

融合深度特征与梯度特征的红外目标跟踪算法

吴 敏, 杨 源, 张园强, 库 涛, 查宇飞, 张胜杰

(空军工程大学航空航天工程学院, 西安, 710038)

摘要 由于红外序列图像目标与背景的对比度低, 图像的边缘模糊并且灰度级动态范围小, 采用何种特征描述目标成为跟踪的关键。深度特征和梯度特征是日前大部分跟踪算法采用的主要特征, 然而深度特征提取的目标语义信息关注类间分类(Intra-Class), 忽略类内差别, 容易受到相似背景(Distractor)干扰; 梯度特征作为局部区域特征不易受背景干扰, 但不能适应目标的剧烈形变。基于这 2 种特征的互补性, 提出一种融合深度特征和梯度特征的红外目标跟踪算法。深度特征与梯度特征被分别用来表征目标的语义信息与局部结构信息, 增强了对任意目标的表征能力; 利用不同特征建立的跟踪模型进一步提高了跟踪的鲁棒性。通过建立模型互助机制, 利用深度特征跟踪模型与梯度特征跟踪模型的互补性, 对目标实施了精准的定位。实验中, 选取了最新的红外视频跟踪数据库(VOT-TIR2016)用来验证文中算法的有效性, 结果表明: 和当今主流跟踪算法相比, 算法在精确度上获得了 3.8% 的提升, 在成功率上获得了 4.3% 的提升, 能够有效处理跟踪中相似背景与形变的影响。

关键词 红外序列图像; 梯度特征; 深度特征; 特征融合

DOI 10.3969/j.issn.1009-3516.2017.06.013

中图分类号 TP391.41; TN911.73 **文献标志码** A **文章编号** 1009-3516(2017)06-0076-07

An Infrared Target Tracking Algorithm Based on the Fusion of Deep Feature and Gradient Feature

WU Min, YANG Yuan, ZHANG Yuanqiang, KU Tao, ZHA Yufei, ZHANG Shengjie

(Aeronautics and Astronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: Based on the complementarity of these two features, this paper presents an infrared target tracking algorithm in combination with the deep feature and the gradient feature. The deep feature and gradient feature are used to represent the semantic information and the local structure information in this paper, enhancing the ability to characterize any target. Next, the tracking models based on different features can further improve the tracking performance. Finally, the paper builds up a model mutual mechanism based on the combination of deep feature tracking model and gradient feature tracking model, implementing positioning precisely to the target. The article selects the latest infrared tracking database and uses the database to verify the effectiveness of the algorithm. The results show that the algorithm in this paper achieves a 3.8% improvement in accuracy and obtains a 4.3% improvement in success rate, enabling to effectively

收稿日期: 2017-03-29

基金项目: 国家自然科学基金(61472442; 61773397; 61701524)

作者简介: 吴 敏(1994—), 男, 湖南常德人, 硕士生, 主要从事视频图像研究。E-mail: 1820304877@qq.com

引用格式: 吴敏, 杨源, 张园强, 等. 融合深度特征与梯度特征的红外目标跟踪算法 [J]. 空军工程大学学报(自然科学版), 2017, 18(6): 76-82. WU Min, YANG Yuan, ZHANG Yuanqiang, et al. An Infrared Target Tracking Algorithm Based on the Fusion of Deep Feature and Gradient Feature [J]. Journal of Air Force Engineering University (Natural Science Edition), 2017, 18(6): 76-82.

handle the effects of the background similarly and deformation.

Key words: the infrared image sequences; gradient feature; deep feature; the fusion of the features

红外目标跟踪是红外预警系统、空载下视觉系统对低空及地面目标的监视和红外寻地制导等军事领域中的关键技术^[1]。由于红外传感器不向空中辐射能量,只是通过接受目标辐射的热量对目标进行探测和跟踪,因而不易被侦察或定位,具有较强的抗干扰能力;同时,由于目标不可避免地要辐射热量,从而又为使用红外传感器对目标探测和跟踪创造了条件。此外,红外传感器还具有测角精度高和目标识别能力强等优点。因此,红外传感器作为一种重要的被动传感器得到了广泛的应用,使红外传感器信号检测和目标跟踪的研究也逐步深入^[2]。

红外图像的目标与背景的对比较低,图像的边缘模糊并且灰度级动态范围小^[3],使得红外图像跟踪成为一项具有挑战性的任务^[4]。为了提高跟踪的精确度,以相关滤波为基础的算法^[5]在跟踪领域得到了很大的发展,并解决了很多实际问题。经典的 KCF^[6]算法采用梯度特征作为输入特征,利用边缘方向分布描述红外目标局部的表象和形状,选取与红外跟踪目标相似度最大的候选目标,从而实现跟踪。但梯度特征作为跟踪目标的局部特征,当目标发生较大形变时,容易导致跟踪目标的丢失。随着深度学习的推广,基于深度特征的 HCF^[7]算法得到广泛的应用,HCF 算法提取的是 CNN 网络的卷积层特征,高层特征体现目标的语义信息,低层特征体现目标的空间特性,但由于 CNN 网络更关注区分类别间物体,忽略类内差别,从而在跟踪特定红外目标上具有一定的局限性。

从深度特征和梯度特征 2 个方面对目标进行特征提取,利用包含红外目标语义信息的深度特征和表征红外目标局部区域信息的梯度特征,采用正则化最小二乘法建立红外目标跟踪模型,根据 2 种特征互补的性质,利用 2 种特征对目标的响应值确定各自的权重,使用互助机制结合不同的预测结果,从而实现对红外目标的跟踪。与 KCF、HCF 相比,本文算法提高了跟踪的精确度和鲁棒性。

1 深度-梯度特征融合

针对红外图像中描述目标的难题,本文采用深度-梯度融合特征,结合深度特征的语义信息和梯度特征边缘信息,精确描述目标轨迹。

本文采用的深度特征是通过组合低层特征形成的更加抽象的高层特征,表示目标的属性、类别或特征,广泛应用于物体检测识别领域。深度特征提取的是 CNN 网络的卷积层特征,关键优势在于高层特征主要反映目标的语义特性^[9],对目标的表现比较鲁棒,低层特征保存了更多的空间特征^[10],对跟踪目标精确定位更有效。以上一帧的预测结果为中心,将目标框若干倍的放大,确定搜索框。用 CNN 网络提取搜索框的特征,采用双线性插值法对特征进行上采样,获取卷积层特征。为了充分提取目标的语义信息,又保留准确的空间信息,提取 Conv3_4, Conv4_4, Conv5_4 3 层卷积层的特征。深度特征提取过程见图 2。

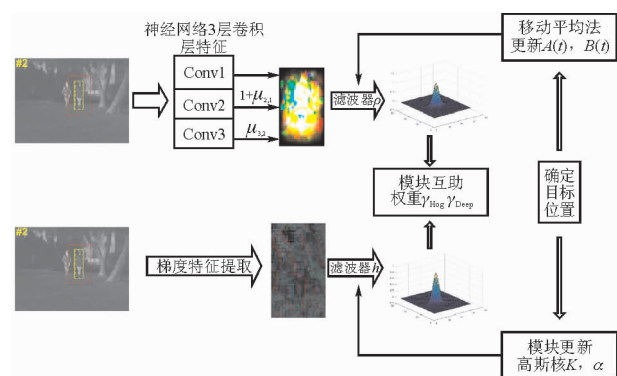


图 1 跟踪算法流程图

Fig. 1 The flow chart of tracking algorithm

针对 2 种算法的优缺点,本文结合大量红外跟踪算法,并参考了 staple^[8]中将 2 种特征融合的方法,提出一种融合深度特征和梯度特征的红外目标跟踪算法。图 1 为跟踪算法流程图,本文算法分别

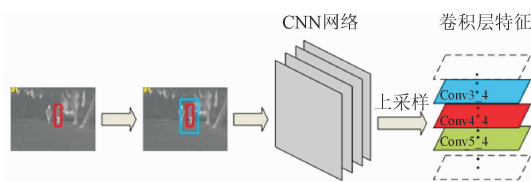


图 2 深度特征提取过程

Fig. 2 The process of extracting deep feature

梯度特征是一种用于表征图像局部梯度方向和梯度强度分布特征的描述符^[11]。在跟踪领域,梯度特征中的梯度方向直方图特征(HOG 特征)被认为是获取边缘和局部形状信息最好的特征之一。为了更好的获取边缘和局部形状信息,提取 HOG 特征作为红外目标的梯度特征。将以预测目标为中心的搜索框归一化,再将搜索框划分为多个 4×4 像素得元胞(cell),每 1 个元胞都按 360° 方向分成 9 个大小一样的区间,统计元胞中的像素在 9 个区间里的梯

度直方图得到1个9维的特征矩阵,相邻的4个元胞可以组成1个块(block),则块内的梯度直方图为 $4 \times 9 = 36$ 维特征矩阵,再将1个块以1个元胞的大小为步长沿水平方向和垂直方向扫描整个搜索框,最后将所得到的块的特征组合在一起即为该搜索框的梯度特征。

由于深度特征提取的是CNN网络的卷积层特征,其所包含的目标语义信息更关注于类间分类,忽略类内差别,容易受到相似背景干扰,而梯度特征作为局部区域特征不易受背景干扰,弥补深度特征的缺陷。因此,本文利用包含红外目标语义信息的深度特征和表征红外目标局部区域信息的梯度特征,采用正则化最小二乘法建立红外目标跟踪模型,利用深度特征与梯度特征的互补性,由2种特征对目标的响应值确定各自的权重,建立互助机制结合不同的预测结果,从而确定目标的位置。深度特征和梯度特征融合见图3。

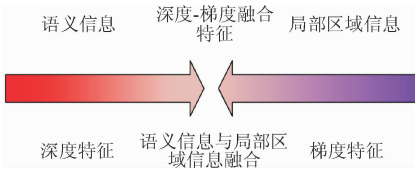


图3 深度特征和梯度特征融合

Fig. 3 The fusion of deep feature and gradient feature

采用深度-梯度融合特征不仅可以提取红外目标语义信息,实现类间跟踪的良好效果,还可以获取目标的局部区域信息,提高在类内跟踪的精确度。

2 融合深度特征与梯度特征的红外目标跟踪算法

2.1 模型的构建

本文的跟踪算法采用先检测后跟踪的策略,即先利用初始帧的目标的位置,训练一个目标检测器,再使用目标检测器去检测下一帧预测位置是否有目标,然后再使用新检测结果去更新训练集进而更新目标检测器。在第 t 帧图像训练目标检测器时,在搜索框 x_t 中,对于同一目标,可能得到多个目标框 S_t ,选取其中响应值最高的目标框为图像中目标的位置 p_t :

$$p_t = \operatorname{argmax}_{p \in S_t} f(T(x_t, p); \mathbf{h}, \boldsymbol{\rho}) \quad (1)$$

式中: $f(T(x_t, p); \mathbf{h}, \boldsymbol{\rho})$ 是依据模型的模型参数 \mathbf{h} 、 $\boldsymbol{\rho}$,计算多个目标框在搜索框 x_t 中对应真实目标框 p 的响应值; $T(x_t, p)$ 函数是提取搜索框 x_t 和目标框 p 的梯度特征和深度特征; $f(T(x_t, p); \mathbf{h}, \boldsymbol{\rho})$ 可以得到预测目标与检测目标之间的匹配情况,响应值

越高,越接近真实目标。

本文通过融合深度特征和梯度特征,根据深度特征的损失值 l_{Hog} 和梯度特征的损失值 l_{Hog} ,采用正则化最小二乘法建立模型的损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$,公式表示为:

$$L(\mathbf{h}, \boldsymbol{\rho}) = \min_{\mathbf{h}, \boldsymbol{\rho}} \gamma_{\text{Hog}} l_{\text{Hog}} + \gamma_{\text{Deep}} l_{\text{Deep}} + \lambda \|\mathbf{h}\|^2 + \beta \|\boldsymbol{\rho}\|^2 \quad (2)$$

式中: γ_{Hog} 为梯度特征的权重; γ_{Deep} 为深度特征的权重; λ 、 β 是正则化参数,通过调整 λ 、 β 的值来控制过拟合。

提取目标框的梯度特征,利用滤波器 \mathbf{h} 将目标框转换为其对应真实目标框的响应值,得到梯度特征的损失值:

$$l_{\text{Hog}} = \sum_i (\mathbf{h}^T \boldsymbol{\phi}_i - y_i)^2 \quad (3)$$

式中: T 为向量的转置; $\boldsymbol{\phi}_i$ 为第 i 个目标框的梯度特征; y_i 为第 i 个目标框的标签。

深度特征提取的是CNN网络中3层卷积层的特征,以由粗粒度到细粒度的方法得到深度特征的损失值:

$$l_{\text{Deep}} = \sum_i \left(\boldsymbol{\rho}^T \sum_{l=3-i, i=0}^{i=1} \frac{1}{1 + \mu_{l, i-1}} (\boldsymbol{\psi}_i^l + \mu_{l, i-1} \boldsymbol{\psi}_i^{l-1}) - y_i \right)^2 \quad (4)$$

式中: $\boldsymbol{\rho}^l$ 为第 l 层的相关滤波器; $\boldsymbol{\psi}_i^l$ 为第 i 个目标框的第 l 层卷积层特征; $\mu_{l, i-1}$ 表示第 l 层的特征对第 $l-1$ 层的特征的约束值。

2.2 梯度特征模块

将目标框的图像分为若干个细胞单元,将各个细胞单元组合成一个块,将所有的块并联,从而得到目标框的梯度特征。

$$\boldsymbol{\phi}_{m+1} = \operatorname{par}(\boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m+1}), m = 1, 2, \dots, n \quad (5)$$

式中: $\boldsymbol{\alpha}_m$ 表示第 m 个块的梯度特征, n 为目标框中块的数目, $\operatorname{par}(\boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m+1})$ 是将第 m 个块和第 $m+1$ 个块的梯度特征并联, $\boldsymbol{\phi}_{m+1}$ 是并联了 $m+1$ 个块的梯度特征。每个块都代表着每个局部的梯度信息,目标框的梯度特征是由若干个局部的梯度特征组成,反映的是目标框的局部区域信息。

通过损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$ 对 \mathbf{h} 求偏导,可以得到有关梯度特征的滤波器 h 的函数。对输入的梯度特征进行循环,循环后的特征为 $\boldsymbol{\phi} = \mathbf{F}^H \operatorname{diag}(\boldsymbol{\phi}) \mathbf{F}$,根据循环矩阵能够被离散傅里叶矩阵对角化,使得矩阵求逆转换为特征值求逆的性质,能够将 h 转换到频域进行计算,从而大大减小了计算的复杂度,并且增加了样本的数目,提高了模型的性能。此时,模型参数 h 的函数为:

$$\mathbf{h} = \frac{\gamma_{\text{Hog}} \mathbf{y} \otimes \boldsymbol{\phi}}{\gamma_{\text{Hog}} \boldsymbol{\phi} \otimes \boldsymbol{\phi} + \lambda} \quad (6)$$

引用一个非线性映射 $\alpha(\phi)$ 列向量,使映射后的样本在新的空间中线性可分, $\alpha(\phi)\alpha(\phi)^T$ 是类似核空间变量的协方差矩阵,令 K 表示核空间的核矩阵。由于高斯核是采用像素邻域的加权均值来代替该点的像素值,而每一邻域像素点的权值是随该点与中心的距离单调递减的,这正好适用于梯度特征这种图像局部特征。通过在新的空间中利用高斯核 $k^{\phi_t, \phi_{t-1}}$ 和构建一个在二维空间的岭回归问题 ∂ , 可以得到:

$$\mathbf{h}^T \phi_t = \hat{\partial} * \hat{k}^{\phi_t, \phi_{t-1}} \quad (7)$$

其中:

$$\partial = \frac{\gamma_{\text{Hog}} \mathbf{y}}{\phi \otimes \phi + \lambda} \quad (8)$$

$$k^{\phi_t, \phi_{t-1}} = \exp\left(-\frac{1}{\sigma^2} \left(\|\phi\|^2 + \|\phi_{t-1}\|^2 - 2F^{-1}\left(\sum \phi_i * \phi_{i-1}\right)\right)\right) \quad (9)$$

式中: ϕ 为第 t 帧的目标框的梯度特征; σ 为高斯核的方差。

2.3 深度特征模块

用 CNN 网络提取目标框的特征,根据神经网络中越靠后的层中图像的语义信息越丰富的性质,为了充分提取目标语义信息,并保留准确空间信息,本文提取了 CNN 网络中 Conv3_4, Conv4_4, Conv5_4 这 3 层卷积层的特征作为目标框的深度特征。

通过损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$ 对 $\boldsymbol{\rho}$ 求偏导,可以得到有关深度特征的第 t 帧的滤波器 $\boldsymbol{\rho}(t)$ 的函数:

$$\boldsymbol{\rho}(t) = \frac{\gamma_{\text{Deep}} \in \mathbf{A}(t)}{\gamma_{\text{Deep}} \mathbf{B}(t) \otimes \mathbf{B}(t) + \beta} \quad (10)$$

其中:

$$\mathbf{A}(t) = \mathbf{y} \otimes \sum_{l=3-i, i=0}^{i=1} \frac{1}{1 + \mu_{l, t-1}} (\boldsymbol{\psi}_i^l + \mu_{l, t-1} \boldsymbol{\psi}_i^{l-1}) \quad (11)$$

$$\mathbf{B}(t) = \sum_{l=3-i, i=0}^{i=1} \frac{1}{1 + \mu_{l, t-1}} (\boldsymbol{\psi}^l + \mu_{l, t-1} \boldsymbol{\psi}^{l-1}) \quad (12)$$

第 t 帧的最优滤波器可以通过由求在所有的跟踪结果中的最小输出误差来描述更新,然而这涉及到求解线性方程组的问题,为了获得鲁棒的近似值,本文通过使用移动平均的方法更新滤波器的分子和分母达到模块更新的目的。

$$\hat{\mathbf{A}}(t) = (1 - \eta_{\text{Hog}}) \hat{\mathbf{A}}(t-1) + \eta_{\text{Hog}} \hat{\mathbf{A}}(t)' \quad (13)$$

$$\hat{\mathbf{B}}(t) = (1 - \eta_{\text{Hog}}) \hat{\mathbf{B}}(t-1) + \eta_{\text{Hog}} \hat{\mathbf{B}}(t)' \quad (14)$$

式中: η_{Hog} 为每一帧的学习速率。

2.4 模块互助

由于深度特征模块和梯度特征模块在对目标位置进行预测时都会存在误差,误差越大其对目标的位置预测结果越不准。为了结合训练的模块得到一

个最终的模型,本文设计了每个模块的权重,并且权重是以每个模块的误差为基础,模块的权重大小与误差的大小成反比。通过使用模块互助的方式结合不同的估计来减少不准确的预测,在一定程度上可以提高最终的模型的跟踪精确度。由于模块的响应值体现了真实目标与候选目标之间的匹配情况,相似度越高的候选框的,其语义信息和局部区域信息越能反映真实目标,因此,本文用每个模块的响应值来对模块进行评价,并以此来定义误差:

$$\boldsymbol{\varepsilon} = 1 - \max_{p \in s_j} f(T(\mathbf{x}_i, \mathbf{p}); \mathbf{h}, \boldsymbol{\rho}) \quad (15)$$

通过利用各个模块的最大响应值与真实目标标签的差值反应模块的误差,结合 2 个模块的误差可以得到每个模块的权重:

$$\gamma_{\text{Hog}} = 1 - \frac{\varepsilon_{\text{Hog}} + \tau}{\varepsilon_{\text{Hog}} + \varepsilon_{\text{Deep}} + \tau} \quad (16)$$

$$\gamma_{\text{Deep}} = 1 - \frac{\varepsilon_{\text{Deep}} + \tau}{\varepsilon_{\text{Hog}} + \varepsilon_{\text{Deep}} + \tau} \quad (17)$$

式中: τ 是一个接近于 0 的常数,其作用是避免当 ε_{Hog} 和 $\varepsilon_{\text{Deep}}$ 为 0 时导致分母为 0,无法得到模块的权重的情况。

2.5 算法总结

结合深度特征与传统特征的红外目标跟踪算法如下:

输入: 红外图像, 初始的目标位置 O_1 ;

输出: 当前帧的目标位置 O_i ;

1) 选取图像的正负样本,分别提取目标的深度特征和梯度特征,根据式(4)中深度特征的损失值 l_{Deep} 和式(5)中梯度特征的损失值 l_{Hog} ,采用正则化最小二乘法建立式(2)中模型的损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$;

2) 通过损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$ 对 \mathbf{h} 求偏导,可以得到有关梯度特征的滤波器 \mathbf{h} 的函数,利用高斯核 $k^{\phi_t, \phi_{t-1}}$ 和构建一个在二维空间的岭回归问题 ∂ ,实现对梯度特征模块的建立与更新;

3) 通过损失函数 $L(\mathbf{h}, \boldsymbol{\rho})$ 对 $\boldsymbol{\rho}$ 求偏导,得到有关深度特征的滤波器 $\boldsymbol{\rho}$,使用移动平均的方法更新滤波器的分子和分母达到模块更新的目的;

4) 通过利用每个模块对目标位置的响应值大小,确定每个模块的误差。根据模块的权重大小与误差的大小成反比的性质,得到每个模块的权重,实现模块互助。

3 实验结果与分析

为验证跟踪算法的有效性,本文进行了大量的仿真实验,实验采用的是核岭回归模型,模型中正则化参数 $\lambda = 0.01$ ^[12-13],高斯核方差 $\sigma = 0.5$,模型学习

速率 $\eta=0.01^{[14]}$ 。为了充分说明本文算法在处理红外目标跟踪的优势,本文选取了4组测试视频和3种对比算法,对比算法分别是利用深度特征的HCF算法,利用梯度特征的KCF算法和基于KCF改进的KCFDP^[15]算法。

3.1 定性分析

图4为红外跟踪结果示意图,其中第1段视频跟踪的目标是1架无人机,分别测试采用3种不同算法进行非生命物体的跟踪效果。实验结果如图所示,在图像的第61帧到第91帧时,目标经过窗户,由于此时窗户的亮度和目标的颜色相近,目标的边缘信息不够明显,此时,发现KCF, KCFDP, HCF算法都很难对目标进行跟踪,而本文算法可以有效的跟踪目标,体现出本文算法对目标和背景的边缘不明显的情况下跟踪精确度好于另外3种算法。

第2段视频是关于动物的跟踪,分别测试采用3种不同算法进行非人生物的跟踪效果。实验结果

如图所示,在图像的第120帧到第240帧时,目标发生了较大形变,此时,KCF和KCFDP算法的跟踪效果直接下降,而HCF和本文算法都能对目标进行有效跟踪,可以看出本文算法在物体发生形变时的跟踪精确度要优于KCF和KCFDP。

第3段视频是关于人的跟踪,分别采用3种不同的算法来进行人体的跟踪。实验结果如图所示,在整个视频的跟踪中,由于跟踪的目标是一个特定的人,而在视频中有其他人存在,这种相似目标会对目标跟踪造成很大的影响,由图中可以看出本文算法的跟踪精确度要优于另外3种算法。

第4段视频是关于小目标的跟踪,分别采用3种不同算法来进行小目标的跟踪。实验结果如图所示,在这整个视频的跟踪中,跟踪的物体较小,这对于提取详细的特征信息会造成很大的麻烦。本文算法结合了深度特征和传统特征的优点,不仅可以提取到目标的梯度信息,还能提取到目标的语义信息,提高了跟踪的精确度和成功率。

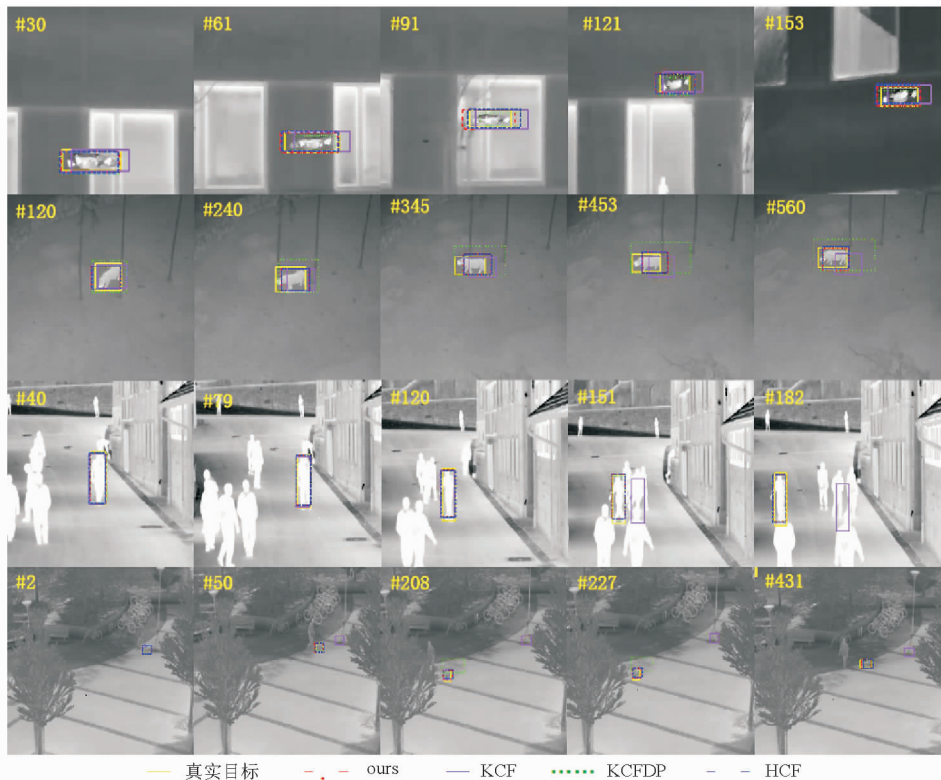


图4 跟踪结果示意图

Fig. 4 Schematic diagram of the tracking results

3.2 定量分析

本文算法采用了深度特征与HOG特征相结合的方法,利用2种特征在不同视频跟踪中的优点,进行优势互补,提高跟踪的效果。在本次实验中,为了更好的表现出本文算法的优越性,本文将3种算法在一个视频库中进行目标跟踪,并比较其跟踪结果,

其结果见图5。

本文以跟踪的精确度和成功率^[16-17]作为跟踪算法的评价指标,图5(a)是跟踪的精确度,其中横坐标为中心距离阈值,纵坐标为精确度,即预测目标的中心到真实目标的中心的距离低于阈值的帧数在整个视频库中所占的比重;图5(b)为跟踪的成功

率,横坐标为覆盖率 O_f 的阈值,纵坐标为成功率,即覆盖率大于阈值的帧数在整个视频库中的比重^[18]。从图中可以看出,本文算法有较好的跟踪性能。

$$O_f = \frac{Area(B_T \cap B_G)}{Area(B_T \cup B_G)} \quad (18)$$

式中: B_T 为算法求得的目标区域, B_G 为标定的真实目标区域, $Area(B_T \cup B_G)$ 为 2 个区域的重叠面积, $Area(B_T \cap B_G)$ 为 2 个区域并集的面积^[19]。

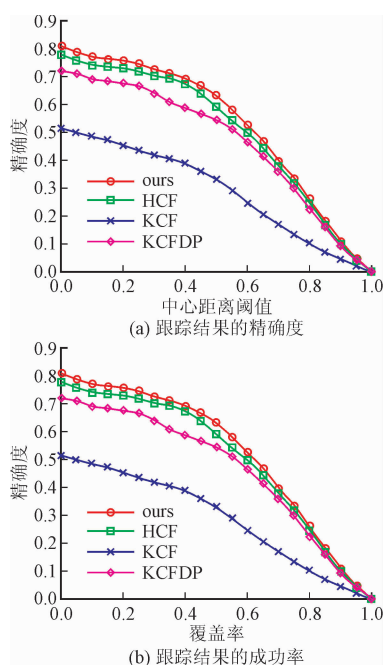


图 5 算法跟踪结果的精确度和成功率比较示意图

Fig. 5 The comparison diagram of the accuracy and success rate

表 1 算法跟踪结果的精确度和成功率

Tab. 1 The results of the accuracy and success rate

跟踪算法	精确度(20)	成功率(0.5)
ours	0.714 6	0.634 3
HCF	0.676 2	0.591 6
KCF	0.403 4	0.332 3
KCFDP	0.640 0	0.545 9

选取中心距离为 20 时的精确度和覆盖率为 0.5 时的成功率作为算法比较的标准^[20],由表 1 可见,基于梯度特征的 KCF 算法具有最差的跟踪效果,其跟踪精确度为 40.34%,成功率为 33.23%;基于深度特征的 HCF 算法的跟踪精确度为 67.62%,成功率为 59.16%,稍高于 KCF 算法;本文算法融合深度特征和梯度特征,具有最高的准确率 71.46%和最好的成功率 63.43%,和其余 3 种算法中跟踪效果最好的 HCF 算法相比,获得了跟踪精确度 3.8%的提升,获得了成功率 4.3%的提升。实验结果证

明,深度-梯度融合特征结合了 2 种特征的优点,实现了优势互补,提高了跟踪的精确度和成功率。

表 2 为 3 种对比算法和本文算法在红外视频跟踪数据库中所得到的跟踪速度, fps 为每秒所跟踪的视频帧数。从结果来看,利用深度特征的 HCF 算法和将深度特征与传统特征融合的本文算法的跟踪速度都要慢于利用传统特征的 KCF 算法和基于 KCF 算法改进的 KCFDP 算法,说明深度特征的提取过程比较复杂,具有较大的计算量^[21];通过将 HCF 算法和本文算法进行比较,发现跟踪的速度相差不大,说明融合了传统特征后对跟踪速度的影响不大。

表 2 跟踪算法的跟踪速度

Tab. 2 Tracking speed of the tracking algorithm

算法	HCF	KCF	KCFDP	ours
fps/(帧·s ⁻¹)	16.31	354.07	49.96	12.61

总的来说,融合了深度特征和传统特征的本文算法在跟踪速度上并没有明显的降低,能达到实时性的效果,但在跟踪精确度和成功率上有了很大的改进。

4 结语

针对在红外目标跟踪中,深度特征忽略类内差别、梯度特征不适应目标形变的问题,本文提出了一种将深度特征和梯度特征融合的方法,利用包含红外目标语义信息的深度特征和表征红外目标局部区域信息的梯度特征,采用正则化最小二乘法建立红外目标跟踪模型^[22],使用互助的方式结合不同的预测结果,从而确定目标的位置。通过实验,与 HCF 算法,KCF 算法和 KCFDP 算法相比的结果,充分证明了本文算法的有效性和优越性。

参考文献(References):

[1] FAN G, VENKATARAMAN V, TANG L, et al. On Boosted and Adaptive Particle Filters for Affine-Invariant Target Tracking in Infrared Imagery [J]. Augmented Vision Perception in Infrared, 2009, 3889(s2-3): 441-466.
 [2] 连浩, 韩传久. 基于 Mean Shift 的红外目标自动跟踪方法 [J]. 微计算机信息, 2008, 24(4): 285-287.
 LIAN H, HAN C J. Infrared Target Auto-Tracking Method Based on Mean Shift Algorithm [J]. Microcomputer Information, 2008, 24(4): 285-287. (in Chinese)
 [3] YILMAZ A, SHAFIQUE K, SHAH M. Target

- Tracking in Airborne forward Looking Infrared Imagery [J]. *Image&Vision Computing*, 2003, 21(7): 623-635.
- [4] DANELLJAN M, KHAN F S, FELSBURG M, et al. Adaptive Color Attributes for Real-Time Visual Tracking [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014: 1090-1097.
- [5] 张园强, 毕笃彦, 查宇飞, 等. 一种利用最大间隔相关滤波的鲁棒目标跟踪算法 [J]. *空军工程大学学报(自然科学版)*, 2017, 18(2): 63-68.
ZHANG Y Q, BI D Y, ZHA Y F, et al. A Robust Object-Tracking Algorithm Using Maximum Margin Correlation Filter [J]. *Journal of Air Force Engineering University (Natural Science Edition)*, 2017, 18(2): 63-68. (in Chinese)
- [6] HENRIQUES J F, RUI C, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 37(3): 583-596.
- [7] MA C, HUANG J B, YANG X, et al. Hierarchical Convolutional Features for Visual Tracking [C] // IEEE International Conference on Computer Vision. IEEE, 2015: 3074-3082.
- [8] DANELLJAN M, HAGER G, KHAN F S, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking [C]//IEEE International Conference on Computer Vision. IEEE, 2015: 4310-4318.
- [9] CHANG O, CONSTANTE P, GORDON A, et al. A Novel Deep Neural Network That Uses Space-Time Features for Tracking and Recognizing a Moving Object [J]. *Journal of Artificial Intelligence and Soft Computing Research*, 2017, 7(2):125-136.
- [10] NEUMANN U, YOU S. Natural Feature Tracking for Augmented Reality [J]. *Multimedia IEEE Transactions on*, 1999, 1(1): 53-64.
- [11] WATANABE T, ITO S, YOKOI K. Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection [J]. *Information and Media Technologies*, 2010, 5(2): 659-667.
- [12] DANELLJAN M, HAGER G, KHAN F S, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking [C]//IEEE International Conference on Computer Vision. Santiago, Chile; IEEE, 2015: 4310-4318.
- [13] DANELLJAN M, HAGER G, KHAN F, et al. Accurate Scale Estimation for Robust Visual Tracking [J]. *British Machine Vision Conference*, 2014. DOI: 10.5244/C.28.65.
- [14] GALOOGAHI H K, SIM T, LUCEY S. Multi-Channel Correlation Filters [C]//IEEE International Conference on Computer Vision. Sydney, NSW; IEEE, 2014: 3072-3079.
- [15] HUANG D. Enable Scale and Aspect Ratio Adaptability in Visual Tracking with Detection Proposals [C] // British Machine Vision Conference, 2015. DOI:10.5244/c.29.185.
- [16] HENRIQUES J F, RUI C, MARTINS P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels [C]//Proceedings of the 12th European Conference on Computer Vision-ECCV. Berlin Heidelberg; Springer, 2012: 702-715.
- [17] WU Y, LIM Y, YANG M H. Online Object Tracking: A Benchmark [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013: 2411-2418.
- [18] WU Y, YANG M H. Object Tracking Benchmark [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, 37(9): 1834-1848.
- [19] VEDALDI A, ZISSERMAN A. Efficient Additive Kernels via Explicit Feature Maps [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2012, 34(3): 480-492.
- [20] GRAY R M. Toeplitz and Circulant Matrices: A Review [J]. *Foundation and Trends® in Communications and Information Theory*, 2006, 2(3):155-239.
- [21] DANELLJAN M, KHAN F S, FELABERG M, et al. Adaptive Color Attributes for Real-Time Visual Tracking [C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014: 1090-1097.
- [22] HENRIQUES J F, CARREIRA J, RUI C, et al. Beyond Hard Negative Mining: Efficient Detector Learning via Block-Circulant Decomposition [C] // IEEE International Conference on Computer Vision. IEEE Computer Society, 2013: 2760-2767.

(编辑: 徐楠楠)