如式 (21) 所示。设 L 为决策树 $\{T_i\}_{i=1}^L$ 的叶结点, 叶结点记录了训练时的偏移量 D_L 和样本类别比例 C_L , 则 $p(d(y) = y - x | c(y) = 1, I(y))$ 可通过高斯 Parzen 窗进行估计, $p(c(y) = 1 | I(y))$ 可直接由 C_L 估计, 如式 (22) 所示, 对于整个霍夫森林, 对所有决策树的概率进行平均, 如式 (23) 所示, 对不同图像块对目标中心进行概率投票的值进行累加, 可得霍夫投票映射到图像空间 $V(x)$ 的值, 如式 (24) 所示, $V(x)$ 越高代表当前区域获得的投票数越多, 即当前区域包含目标的概率越高, 通过高概率区域范围进一步确定目标位置。

$$\begin{aligned} p(E(x) | I(y)) &= p(E(x), c(y) = 1 | I(y)) = \\ &= p(E(x) | c(y) = 1, I(y)) \cdot p(c(y) = 1 | I(y)) = \\ &= p(d(y) = y - x | c(y) = 1, I(y)) \cdot p(c(y) = 1 | I(y)) \end{aligned} \quad (21)$$

$$\begin{aligned} p(E(x) | I(y); T) &= \\ &= \left[\frac{1}{|D_L|} \sum_{d \in D_L} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|(y-x) - d\|^2}{2\sigma^2}\right) \right] \cdot C_L \end{aligned} \quad (22)$$

$$p(E(x) | I(y); \{T_i\}_{i=1}^T) = \frac{1}{T} \sum_{i=1}^T p(E(x) | I(y); T_i) \quad (23)$$

$$V(x) = \sum_{y \in B(x)} p(E(x) | I(y); \{T_i\}_{i=1}^T) \quad (24)$$

4 红外目标跟踪

基于多模态背景模型和霍夫森林的红外目标跟踪的算法整体流程如图 3 所示, 首先基于建立的特征描述子提取红外图像特征, 然后通过计算各区域序列的特征距离和背景模式列表确定匹配的背景模式, 进而由背景建模提取前景区域和背景区域。以背景模型提供的区域信息为基础进行区域采样, 在前景区和背景区分别采集图像块作为正负样本集送入随机森林进行训练, 训练过程中根据图像块的类别信息和位置信息对决策树进行构建, 最终训练完成的决策树叶结点记录了霍夫投票所需的信息, 通过统计不同图像块对目标中心进行概率投票的值得到目标置信度图, 置信度图反映了目标处于不同位置的概率, 最后根据置信度图完成目标区域定位。

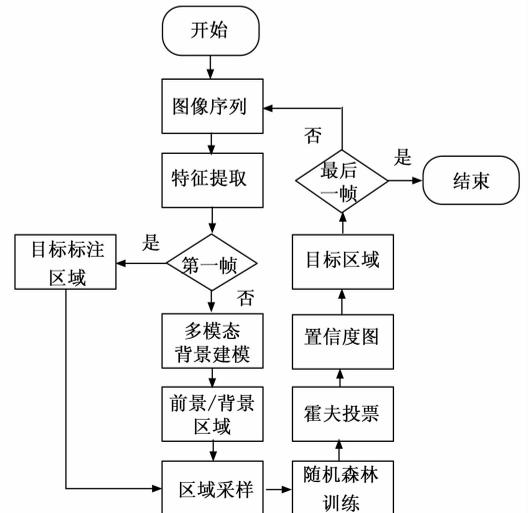


图 3 红外目标跟踪算法流程图

5 实验结果分析

实验硬件系统为 Intel (R) Core (TM) i3-4030U 1.90 GHz 的 CPU, 4 GB 内存, 操作系统为 Windows 7, 运行软件为 Visual Studio 2010。实验采用 OTCBVS (Object Tracking and Classification in and Beyond the Visible Spectrum) 红外公开数据集中的行人序列进行测试分析, 图像分辨率为 320 * 240。通过对跟踪结果的定性分析和定量分析, 对本文提出的算法与 Godec M^[6] 提出的基于霍夫森林的跟踪算法进行比较。

选取跟踪序列中部分帧的跟踪结果对算法的性能进行分析, 图 4 为第 16 帧的目标区域置信度分布对比, 左图为原始的基于霍夫森林的跟踪算法置信度分布, 图中的高亮区即为算法的高置信度区。从左图中可以看出, 经过霍夫投票后的目标置信度图受背景杂波影响, 高置信度区分布较分散, 部分背景区域也覆盖于高置信度区域内。图 4 的右图为本文提出的融入背景建模后置信度区的分布图, 对比第 16 帧的目标实际位置 (见图 5 中的第 16 帧跟踪结果), 可以看出经过背景建模后, 置信度分布图排除了大量背景杂波的干扰, 高置信度区域得到了有效的收缩, 将关注区更集中于目标区域。

图 5 分别为第 16、33 和 82 帧的跟踪结果, 左侧为原始算法的跟踪结果, 右侧为本文提出改进算法后的跟踪结果。从左侧的跟踪结果中可以看出当目标靠近背景中的高亮干扰源时, 由于背景杂波的灰度分布特征与目标本身的灰度分布存在一定的相似性, 原始算法易将其误认为目标区, 并错误地将其引入了后续的模式更新过程中, 造成后续跟踪过程的漂移。右侧的跟踪结果中, 经过背景建模后霍夫投票的置信度分布图排除了背景中的干扰源, 较好地保留了目标的区域信息, 将高概率区集中于目标本身, 跟踪区域能够更有效地锁定目标。

定义跟踪误差为每帧目标中心标注位置 (x_{gt}, y_{gt}) 与跟踪的目标位置 (x_t, y_t) 之间的欧氏距离, 如式 (25) 所示, 跟踪误差曲线如图 6 所示, 其中加 * 号曲线为本文提出的算法。从原始算法的误差跟踪曲线中可以看出原始算法的跟踪误差波动范围较大, 进一步地对误差曲线进行分析, 可以看到原始误差曲线有两处较明显的波峰, 是因为目标向背景中的高亮区域靠近, 原始算法将背景中的这部分干扰区域和目标同时锁定 (见图 5 左侧的跟踪结果), 造成跟踪误差曲线的波动。相较于原始算法, 融合背景建模后误差曲线 (图 6 中加 * 号曲线) 有了明显下降, 原先的误差曲线两处较大的波动也得到了有效的抑制。



图 4 目标区域置信度图

$$e_r = \sqrt{(x_{gt} - x_t)^2 + (y_{gt} - y_t)^2} \quad (25)$$

6 结论

本文针对红外目标跟踪过程中背景杂波和目标形变对目标

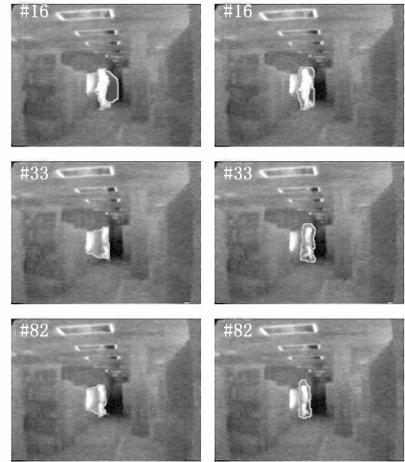


图 5 部分帧跟踪结果

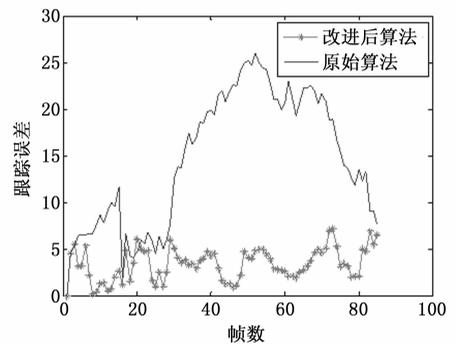


图 6 跟踪误差曲线

跟踪引起的干扰, 提出了一种基于多模态背景模型和霍夫森林的红外目标跟踪算法, 通过特征描述和多模态背景建模提取背景区域和前景区域作为霍夫森林的训练输入, 通过区域分块采样和在线样本更新, 获得目标信息的鲁棒表征。由霍夫投票输出对目标中心的投票概率图, 生成目标区域分布置信度图, 最后根据置信度图完成目标区域定位。相较于基于矩形框对目标进行跟踪的算法, 本文提出的算法提高了对目标区域定位的准确性, 通过背景建模和目标区域分解较好地保留了目标的结构和形状信息, 避免了模型更新时过多地引入背景杂波对跟踪过程造成的影响。实验结果表明在背景杂波和目标形变对目标跟踪造成一定程度干扰的情况下, 所提算法仍能保持跟踪的稳定性。

参考文献:

[1] Wang N, Shi J, Yeung D Y, et al. Understanding and diagnosing visual tracking systems [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. 2015: 3101-3109.
 [2] 杨福才, 杨德东, 毛宁, 等. 基于稀疏编码直方图的稳健红外目标跟踪 [J]. 光学学报, 2017, 37 (11): 1115002.
 [3] 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.

(下转第 233 页)