

基于稀疏编码直方图的稳健红外目标跟踪

杨福才, 杨德东, 毛 宁, 李雪晴

河北工业大学控制科学与工程学院, 天津 300130

摘要 充分利用红外图像信息建立有效的观测模型是实现稳健红外目标跟踪的基础。影响红外目标跟踪结果的因素除可见光目标跟踪也会面临的干扰因素之外, 还有诸如边缘和纹理信息缺失、信噪比低和背景噪声影响等特有因素。提出基于稀疏编码直方图(HSC)特征和扰动感知模型(DAM)的红外目标跟踪方法, 使用 K-奇异值分解算法得到过完备字典, 利用该字典计算得到每个像素点的稀疏编码, 并组成 HSC 对目标进行表达, 同时通过引入 DAM 增强算法抗背景干扰能力。该方法充分利用了红外图像中运动目标的结构特性, 能够有效去除背景干扰。与其他跟踪器相比, 在 VOT-TIR2015 数据集上, 该方法的精确度和成功率指标分别获得 3.8% 和 4.4% 的提升, 具有较高的研究价值和实用价值。

关键词 机器视觉; 红外图像; 目标跟踪; 稀疏编码直方图; 扰动感知模型

中图分类号 TP391.4 **文献标识码** A

doi: 10.3788/AOS201737.1115002

Robust Infrared Target Tracking Based on Histograms of Sparse Coding

Yang Fucui, Yang Dedong, Mao Ning, Li Xueqing

School of Control Science and Engineering, Hebei University of Technology, Tianjin 300130, China

Abstract Making use of information in infrared images to build an effective observation model is the basis for realizing robust infrared target tracking. Besides the regular factors that have adverse influence on visual target tracking, infrared target tracking is faced with other difficulties as well, such as lack of edge and texture information, low signal-to-noise ratio and background clutter. An infrared target tracking algorithm based on histograms of sparse coding (HSC) and the distractor-aware model (DAM) is proposed, which exploits K singular value decomposition algorithm to obtain an overcomplete dictionary. With the dictionary, sparse code of every pixel is computed to compose HSC as a descriptor, and DAM is utilized to strengthen resistance against background clutter. The proposed algorithm does not only use structural information of tracked object but also eliminates the influence of background clutter. Compared with other tracking algorithms, the proposed algorithm achieves 3.8% and 4.4% enhancement on VOT-TIR2015 dataset with respect to precision and success rate, respectively, possessing high research and practical value.

Key words machine vision; infrared image; target tracking; histogram of sparse coding; distractor-aware model

OCIS codes 150.0155; 150.1135

1 引 言

红外目标跟踪广泛应用于军事及生产生活中, 具有重要的价值。红外目标跟踪与可见光目标跟踪相比, 除了需要解决目标形变、发生遮挡、运动模糊等问题之外, 还需要解决更加独特的问题, 比如信噪比低、纹理信息缺乏、背景复杂混乱^[1]。这些特有的影响因素使红外目标跟踪更具有挑战性。

1.1 相关算法

目前国内外学者已提出很多红外目标跟踪算法。Huang 等^[2]提出使用 GrabCut 方法对红外目标进行

收稿日期: 2017-03-27; 收到修改稿日期: 2017-05-10

基金项目: 国家自然科学基金(61203076)、天津市自然科学基金(13JCQNJC03500)、河北省自然科学基金(F2017202009)

作者简介: 杨福才(1990—), 男, 硕士研究生, 主要从事机器学习、目标跟踪方面的研究。E-mail: 18622198313@163.com

导师简介: 杨德东(1977—), 男, 博士, 副教授, 硕士生导师, 主要从事智能感知与控制、目标检测与跟踪等方面的研究。

E-mail: ydd12677@163.com(通信联系人)

精确切割,之后使用支持向量机进行在线更新。Demir 等^[3]提出使用区域共差方法对目标进行跟踪。张辉等^[4]提出使用灰度特征、局部标准差特征和梯度特征以直方图的形式描述目标外观,在粒子滤波框架下估计目标运动状态。管志强等^[5]提出由均值漂移算法快速找到局部最优解,然后利用边缘不变矩作为特征修正均值平移算法产生的误差。孙瑾等^[6]利用基于梯度方向的局部二值模式特征为基础的粒子滤波跟踪算法,建立局部和全局的特征直方图描述实现对手部目标的跟踪。赵爱罡等^[7]提出以边缘曲线的区域质心为基准点建立局部坐标系,在此坐标系下边缘点的坐标值具有仿射不变性,将边缘的坐标集合作为描述子,利用平均 Hausdorff 距离度量边缘的相似性。很多用于红外目标跟踪的算法是由可见光目标跟踪算法发展而来的,如 l1 最小化模型跟踪器^[8]、颜色命名(CN)方法^[9]、时空上下文学习器(STC)^[10]、结构化输出跟踪(Struck)^[11]、多实例学习(MIL)^[12]、边缘盒跟踪(EBT)^[13]和核相关滤波器(KCF)^[14]算法。Struck 算法使用核结构化输出支持向量机对目标进行预测,这种支持向量机的输出并不是通常的±1,而是输出对于目标变化的预测。EBT 方法提出使用边缘检测方法确定疑似目标边缘盒,使用在线更新的支持向量机确定目标框的可能位置。除此之外,相关滤波方法将目标跟踪问题视作连续时间域内的检测问题,通过快速傅里叶变换将时域内的卷积计算转换为频域内的哈达玛乘积运算,近期取得了良好的结果。

1.2 相关滤波器及稀疏编码直方图

基于相关滤波的跟踪算法源自平方误差和最小化输出(MOSSE)算法^[15],MOSSE 算法对初始帧目标赋予二维高斯输出,在频域进行训练,并且采用创新的更新策略。Henriques 等^[14]引入核方法对 MOSSE 算法进行改进,提出 KCF,KCF 使用循环移位的方法生成训练样本,并与核方法结合。多特征尺度自适应跟踪器(SAMF)^[16]、判别式尺度空间跟踪器(DSST)算法^[17]采样不同尺度的图像,并使用双线性插值方法缩放成初始目标大小并对目标尺度进行估计。空间正则化判别式相关滤波器(SRDCF)算法^[18]引入空间正则化方法,依据位置信息惩罚滤波器参数,在不破坏正样本的条件下使相关滤波器通过大量负样本进行学习。

方向梯度直方图(HOG)特征仅利用梯度信息,缺少对其他信息进行描述的能力^[19]。相关研究证明稀疏编码直方图(HSC)特征在 PASCAL 数据集和 INRIA 数据集上获得的检测结果均优于 HOG 特征^[20]。使用基于 K-奇异值分解(K-SVD)^[21]算法得到的字典对每个像素进行稀疏编码,然后将这些编码按照一定规则集中形成 HSC,用来对目标进行表达。

受上述跟踪算法以及模板与像素级学习器之和(STAPLE)算法^[22]的启发,本文提出使用基于 HSC 和扰动感知模型(DAM)^[23]的红外目标跟踪方法。对于细节缺失的红外图像,提取具有更强表达能力的特征尤其重要,而 HSC 特征的表达能力比 HOG 特征更为有效,而且需要确定的参数更少。使用 DAM 减轻背景干扰的影响,可以增强跟踪器对目标形变的稳健性。两者的结合可使算法能够对红外目标进行准确、稳健的跟踪。

2 基本方法

针对红外目标跟踪存在的问题,提出使用稀疏编码的方法对红外目标进行表达,利用基于灰度特征的 DAM 剔除背景噪声影响,使用特定得分函数评价二者得分并将其综合,以此确定目标在后续视频图像帧中的位置和尺度。

2.1 跟踪方法

采用检测跟踪方法,在视频序列的第 t 帧,可通过最大化得分函数确定目标在图像 I_t 中的中心位置 p_t :

$$p_t = \arg \max_{p \in S_t} f[T(I_t, p); \theta_{t-1}], \quad (1)$$

式中函数 T 表示对图像进行特征提取操作, $f[T(I_t, p); \theta]$ 根据模型参数 θ 对图像 I 中的矩形窗 p 计算得分。模型参数 θ 可以通过最小化损失函数 $L(\theta; \mathcal{J}_t)$ 得到,该损失函数根据 t 时刻之前的图像以及目标位置 $\mathcal{J}_t = \{(I_i, p_i)\}_{i=1}^t$ 得到,通过最小化损失函数可以确定模型参数 θ :

$$\theta_t = \arg \max_{\theta \in \mathcal{Q}} \{L(\theta; \mathcal{J}_t) + \lambda R(\theta)\}, \quad (2)$$

式中模型参数空间由 \mathcal{Q} 表示,使用正则项 $R(\theta)$ 以及正则化系数 λ 限制模型的复杂度并防止过拟合发生。第一帧中目标的状态 p_1 是给定的。采用文献[22]提出的得分函数,由模板得分 $f_{\text{tmpl}}(I)$ 和直方图得分 $f_{\text{hist}}(I)$ 线性组合而成:

$$f(I) = \gamma_{\text{tmpl}} f_{\text{tmpl}}(I) + \gamma_{\text{hist}} f_{\text{hist}}(I), \quad (3)$$

式中 γ_{templ} 和 γ_{hist} 分别表示模板得分和直方图得分的融合系数。模板得分根据基于 HSC 的相关滤波算法求得,直方图得分由 DAM 得出,将得分最大的位置视为目标的中心位置。具体流程如图 1 所示。

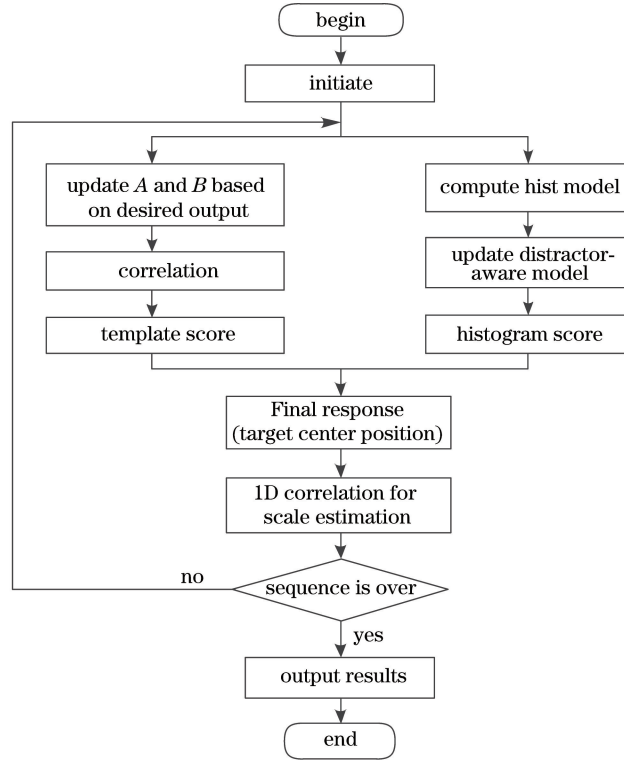


图 1 整体算法流程图

Fig. 1 Flow chart of overall algorithm

2.2 模板得分

模板得分由相关滤波算法获得。在基于最小二乘的相关滤波算法中,损失函数定义为

$$l_{\text{templ}}(\mathbf{I}, \mathbf{p}, \mathbf{h}) = \left\| \sum_{q=1}^Q \mathbf{h}^q \odot \boldsymbol{\phi}^q - \mathbf{y} \right\|^2 + \lambda \sum_{q=1}^Q \|\mathbf{h}^q\|^2, \quad (4)$$

式中 \mathbf{h}^q 表示相关滤波器 h 中第 q 维的参数, \odot 表示卷积运算, $\boldsymbol{\phi}^q$ 表示多通道特征图 $\boldsymbol{\phi}$ 中的第 q 维(共 Q 维), $\boldsymbol{\phi}$ 是 $\phi_{T(\mathbf{I}, \mathbf{p})}$ 的缩写, \mathbf{y} 表示关于训练样本 $\boldsymbol{\phi}$ 的相关滤波器的期望输出, 参数 $\lambda > 0$ 对正则化项进行控制。这是一个岭回归问题。

使用 $\phi_{T(\mathbf{I}, \mathbf{p})}$ 表示目标的 HSC 特征, 通过最小化 l_{templ} 可以得到相关滤波器的参数 \mathbf{h} 。直接进行卷积计算会消耗过多的计算资源和时间, 为了追求解决问题的快速方案, 在频域中进行相关性计算。首先, 将输入图像的特征矩阵和滤波器进行快速傅里叶变换: $\boldsymbol{\Phi} = \mathcal{F}(\boldsymbol{\phi})$, $\mathbf{H} = \mathcal{F}(h)$ 。由卷积定理可知, 相关性计算在傅里叶域变为哈达玛乘积运算, 减少了计算量和计算所需时间, 算法的时间复杂度为 $O(Q\mathcal{P}\log\mathcal{P})$, 其中 \mathcal{P} 为跟踪窗口的像素个数。根据求导结果为零可得 l_{templ} 的最小化结果为

$$\mathbf{H}^i = \bar{\mathbf{y}}^* \boldsymbol{\Phi}^i / \left(\sum_{q=1}^Q \boldsymbol{\Phi}^{q*} \boldsymbol{\Phi}^q + \lambda \right), \quad (5)$$

式中 $\bar{\mathbf{y}}$ 表示 \mathbf{y} 的傅里叶变换, 即 $\bar{\mathbf{y}} = \mathcal{F}(\mathbf{y})$, $\bar{\mathbf{y}}^*$ 表示 $\bar{\mathbf{y}}$ 的复共轭矩阵, $\boldsymbol{\Phi}^*$ 是 $\boldsymbol{\Phi}$ 的复共轭矩阵。令(5)式分子表示为 $\mathbf{A}_i^i = \bar{\mathbf{y}}^* \boldsymbol{\Phi}^i$, 分母表示为 $\mathbf{B}_i = \sum_{q=1}^Q \boldsymbol{\Phi}^{q*} \boldsymbol{\Phi}^q + \lambda$ 。

使用线性插值的更新策略可得

$$\mathbf{A}_i^i = (1 - \eta)\mathbf{A}_{i-1}^i + \eta\bar{\mathbf{y}}_i^* \boldsymbol{\Phi}_i^i, \quad (6)$$

$$\mathbf{B}_i = (1 - \eta)\mathbf{B}_{i-1} + \eta \sum_{q=1}^Q \boldsymbol{\Phi}_i^{q*} \boldsymbol{\Phi}_i^q, \quad (7)$$

式中 η 为学习率参数。相关滤波器在特征图矩形区域的得分 y' 的计算公式为

$$y' = \mathcal{F}^{-1} \left(\frac{\sum_{q=1}^Q \mathbf{A}^{q*} \Phi^q}{\mathbf{B} + \lambda} \right), \quad (8)$$

式中 y' 就是模板得分 f_{templ} , 即

$$f_{\text{templ}}(\mathbf{I}) = y'. \quad (9)$$

2.3 稀疏编码直方图

2.3.1 字典学习

字典经由 K-SVD 算法获得。对于给定的信号 $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$, K-SVD 算法通过解决

$$\min_{\mathbf{D}, \mathbf{C}} \|\mathbf{Y} - \mathbf{DC}\|_F^2 \quad \text{s. t. } \forall i, \|\mathbf{c}_i\|_0 \leq \epsilon \quad (10)$$

表示的最优化问题来寻找维度为 K (即包含 K 个单词) 的字典 $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^K$ 以及对应的稀疏编码矩阵 $\mathbf{C} = \{\mathbf{c}_i\}_{i=1}^N$ 。(10) 式中 $\|\cdot\|_F$ 为 Frobenius 范数, $\|\cdot\|_0$ 为零范数, 用来表示稀疏列向量 \mathbf{c}_i 中非零项个数的最大值, ϵ 为设定的稀疏度。K-SVD 算法通过交替计算 \mathbf{C} 和 \mathbf{D} 解决该最优化问题。针对给定字典 \mathbf{D} , 采用贪婪算法中的正交匹配跟踪 (OMP) 算法可以高效计算 \mathbf{C} ; 针对确定的 \mathbf{C} , 通过迭代使用奇异值分解对字典 \mathbf{D} 进行更新, 当算法收敛时, 得到的 \mathbf{D} 即为所需字典。

利用本文算法对字典进行训练时, 从 BSDS500 数据集的图像中 ($1 \leq i \leq N, i \in \mathbf{Z}$) 随机采样 N 个大小为 $p_{\text{patchsize}} \times p_{\text{patchsize}}$ 的小图像, 并将其按照先行后列的方式展开为列向量 \mathbf{y}_i 组成 \mathbf{Y} , 同时设置字典的维度以及稀疏度用来控制字典的表达能。不同图像块尺寸和维度的字典可视化图像如图 2 所示。随着图像块尺寸和字典尺度的增加, 字典可以表达更加复杂的模式。使用 K-SVD 算法在 \mathbf{Y} 上进行非监督学习就可以得到字典 \mathbf{D} 。

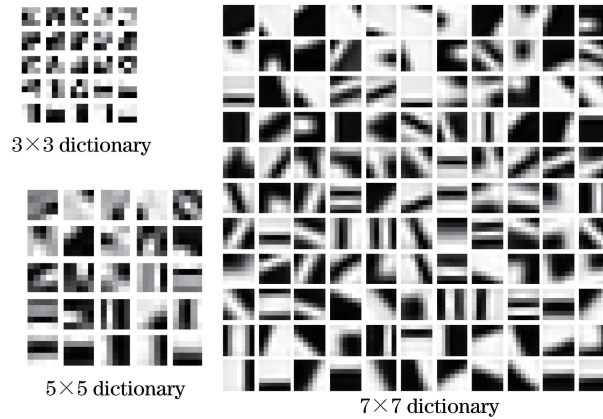


图 2 由 K-SVD 算法获得的三种不同图像块尺寸的字典

Fig. 2 Dictionaries learned through K-SVD algorithm for three patch sizes

2.3.2 HSC 特征提取

得到字典 \mathbf{D} 之后, 再次使用 OMP 算法即可计算 HSC, 具体步骤如下。将跟踪目标区域划分为若干个一定尺度的单元块, 对于某一单元块中的像素 e_i , 取其 $p_{\text{patchsize}} \times p_{\text{patchsize}}$ 大小的邻域空间, 将其按照先行后列的方式展开为 \mathcal{Y}_i , 根据预先训练完毕的字典 \mathbf{D} 求取稀疏编码 $\mathbf{C}_i \in \mathbf{R}^{K \times 1}$ 作为像素 e_i 的稀疏编码:

$$\min_{\mathbf{C}_i} \|\mathcal{Y}_i - \mathbf{D}\mathbf{C}_i\|_F^2 \quad \text{s. t. } \|\mathbf{C}_i\|_0 \leq \epsilon, \quad (11)$$

依次求取单元块中所有像素的稀疏编码组成该单元块的稀疏编码 \mathcal{C} , 对 \mathcal{C} 中的所有元素取其绝对值并取均值得到该单元块的 HSC $\mathcal{C}_{\text{cell}}$:

$$\mathcal{C}_{\text{cell}} = \frac{\sum_{i=1}^{p_{\text{patchsize}} \times p_{\text{patchsize}}} |\mathcal{C}_i|}{p_{\text{patchsize}} \times p_{\text{patchsize}}}, \quad (12)$$

之后对 $\mathcal{C}_{\text{cell}}$ 中的非零项 $\mathcal{C}_{\text{cell}}^i$ 进行正则化计算, 得到该单元块最终的 HSC, 即

$$\mathcal{C}_{\text{cell}} = \frac{\mathcal{C}_{\text{cell}}^i}{\sqrt{\sum_i \mathcal{C}_{\text{cell}}^i{}^2}}. \quad (13)$$

HSC 生成步骤如图 3 所示。

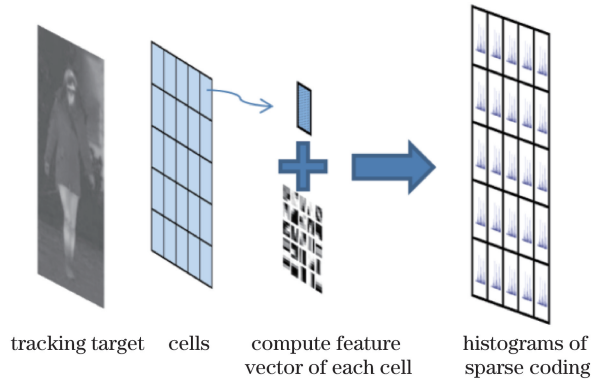


图 3 HSC 的计算过程

Fig. 3 Process of computing HSC

2.4 直方图得分

理想状态下,直方图得分需要从 t 时刻之前的所有图像中学习得到。出于稳定性和计算效率方面的考虑,使用当前帧图像的直方图信息,并且按照一定的规则进行更新。为了将属于目标的像素 $x \in \mathcal{O}$ 与背景区域 B 区分开,采用基于灰度直方图的贝叶斯滤波器对目标的灰度信息进行建模。使用 $H_{\Omega}^I(b)$ 表示图像 I 中区域 Ω 的第 b 个灰度区间的直方图,使用 b_x 表示像素 x 的亮度值 $V(x)$ 处于灰度区间 b 。对于给定的矩形目标区域 O 及其背景区域 B ,由贝叶斯原理可得像素 x 属于目标物体的似然概率为

$$P(x \in \mathcal{O} | O, B, b_x) \approx \frac{P(b_x | x \in O)P(x \in O)}{\sum_{\Omega \in \{O, B\}} P(b_x | x \in \Omega)P(x \in \Omega)} \quad (14)$$

使用灰度直方图表示似然概率,即 $P(b_x | x \in O) \approx H_O^I(b_x) / |O|$, 并且 $P(b_x | x \in B) \approx H_B^I(b_x) / |B|$, 其中 $|\cdot|$ 表示像素个数,先验概率可以使用 $P(x \in \Omega) \approx |O| / (|O| + |B|)$ 近似。则(14)式可以简化为

$$P(x \in \mathcal{O} | O, B, b_x) = \begin{cases} \frac{H_O^I(b_x)}{H_O^I(b_x) + H_B^I(b_x)}, & \text{if } V(x) \in O \cup B \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

该模型即为目标-背景模型,可将目标物体与背景区分,然而在周围区域 S 中仍然存在干扰因子,为提高跟踪的准确性,需要减少背景干扰信息。对于与目标区域 O 颜色相似的周围环境 S ,使用贝叶斯方法判断像素 x 属于目标还是干扰,因此可以定义目标-干扰模型:

$$P(x \in \mathcal{O} | O, S, b_x) = \begin{cases} \frac{H_O^I(b_x)}{H_O^I(b_x) + H_S^I(b_x)}, & \text{if } I(x) \in I(O \cup S) \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

将目标-背景模型与目标-干扰模型结合得到 DAM:

$$P(x \in \mathcal{O} | b_x) = \lambda_p P(x \in \mathcal{O} | O, S, b_x) + (1 - \lambda_p) P(x \in \mathcal{O} | O, B, b_x), \quad (17)$$

式中 λ_p 为预设权重参数。该模型能够有效减少背景噪声的干扰,非常适用于红外目标跟踪。为进一步适应目标外观变化及光照变化,使用如下更新策略对模型进行更新:

$$P_{1:t}(x \in \mathcal{O} | b_x) = \eta_p P(x \in \mathcal{O} | b_x) + (1 - \eta_p) P_{1:t-1}(x \in \mathcal{O} | b_x), \quad (18)$$

式中 η_p 为学习率。

得到 DAM 之后,在视频序列第 t 帧图像中以第 $t-1$ 帧图像目标中心位置为中心,选定搜索框 $\mathcal{W} \in I$, 且 $\mathcal{W} \subset \mathbf{Z}^2$, 对于其中的每个像素 $x(i, j) \in \mathcal{W}$, 计算其属于目标物体 \mathcal{O} 的概率:

$$P_t(x \in \mathcal{O}) = P_{1:t}(x \in \mathcal{O} | b_x), \quad (19)$$

式中 x 为 $x(i, j)$ 的简写。本文算法不使用所得概率作为直方图在该像素点的得分,而是继续计算 $P_t(x \in \mathcal{O})$ 的积分图 $\sum_{x(i, j) \in \mathcal{W}} P_t[x(i, j)]$, 并定义直方图得分为

$$f_{\text{hist}}(I) = \frac{1}{|\mathcal{W}|} \sum_{x(i, j) \in \mathcal{W}} P_t[x(i, j)]. \quad (20)$$

2.5 确定最大响应位置

分别求出模板得分 $f_{\text{templ}}(I)$ 和直方图得分 $f_{\text{hist}}(I)$ 后,取 $\gamma_{\text{hist}} = \alpha, \gamma_{\text{templ}} = 1 - \alpha$, 则最终得分为

$$f(I) = (1 - \alpha)f_{\text{templ}}(I) + \alpha f_{\text{hist}}(I). \quad (21)$$

根据最终得分情况,对于搜索区域 $R_s \in \mathbf{Z}^2$,取得分最大处的位置为目标中心位置 p_t ,即

$$p_t = \arg \max_{p \in R_s} f(I). \quad (22)$$

2.6 确定目标尺度

采用尺度金字塔方法对目标尺度进行估计。得到目标在当前第 t 帧的中心位置 p_t 之后,通过训练另一个独立的相关滤波器对目标尺度 z_t 进行估计。使用 $W \times H$ 表示第 $t-1$ 帧的目标尺度,使用 s 表示尺度滤波器的大小。对于每一个 $n \in \left\{ \left[-\frac{s-1}{2}, \left[-\frac{s-2}{2}, \dots, \left[\frac{s-2}{2}, \left[\frac{s-1}{2} \right] \right] \right] \right\}$, 提取中心位置为 p_t 、尺度为 $\beta^n W \times \beta^n H$ 的一组图像簇 J_n , 其中 β 为尺度因子。对于图像簇 J_n 中的每一幅图像 j_n , 将其统一缩放成尺度为 $W \times H$ 的图像,提取其 HSC 特征,根据(6)式和(7)式得出的 A_{t-1} 和 B_{t-1} 计算相关运算的结果 Y_n , 取 $k = \arg \max_{n \in \left\{ \left[-\frac{s-1}{2}, \left[-\frac{s-2}{2}, \dots, \left[\frac{s-2}{2}, \left[\frac{s-1}{2} \right] \right] \right] \right\}} Y_n$, 则可以得到当前帧的目标尺度为

$$z_t = \beta^k W \times \beta^k H. \quad (23)$$

3 实验仿真

3.1 数据集与对比算法

使用 VOT-TIR2015 红外目标跟踪数据集^[24] 对该算法进行验证,该数据集共有 20 个视频序列,从不同的场景中采集得到,每个视频序列的平均长度为 563 帧。跟踪目标包括行人、车辆、动物、无人机等,全部视频序列均提供跟踪目标真实值以及跟踪目标的属性,这些属性或者问题都是极为常见和难以处理的,包括光照变化、平面外旋转、尺度变化、遮挡、形变、运动模糊、快速运动、平面内旋转、出视野、背景杂乱、低分辨率^[25]。

为了证明该算法的有效性和稳健性,比较本文算法与另外 22 种红外目标跟踪算法,即 RPT^[24], TGPR^[26], PDMFD^[24], SAMF_AT^[16], SAMF^[16], MEEM^[27], KCF^[14], DSST^[17], ACT^[24], DST^[24], CCT^[24], KCFDP^[14], KCF_MTSA^[14], HRP^[24], IVT^[24], ASLA^[24], SCM^[24], L1APG^[24], DFT^[24], MTT^[24], LOT^[24], CSK^[24]。

3.2 参数设置

使用 BSDS500 数据集上预先训练完成的字典计算跟踪目标的 HSC 特征,将跟踪目标划分成大小为 8×8 的单元块,字典的大小选择为 100,设置稀疏度 $\epsilon = 2, p_{\text{patchsize}} = 7$ 。直方图得分权重设置为 $\gamma_{\text{hist}} = \alpha = 0.35$ 。计算模板得分时,将灰度范围为 $[0, 255]$ 的像素投影到具有 32 个灰度区间的直方图中,设置正则化系数 $\lambda = 10^{-3}$,在更新环节中学习率设置为 $\eta = 0.015$ 。计算 DAM 时,设置物体目标-背景模型和目标-干扰模型的融合系数 $\lambda_p = 1.3$,在更新环节中设置学习率 $\eta_p = 0.04$ 。在尺度估计步骤中取尺度滤波器大小 $s = 33$,取尺度因子 $\beta = 1.02$ 。

3.3 衡量标准

使用精确度和成功率衡量算法结果。定义中心位置误差为由算法得到的目标框中心 o_t 与跟踪目标真实值目标框中心 o_{GT} 之间的欧氏距离,即 $D_t = \sqrt{|o_t - o_{\text{GT}}|}$ 。定义某一算法在一个视频序列上的精确度为小于位置误差阈值的视频帧数占总帧数的比例。定义重叠度为由算法得到的目标框 r_t 与跟踪目标真实值目标框 r_{GT} 之间交集与并集的比值,即 $S_t = |r_t \cap r_{\text{GT}}| / |r_t \cup r_{\text{GT}}|$, 其中 $|\cdot|$ 表示区域内像素的个数。定义成功率为大于重叠度阈值的图像帧数占总帧数的比例^[24]。

3.4 定量分析

采用文献^[25]提出的 OPE(One-Pass Evaluation)方法在所有视频序列上进行一次测试。取中心位置误差阈值为 20 来衡量精确度,取重叠度阈值为 0.5 来衡量成功率。为显示清晰,图 4 和图 5 中给出排名前十的算法结果。就精确度而言,本文算法结果比排名第二的算法高 3.8%,在成功率方面本文算法结果比排名第二的算法高 4.4%,两项指标均有明显提升,体现了本文算法的有效性。

各种算法所有属性的精确度如图 6 所示,本文算法在处理形变、遮挡、尺度变化、平面外旋转、低分辨率、背景杂乱、运动模糊等复杂情况时均表现出良好的效果,在发生快速运动和平面内旋转时,以微小差异排名

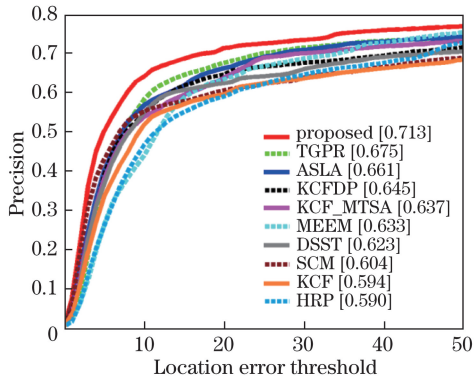


图 4 20 个视频序列上得到的不同算法的精确度
Fig. 4 Precision plot of different algorithms tested on 20 sequences

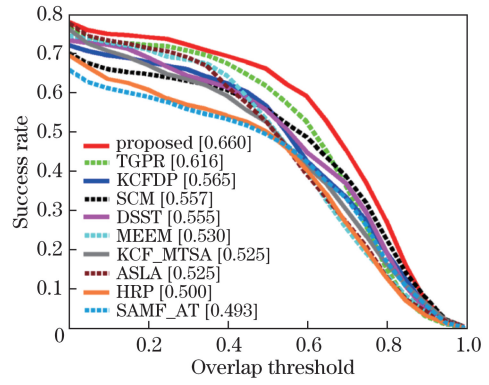


图 5 20 个视频序列上得到的不同算法的成功率
Fig. 5 Success rate plot of different algorithms tested on 20 sequences

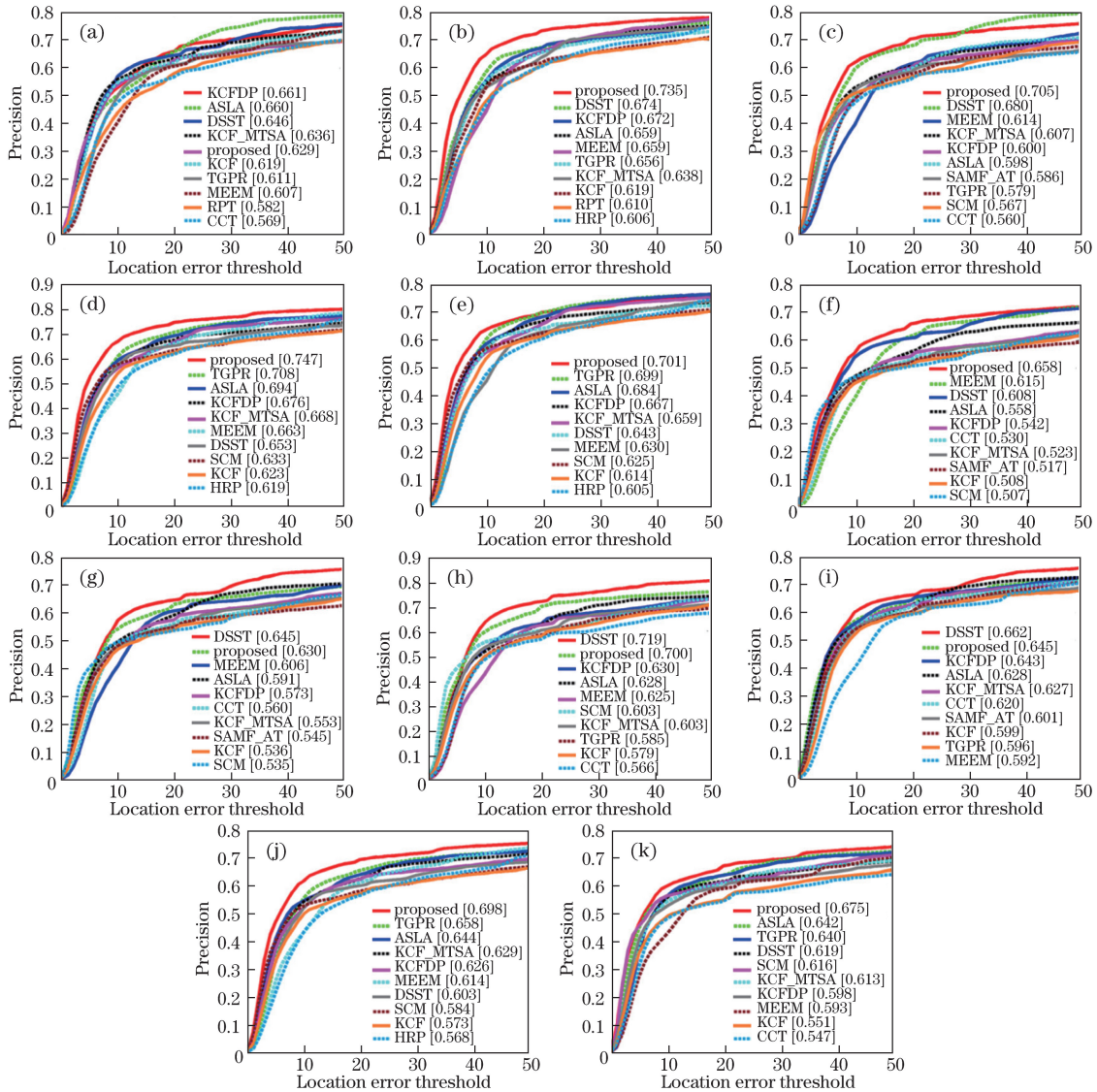


图 6 所有属性对应的精确度图。(a)光照变化;(b)平面外旋转;(c)尺度变化;(d)形变;(e)遮挡;(f)运动模糊;
(g)快速运动;(h)平面内旋转;(i)出视野;(j)背景杂乱;(k)低分辨率
Fig. 6 Precision plots of all attributions. (a) Illumination variation; (b) out-of-plane rotation; (c) scale variation;
(d) deformation; (e) occlusion; (f) motion blur; (g) fast motion; (h) in-plane rotation; (i) out of view;
(j) background clutter; (k) low resolution

第 2。成功率结果如图 7 所示。对于光照变化和出视野之外的属性,比排名第 2 的算法均有显著提高。在光照变化的情况下,本文方法在两种评价标准中的排名都不高,原因在于引入灰度信息的同时不可避免地增加了算法对光照变化的敏感度,但是在其他属性上得到的结果证明,引入灰度信息的好处大于弊端。

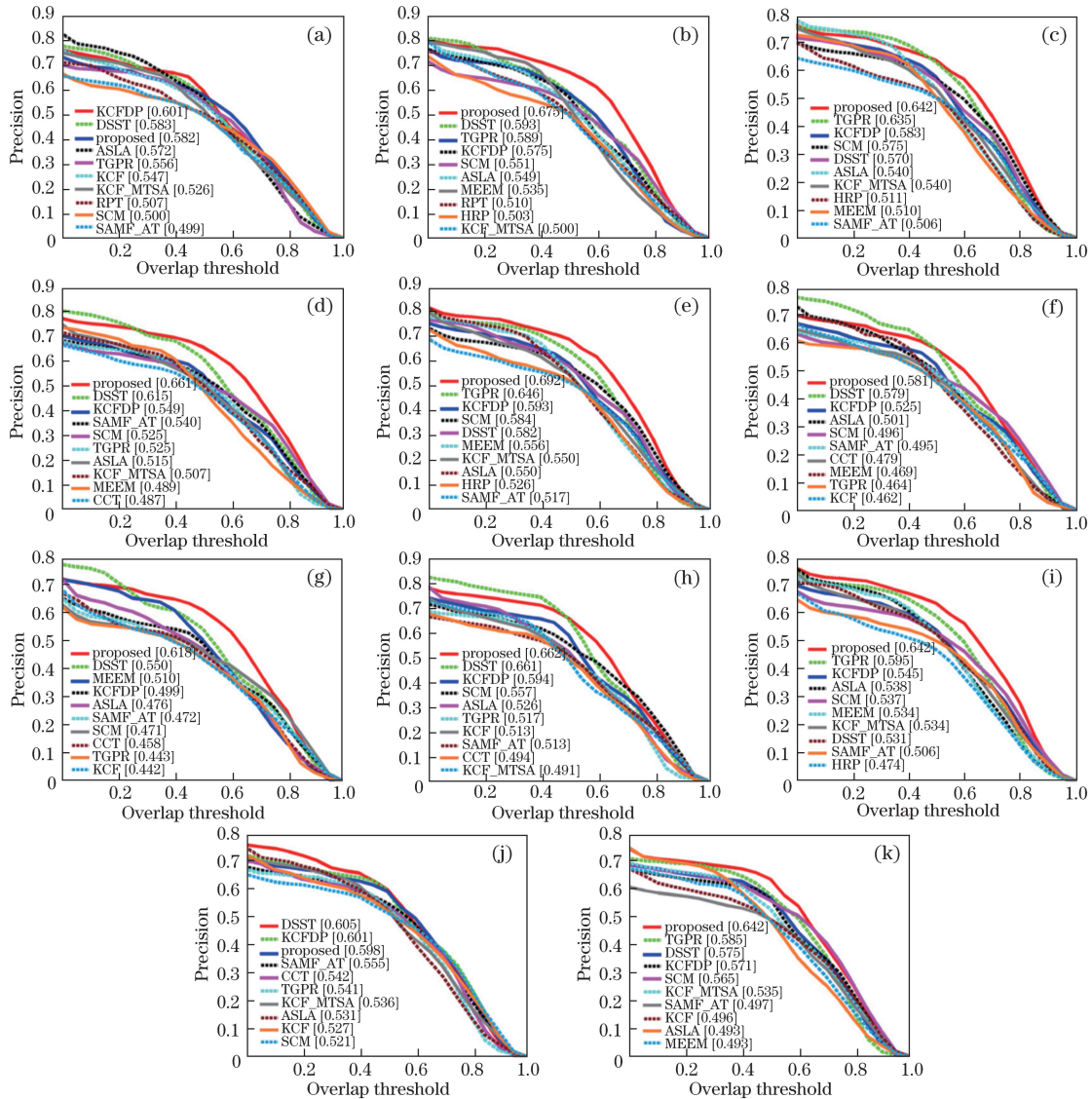


图 7 全部属性的成功率图。(a)光照变化;(b)平面外旋转;(c)遮挡;(d)尺度变化;(e)形变;(f)快速运动;(g)运动模糊;(h)平面内旋转;(i)背景杂乱;(j)出视野;(k)低分辨率

Fig. 7 Success rate plots of all attributions. (a) Illumination variation; (b) out-of-plane rotation; (c) occlusion; (d) scale variation; (e) deformation; (f) fast motion; (g) motion blur; (h) in-plane rotation; (i) background clutter; (j) out of view; (k) low resolution

3.5 定性分析

采用本文算法与其他 9 种算法分别对 9 个视频序列中的目标进行跟踪,当目标经历不同情况时,分析不同算法对应的跟踪结果,如图 8 所示。

1) 平面外旋转。当跟踪目标经历平面外旋转时,本文算法与其他算法的对比情况如图 8(a)所示,在视频序列 birds 中,当该女士转身的时候,很多跟踪算法发生了漂移,甚至有算法将目标定位于另一位男士身上,而本文算法凭借 HSC 特征对目标极强的表达能力以及 DAM 有效排除背景干扰的能力可以精确地对目标进行跟踪。在 running_rhino 视频序列中,当犀牛转弯的时候,有些算法对尺度的估计不准确,将很多冗余背景信息当作目标进行框选,导致后续跟踪失败,有些算法发生了漂移,而本文算法不仅未发生漂移,而且

对尺度的估计也很准确,确保了后续精确跟踪。综上所述,对于平面外旋转的情况本文算法有足够能力对目标进行表达,确保漂移和尺度估计错误不会发生。

2) 形变。如图 8(b)所示,在视频序列 birds 中,当目标发生严重的形变时,只有少数算法没有发生漂移,本文算法为其中之一。得益于 DAM 的使用,本文算法可以在目标发生严重形变时仍然具有足够的表达能力,实现对目标的追踪,相同的情况适用于 crouching 视频序列。当跟踪目标(人)蹲伏时,有一些算法的尺度估计产生严重的错误,这些错误会在后续的跟踪中累积,导致跟踪失败,而本文算法仍然可以准确地预测目标的尺度和位置,为后续跟踪提供可靠的信息,证明了本文算法的稳健性。

3) 背景杂乱。红外目标跟踪中,背景杂乱的影响十分严重。受红外成像原理的影响,红外目标缺少颜色、纹理等有效的信息,而这些信息可以为目标跟踪提供很大帮助。当背景中存在与目标相似的物体时,很多跟踪算法容易发生漂移,例如 crowd 和 mixed_distractor 视频序列的情况。如图 8(c)所示,在 crowd 视频序列中,很多人向相同的方向移动,在缺少颜色、纹理等信息时,绝大多数跟踪算法发生了漂移,一些算法跟踪了错误的目标,一些算法跟踪了部分背景。同样的情况出现在 mixed_distractor 视频序列中,当跟踪目标经过第一个背景中的人时绝大多数跟踪算法产生了漂移,此时,本文算法仍然可以对目标进行精确跟踪,而在目标经历形变和平面外旋转之后,又有几个跟踪算法对尺度的估计错误,此时本文算法仍然保持了对目标的精确跟踪。

4) 遮挡。遮挡有部分遮挡和全部遮挡的情形,如图 8(d)所示,street 视频序列中目标被部分遮挡,hiding 序列中目标被全部遮挡。可以看出相似物体遮挡目标的情况与可见光图像相比更加难以处理,因为算法有很大概率会发生漂移。本文算法在两个视频序列中均表现优秀,不仅可以处理部分遮挡,还可以处理全部遮挡。

5) 尺度变化。如图 8(e)所示,视频序列 jacket 和 selma 中,跟踪目标均经历了较大的尺度变化,当目标尺度持续变化时,很多跟踪算法不能有效地对目标尺度进行估计,在后续跟踪中产生不同程度的错误,这些错误会逐渐累积,导致跟踪的失败。本文算法对于尺度的准确估计使其获得更加精确的跟踪结果。

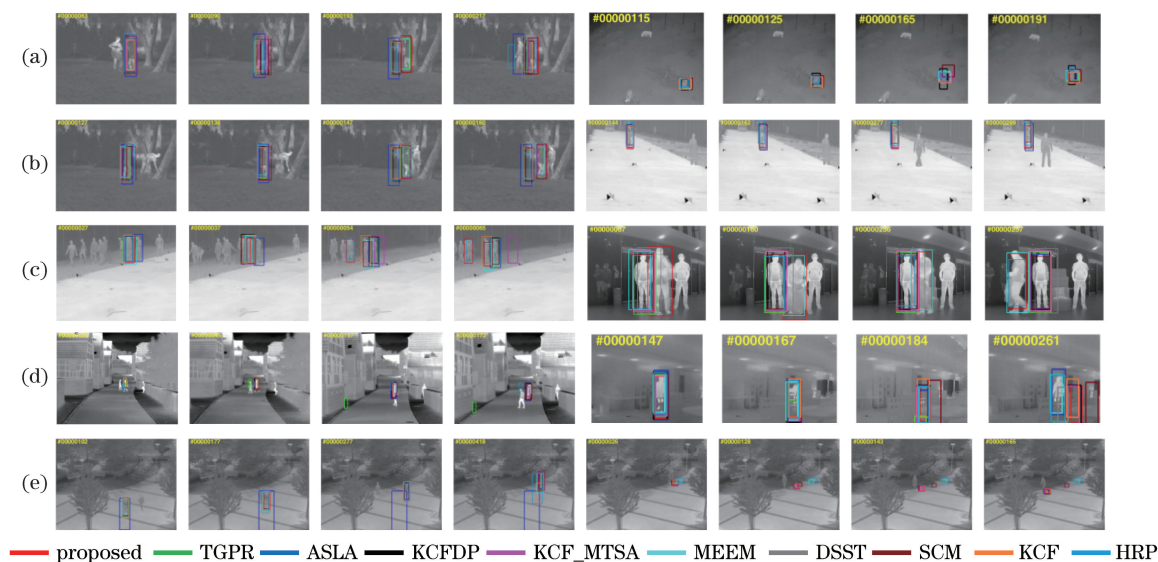


图 8 本文算法与其他 9 种算法在 9 个视频序列上的跟踪结果。(a)平面外旋转(birds 和 running_rhino 序列);
 (b)形变(birds 和 crouching 序列);(c)背景杂乱(crowd 和 mixed_distractors 序列);
 (d)遮挡(street 和 hiding 序列);(e)尺度变化(jacket 和 selma 序列)

Fig. 8 Tracking results of the proposed algorithm and another nine trackers on nine image sequences.
 (a) Out-of-plane rotation (birds and running_rhino sequences); (b) deformation (birds and crouching sequences);
 (c) background clutter (crowd and mixed_distractors sequences); (d) occlusion (street and hiding sequences);
 (d) scale variation (jacket and selma sequences)

4 结 论

提出一种基于 HSC 特征和 DAM 的红外目标跟踪方法。HSC 特征的表达能力比 HOG 特征更强,可以对目标物体进行更加详细的表达,因此适用于缺少细节和颜色特征的红外目标跟踪任务。利用 DAM 降低红外图像的背景干扰,既消除噪声影响,又增加对图像灰度信息的利用。两种方法的结合使本文算法在 VOT-TIR2015 数据集上取得比其他 22 种红外目标跟踪算法更好的结果。本文算法目前不能处理目标纵横比变化的情况,并且两种得分的融合方式可以进行自适应改进,这将是未来改进的主要方向。

参 考 文 献

- [1] He Y J, Li M, Zhang J L, *et al.* Infrared target tracking via weighted correlation filter[J]. *Infrared Physics & Technology*, 2015, 73: 103-114.
- [2] Huang Y, Zheng H, Yang H. Improving an object tracker for infrared flying bird tracking[C]. *International Conference on Image Processing*, 2016: 1699-1703.
- [3] Demir S, Cetin E. Co-difference based object tracking algorithm for infrared videos[C]. *International Conference on Image Processing*, 2016: 434-438.
- [4] Zhang Hui, Zhao Baojun, Tang Linbo, *et al.* Infrared object tracking based on adaptive multi-features integrations[J]. *Acta Optica Sinica*, 2010, 30(5): 1291-1296.
张辉, 赵保军, 唐林波, 等. 基于自适应多特征整合的红外目标跟踪[J]. *光学学报*, 2010, 30(5): 1291-1296.
- [5] Guan Zhiqiang, Chen Qian, Qian Weixian, *et al.* Infrared target tracking algorithm based on algorithm fusion[J]. *Acta Optica Sinica*, 2008, 28(5): 860-865.
管志强, 陈钱, 钱惟贤, 等. 一种基于算法融合的红外目标跟踪算法[J]. *光学学报*, 2008, 28(5): 860-865.
- [6] Sun Jin, Ding Yonghui, Zhou Lai. Visually interactive hand tracking algorithm combined with infrared depth tracking[J]. *Acta Optica Sinica*, 2017, 37(1): 0115002.
孙瑾, 丁永辉, 周来. 融合红外深度信息的视觉交互手部跟踪算法[J]. *光学学报*, 2017, 37(1): 0115002.
- [7] Zhao Aigang, Wang Hongli, Yang Xiaogang, *et al.* An affine invariant method of forward looking infra-red target recognition[J]. *Laser & Optoelectronics Progress*, 2015, 52(7): 071501.
赵爱罡, 王宏力, 杨小冈, 等. 一种仿射不变的前视红外目标识别方法[J]. *激光与光电子学进展*, 2015, 52(7): 071501.
- [8] Bao C L, Wu Y, Ling H B, *et al.* Real time robust l1 tracker using accelerated proximal gradient approach[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2012: 1830-1837.
- [9] Danelljan M, Khan F S, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1090-1097.
- [10] Zhang K H, Zhang L, Liu Q S, *et al.* Fast visual tracking via dense spatio-temporal context learning[C]. *European Conference on Computer Vision*, 2014: 127-141.
- [11] Hare S, Golodetz S, Saffari A, *et al.* Struck: structured output tracking with kernels[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(10): 2096-2109.
- [12] Babenko B, Yang M H, Belongie S. Visual tracking with online multiple instance learning[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 983-990.
- [13] Zhu G, Porikli F, Li H. Beyond local search: tracking objects everywhere with instance-specific proposals[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 943-951.
- [14] Henriques J F, Caseiro R, Martins P, *et al.* High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.
- [15] Bolme D S, Beveridge J R, Draper B A, *et al.* Visual object tracking using adaptive correlation filters[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010: 2544-2550.
- [16] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[C]. *European Conference on Computer Vision, Workshop VOT2014*, 2014: 254-265.
- [17] Danelljan M, Häger G, Khan F S, *et al.* Accurate scale estimation for robust visual tracking[C]. *British Machine Vision Conference*, 2014.
- [18] Danelljan M, Häger G, Khan F S, *et al.* Learning spatially regularized correlation filters for visual tracking[C]. *IEEE International Conference on Computer Vision*, 2015: 4310-4318.
- [19] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. *IEEE Computer Society Conference on*

- Computer Vision and Pattern Recognition, 2005: 886-893.
- [20] Ren X F, Ramanan D. Histograms of sparse codes for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 3246-3253.
- [21] Aharon M, Elad M, Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.
- [22] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking[C]. International Conference on Computer Vision and Pattern Recognition, 2016: 1401-1409.
- [23] Possegger H, Mauthner T, Bischof H. Indefense of color-based model-free tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2113-2120.
- [24] Alahari K, Berg A, Hager G, *et al.* The thermal infrared visual object tracking VOT-TIR2015 challenge results[C]. IEEE International Conference on Computer Vision Workshop, 2015: 639-651.
- [25] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.
- [26] Gao J, Ling H B, Hu W M, *et al.* Transfer learning based visual tracking with Gaussian process regression[C]. European Conference on Computer Vision, 2014, 8691: 188-203.
- [27] Zhang J M, Ma S G, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization[C]. European Conference on Computer Vision, 2014, 8694: 188-203.