

基于双核模型上下文的流形正则相关滤波跟踪算法

崔丽群¹, 贺情杰^{1*}, 何牧泽²

¹ 辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105;

² 北京超星公司, 北京 100000

摘要 为了兼顾算法的跟踪速度与精度,提出了一种基于双核模型上下文的流形正则相关滤波跟踪算法。其中,结合上下文相关框架与相关滤波算法的主模块承担主要跟踪任务,可弥补相关滤波学习模型中余弦窗过滤的背景信息。对上下文相关样本进行流形正则处理,可达到惩罚上下文相关框架、优化主模块模型的目的。辅助模块则结合了核相关滤波算法与卷积特征,当跟踪目标发生遮挡、形变或超出视距等情况,跟踪置信度低于经验阈值时启用辅助模块,防止主模块模型发生漂移。由于主模块的跟踪速度快、精度低,而辅助模块跟踪速度慢、精度高,两个模块可在速度和精度方面优势互补。在 OTB2015 和 VOT2016 数据集上的测试结果表明,本算法在精度和鲁棒性方面均超过了其他相关滤波算法。

关键词 图像处理; 目标跟踪; 相关滤波; 上下文; 流形正则

中图分类号 TP301

文献标志码 A

doi: 10.3788/LOP202158.1210011

Manifold Regular Correlation Filter Tracking Algorithm Based on Dual-Core Model Context

Cui Liqun¹, He Qingjie^{1*}, He Muze²

¹ College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China;

² Beijing Chaoxing Company, Beijing 100000, China

Abstract In this study, a manifold regular correlation filter tracking algorithm based on dual-core model context is proposed to balance the tracking speed and accuracy of the algorithm. The main module combines the context-related framework and the relevant filtering algorithm is responsible for the main tracking task, which can compensate for the background information filtered using the cosine window in the relevant filter-learning model. The manifold regular processing of context-related samples can achieve the purpose of penalizing context-related framework and optimizing the main module model. The auxiliary module combines kernel correlation filtering algorithms and convolution features. When the tracking target is occluded, deformed or exceeds the line of sight, the auxiliary module is activated when the tracking confidence is lower than the empirical threshold to prevent the drifting of main module model. The main module has fast-tracking speed and low accuracy, whereas the auxiliary module has slow tracking speed and high accuracy. These modules can complement each other in terms of speed and accuracy. The test results on the OTB2015 and VOT2016 data sets show that the algorithm has better accuracy and robustness than other correlation filtering algorithms.

Key words image processing; target tracking; correlation filtering; context; manifold regularization

OCIS codes 100.3008;100.4999

1 引言

视觉目标跟踪是计算机视觉的一个重要分支,

在智能视频交通监控、机器人、人机交互等领域得到了广泛的应用^[1]。根据目标外观模型将目标跟踪算法分为生成式跟踪算法与判别式跟踪算法^[2]。近年

收稿日期: 2020-09-01; 修回日期: 2020-09-20; 录用日期: 2020-09-30

基金项目: 国家自然科学基金(41701479)、辽宁省自然科学基金(20180550529)

* E-mail: 1932867699@qq.com

来,随着机器学习的兴起,判别式跟踪算法逐渐能同时利用目标信息和目标周围背景信息,得到了人们的广泛研究^[3]。

判别式跟踪算法也被称为检测跟踪算法,通过训练数据学习到一个分类模型,以区分前景和背景^[4]。以目标所在区域为正样本,背景区域为负样本进行分类器的训练,分类器预测的最高分数所在位置就是目标所在的位置。代表性的判别式跟踪算法有基于分类跟踪的深度学习算法和基于支持向量机(SVM)的跟踪算法^[5]。Bolme等^[6]将相关滤波应用于目标跟踪领域,利用快速傅里叶变换和卷积定理在频域中最小化期望响应和滤波器与目标区域的均方误差和,得到最小平方和误差(MOSSE)滤波器。MOSSE算法的跟踪速度每秒高达600多帧,虽然跟踪精度较低,但算法的实时性较高^[6]。Henriques等^[7]将循环矩阵及核技巧引入MOOSE算法中,利用灰度特征构建循环结构核(CSK)跟踪算法,提高了MOOSE算法的跟踪精度。Danelljan等^[8]利用颜色属性表示输入数据,自适应选择鲁棒的颜色特征CN进行跟踪。Henriques等^[9]提出了核相关滤波(KCF)算法,在CSK中引入不同的核方法以及方向梯度直方图(HOG)特征,提高了相关滤波视频目标跟踪算法的精度。Bertinetto等^[10]提出了Staple算法,将HOG特征和颜色直方图特征进行融合,在VOT2016数据集上获得了较好的跟踪效果。Qi等^[11]提出将深度卷积神经网络获取的特征用于相关滤波目标跟踪中,也获得了较好的跟踪效果。但提取卷积特征的计算开销较大,无法满足目标跟踪的实时性要求,如高效卷积算子(ECO)^[12]、连续卷积算子跟踪(CCOT)^[13]。Mueller等^[14]提出了上下文相关框架,解决了传统相关滤波算法背景信息丢失严重的问题,允许显式地将目标周围的上下文信息合并到相关滤波学习中。但上下文相关框架带来的正负样本不均衡问题,会导致学习模型的泛化能力下降。

为解决上下文相关框架过量学习目标背景信息和基于深度学习的跟踪算法速度慢的问题,本文提出了一种基于双核模型上下文的流形正则相关滤波跟踪算法。该算法包含主模块和辅助模块,主模块采用上下文流形正则相关滤波模型,可以很好地解决上下文相关框架过量学习目标背景信息的问题;辅助模块采用跟踪效果较好的KCF算法。主模块和辅助模块交替进行跟踪,减少了辅助模块深度特

征的提取帧数,只在相关帧数目标发生遮挡、形变或超出视距等问题时,辅助模块才会利用深度特征校正主模块的跟踪结果。本算法既能有效防止主模块模型发生漂移,又不会严重影响主模块的跟踪速度。

2 上下文相关跟踪

Mueller等^[14]提出的框架可显式地将目标周围的上下文信息合并到相关滤波学习中。在跟踪目标 P_0 周围提取出 k 个上下文相关图像块 P_i ,利用特定的特征提取算法建立 P_0 及 P_i 的外观模型,即 $P_0 \rightarrow X_0 \in \mathbf{R}^n, P_i \rightarrow X_i \in \mathbf{R}^n$ 。对 X_0 和 X_i 进行循环移位,建立正样本空间 x_0 和负样本空间 x_i 。负样本空间的样本标签用0表示,上下文相关跟踪旨在学习一个滤波器 w ,使模型对目标区域的评分尽量高,对目标周围区域的评分尽量低,需要优化的目标函数可表示为

$$\min_w \|x_0 w - y\|_2^2 + \lambda \|w\|_2^2 + \lambda_1 \|x_i w\|_2^2, \quad (1)$$

式中, y 为二维高斯矢量化回归目标, $\lambda, \lambda_1 > 0$ 为正项系数。将上下文相关正项和回归目标合并为

$$\min_w \|Aw - \bar{y}\|_2^2 + \lambda \|w\|_2^2, \quad (2)$$

式中, $A = \begin{bmatrix} x_0 \\ \sqrt{\lambda_1} x_1 \\ \vdots \\ \sqrt{\lambda_1} x_k \end{bmatrix}, \bar{y} = \begin{bmatrix} y \\ 0 \\ \vdots \\ 0 \end{bmatrix}, x_1, \dots, x_k$ 为 k 个负

样本特征向量。在傅里叶域中得到(2)式的闭式解为

$$\hat{w} = \frac{\hat{x}_0^* \odot \hat{y}}{\hat{x}_0^* \odot \hat{x}_0 + \lambda + \lambda_1 \sum_{i=1}^k \hat{x}_i^* \odot \hat{x}_i}, \quad (3)$$

式中, \hat{x}_0 为 x_0 的傅里叶变换, \hat{x}_i 为 x_i 的傅里叶变换, \hat{x}_0^* 为 \hat{x}_0 的共轭, \hat{x}_i^* 为 \hat{x}_i 的共轭, λ 为所有元素均为 λ 的矩阵。与传统的相关滤波跟踪算法相同,当新的图像块 z 出现时,将跟踪结果确定为最大响应 r 的位置,在频域中进行快速计算,得到

$$\hat{r} = \hat{z} \odot \hat{w}, \quad (4)$$

式中, \hat{r} 为频域内的相关响应图, \hat{w} 为频域内的位置滤波器。

3 基于双核模型上下文的流形正则相关滤波跟踪

主模块采用上下文流形正则相关滤波模型,特征采用HOG+CN的自适应融合。辅助模块采用

KCF 算法, 卷积层为 Visual Geometry Group (VGG) 实验室用 ImageNet 数据集训练的卷积神经网络 VGG-16 的卷积层 Conv4-3。跟踪置信度评价指标采用最大响应值与平均峰值相关能量 (APCE)^[15]。当主模块的最大响应值和 APCE 中的一个低于经验阈值时, 停止主模块模型的更新, 启

用辅助模块跟踪。利用深度特征强大的表征能力在目标发生遮挡、变形、超出视距等情况下完成跟踪, 有效防止主模块模型漂移。主模块与辅助模块协同作用, 使算法的跟踪精度高于基于手工特征的相关滤波算法, 速度也达到了实时跟踪的要求, 本算法的流程图如图 1 所示。

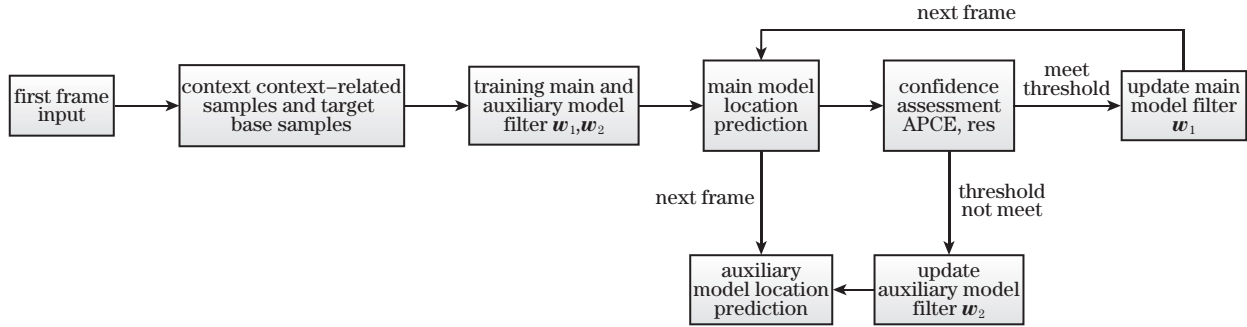


图 1 本算法的流程图

Fig. 1 Flow chart of our algorithm

3.1 上下文流形正则相关滤波

引入循环移位的近似密集取样方法可以解决 MOSSE 算法导致的样本冗余问题, 但相关滤波算法引入的循环矩阵也会带来边界效应。可通过添加余弦窗的方式缓解边界效应, 但余弦窗会过滤掉大量的背景信息。因此, 将上下文相关框架引入相关滤波算法中, 使模型学习到更多的背景信息。为防止模型过量学习背景信息, 提出了一种流形正则方法。即对上下文相关样本进行流形正则处理, 以达到惩罚上下文相关框架、优化主模块模型的目的。

3.1.1 上下文相关样本流形建模

采用拉普拉斯映射方法建立上下文相关样本的流形结构优化模型, 并将该模型作为流形正则项引入(1)式。将上下文样本作为目标基样本的增广样本, 上下文重叠采样方法如图 2 所示。首先, 给定跟

踪目标框, 并将其扩展到目标周围区域; 然后, 将扩展区域按照目标框尺寸分割成 4 个重叠的图像块, 对每个图像块进行循环移位, 产生一系列虚拟样本 $\{x_i \in \mathbf{R}^n\}_{i=1}^m$, 其中, m 为顶点数; 最后, 利用图像的拉普拉斯正则项, 模拟这些样本分布的局部流形结构。拉普拉斯正则项可以度量不同样本之间的相似性, 如果样本 x_i 和 x_j 在高维特征空间的内部几何流形结构相近, 则 x_i 和 x_j 的相关映射 $f(x_i)$ 和 $f(x_j)$ 也相近。因此, 利用虚拟样本 $\{x_i \in \mathbf{R}^n\}_{i=1}^m$ 构建一个具有 m 个顶点的最近邻无向图 G , 每个顶点表示一个样本。每个顶点 x_i 都与其最近邻顶点 $\{x_j\}$ 相连接。文献[16]用高斯核权重因子定义无向图 G 的带权邻接矩阵 W , 如果结点 i 与结点 j 相互连接, 则 W_{ij} 可表示为

$$W_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma}\right), \quad (5)$$

式中, $\sigma > 0$ 为控制权重 W_{ij} 下降速度的参数。 x_i 的度数 $d_i = \sum_{j=1}^m W_{ij}$, 则度数矩阵 $D = \text{diag}(d_1, \dots, d_m)$ 。为了将加权图 G 映射到一组表示系数上, 拉普拉斯特征映射的学习准则最小化目标可表示为

$$\sum_{i,j=1}^m [f(x_i) - f(x_j)]^2 W_{ij} = Y^T L Y, \quad (6)$$

式中, $Y = [f(x_1), \dots, f(x_m)]^T$, $L = D - W$ 为拉普拉斯矩阵。

3.1.2 上下文相关样本流形正则化

(1)式给出了上下文相关跟踪框架的优化目标函数, 其含义是在相关滤波基样本的基础上增加上

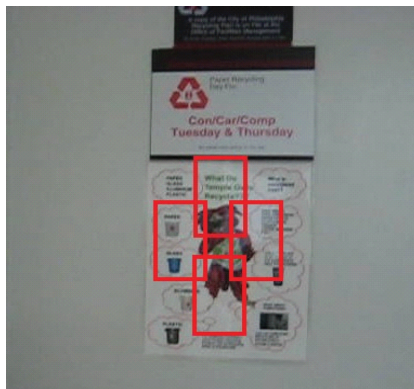


图 2 上下文重叠采样的示意图

Fig. 2 Schematic diagram of the context overlap sampling

下文相关样本,将目标基样本视为正样本,上下文相关样本视为负样本,最优结果是训练出一个滤波器 w 。但上下文相关样本的数量是目标基样本的 4 倍,优化模型的结构风险误差较大,训练出来的滤波器 w 泛化能力还有待提高。为平衡正负样本不均匀带来的模型结构误差,采用增加正则项的方法降低上下文相关样本的训练权重。给定 $p+1$ 个标签 y_i 为正的样本 $\{(x_i, y_i=1)\}_{i=0}^{p=0}$, $q-p$ 个标签 y_i 为负的样本 $\{(x_i, y_i=0)\}_{i=p+1}^{q=4}$ 。通过优化模型找到一个最优的线性分类器 $f^* \in H_k$, 优化模型可表示为

$$\min_{f^* \in H_k} \left\{ \sum_{i=0}^n [f(x_i) - y_i] + \lambda \|f\|_k^2 + \lambda_1 \sum_{i=n+1}^m f(x_i) + \lambda_2 Y^T L Y \right\}, \quad (7)$$

式中, H_k 为再生核希尔伯特空间, $\lambda, \lambda_1, \lambda_2 > 0$ 为正则项系数。为了快速求解(7)式,用一个线性函数表示回归目标,即

$$f(x) = w^T x. \quad (8)$$

将(8)式代入(7)式,得到

$$\min_w L(w) = \|Aw - \bar{y}\|_2^2 + \lambda \|w\|_2^2 + \lambda_2 w^T \bar{A}^T L \bar{A} w, \quad (9)$$

式中, $\bar{A} = \begin{bmatrix} x_0 \\ \vdots \\ x_k \end{bmatrix}$ 。令 $\frac{\partial L(w)}{\partial w} = 0$, 则上下文流形正则

相关滤波模型在傅里叶域的解为

$$\hat{w} = \frac{\hat{x}_0^* \odot \hat{y}^*}{\hat{x}_0} \odot \hat{x}_0 + \lambda_1 \sum_{l=1}^k \hat{x}_l^* \odot \hat{x}_l + \lambda + \lambda_2 \sum_{i,j=0}^k W_{ij} (\hat{x}_i - \hat{x}_j)^* \odot (\hat{x}_i - \hat{x}_j), \quad (10)$$

式中, $k=4$ 为上下文相关样本的个数。

3.1.3 快速检测

上下文流形正则相关滤波基于判别式相关滤波跟踪框架,给定一个具有 D 通道特征表示 $\{z^d\}_{d=1}^D$ 的新图像块 z , 其相关响应矩阵 r 可表示为

$$r(z) = \mathcal{F}^{-1} \left(\sum_{d=1}^D \hat{z}^d \odot \hat{w}^d \right), \quad (11)$$

式中, \mathcal{F}^{-1} 为傅里叶逆变换, \hat{z}^d 为新图像块 z 频域内 d 通道的特征, \hat{w}^d 为频域内第 d 通道的分类器参数。

3.1.4 HOG 与 CN 的自适应融合

KCF^[9] 算法采用多通道 HOG 特征,能很好地适应光照变化、平面内旋转等情况。颜色特征则是从全局角度对目标进行描述,具有旋转不变

性,且不易受目标尺度或形状变化的影响。CN 特征是颜色特征的一种,通过概率映射方式将图像从原始的三维 RGB (Red, Green, Blue) 空间转换到 11 维的颜色特征空间。相比其他颜色特征, CN 特征的目标描述能力更强。因此,利用现有的跟踪置信度评判标准对 HOG 和 CN 特征在响应层进行融合。文献[6]提出用峰值旁瓣比 R_{ps} 衡量目标跟踪的置信度。相关滤波峰值 x 处的 R_{ps} 为

$$R_{ps}(x) = \frac{\max(r) - \mu(r)}{\sigma(r)}, \quad (12)$$

式中, $\max(r)$ 为相关响应矩阵 r 的最大响应值, $\mu(r)$ 为 r 的均值, $\sigma(r)$ 为 r 的标准差。文献[17]用相关滤波响应一致性 C_{CFR} 衡量目标跟踪的置信度,可表示为

$$C_{CFR} = \|\hat{f}_t(u, v) - \hat{f}_{t-1}(u + \Delta u, v + \Delta v)\|_2^2, \quad (13)$$

式中, $\hat{f}_t(u, v)$ 和 $\hat{f}_{t-1}(u + \Delta u, v + \Delta v)$ 分别为目标第 t 帧和第 $t-1$ 帧时傅里叶域的相关滤波响应, $(\Delta u, \Delta v)$ 为相邻两帧图像目标位置的相对位移。 R_{ps} 越大,表明目标跟踪的置信度越高; C_{CFR} 越小,表明目标跟踪的置信度越高。将 R_{ps} 和 C_{CFR} 线性组合成跟踪置信度函数 $f(R_{ps}, C_{CFR})$, 可表示为

$$f(R_{ps}, C_{CFR}) = (1 - \rho) R_{ps} + \rho \frac{1}{C_{CFR} + \epsilon}, \quad (14)$$

式中, ρ 为 R_{ps} 和 C_{CFR} 的权重调节系数, $\epsilon = 0.01$ 为防止分母为 0 的参数。则 HOG 特征的跟踪置信度 f_{HOG} 可表示为

$$f_{HOG} = (1 - \rho) R_{ps, HOG} + \rho \frac{1}{C_{CFR, HOG} + \epsilon}, \quad (15)$$

同理,得到 CN 特征的跟踪置信度 f_{CN} 为

$$f_{CN} = (1 - \rho) R_{ps, CN} + \rho \frac{1}{C_{CFR, CN} + \epsilon}. \quad (16)$$

将 f_{HOG} 和 f_{CN} 作为特征融合的权重因子,得到融合后的相关滤波响应为

$$\hat{f}_{final} = \frac{f_{CN}}{f_{CN} + f_{HOG}} \hat{f}_{CN} + \frac{f_{HOG}}{f_{CN} + f_{HOG}} \hat{f}_{HOG}, \quad (17)$$

式中, \hat{f}_{final} 为融合后的频域相关滤波响应, \hat{f}_{HOG} 和 \hat{f}_{CN} 分别为 HOG 特征和 CN 特征对应的频域相关滤波响应。

3.2 核相关滤波

辅助模块采用的 KCF^[9] 目标跟踪算法是一种判别式跟踪算法,算法在第一帧图像目标所在区域训练一个目标检测器,用于预测下一帧图像的目标所在位置,再由新的预测结果更新目标检测器。在

训练滤波器模板时选取的目标区域为正样本,对目标区域进行循环移位获得负样本,构造循环矩阵。利用傅里叶变换与循环矩阵的相关性质,大大缩短了训练滤波器所需的时间。

3.2.1 岭回归

KCF 算法使用岭回归训练分类器,原因是该分类器包含一个简单的闭式解,且可以实现与 SVM 等复杂方法相近的性能。训练分类器的目的是找到一个类似 $f(\mathbf{X}) = \mathbf{w}^T \mathbf{X}$ 的函数,使样本 x_i 与其回归目标 y_i 的平方误差最小,同时加入正则化项防止分类器过拟合,可表示为

$$\min_{\mathbf{w}} \sum_i [f(\mathbf{x}_i) - y_i]^2 + \lambda_0 \|\mathbf{w}\|^2, \quad (18)$$

式中, y_i 为样本类别, λ_0 为正则项系数,可用矩阵表示为

$$\min_{\mathbf{w}} \sum (\mathbf{w}^T \mathbf{X} - \mathbf{Y})^2 + \lambda_0 \|\mathbf{w}\|^2, \quad (19)$$

式中, \mathbf{X} 为样本集合, \mathbf{Y} 为样本类别向量。

3.2.2 核技巧及快速检测

KCF 算法将核技巧引入目标检测器的训练中,避免了因样本当前维线性不可分而无法训练出好的目标检测器的问题,提高了算法的鲁棒性。其中,核岭回归分类函数可表示为

$$f(\boldsymbol{\alpha}) = \mathbf{w}^T \boldsymbol{\alpha} = \sum_{i=1}^n \alpha_i k(\boldsymbol{\alpha}, \mathbf{x}_i), \quad (20)$$

式中, α_i 为训练样本 \mathbf{x}_i 对应的系数, $k(\boldsymbol{\alpha}, \mathbf{x}_i)$ 为高斯核函数, $\boldsymbol{\alpha}$ 为需要分类的图像块。根据文献[9]得到 $\boldsymbol{\alpha}$ 在频域上的解为

$$\hat{\boldsymbol{\alpha}} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}} + \lambda_0}, \quad (21)$$

式中, $\hat{\mathbf{k}}$ 为训练样本生成的核矩阵, $\hat{\boldsymbol{\alpha}}$ 为系数矩阵, λ_0 为正则项系数。为了检测目标位置, KCF 算法使用相关滤波器在 \mathbf{z} 上进行卷积运算,可表示为

$$f(\mathbf{z}) = \mathcal{F}^{-1}(\hat{\mathbf{k}}_{\mathbf{z}\mathbf{x}} \odot \hat{\boldsymbol{\alpha}}), \quad (22)$$

式中, $\hat{\mathbf{k}}_{\mathbf{z}\mathbf{x}} = \hat{\mathbf{k}}(\mathbf{z}, \mathbf{x})$ 。计算结果中响应值最大的位置即为目标位置。

3.2.3 深度特征的选择

Image-VGG-16^[18] 是 VGG 实验室用 ImageNet 数据集训练的卷积神经网络,该网络使用卷积核提取具有语义信息的卷积特征。随着卷积层数的加深,卷积特征中的语义信息也越来越丰富。池化层则会压缩特征图,加快计算速度并防止过拟合,但也会降低卷积特征的空间分辨率。VGG-16 不同层特征的特点不同,低层 Conv4-3 特征擅长区分不同类

别的物体,对目标的形变和遮挡鲁棒;高层 Conv5-4 特征更关注目标的局部细节,可用来区分背景中与目标相似的错误选择。通过对特征层 Conv2-2、Conv3-3、Conv4-3、Conv5-4 的实验验证,并综合考虑了算法的实时性和精度,选择 Conv4-3 作为辅助模块的特征层。

3.3 模型更新

3.3.1 主模块模型更新

当跟踪置信度指标高于经验阈值时,对主模块模型在频域上进行更新。由于主模块并未使用核技巧,因此,直接对滤波器 \mathbf{w} 进行更新,即用新的权重系数 $\hat{\mathbf{w}}$ 更新旧的权重系数 \mathbf{w}^{t-1} ,可表示为

$$\hat{\mathbf{w}}_{\text{HOG}}^t = (1 - \beta) \hat{\mathbf{w}}_{\text{HOG}}^{t-1} + \beta \hat{\mathbf{w}}_{\text{HOG}}^t, \quad (23)$$

$$\hat{\mathbf{w}}_{\text{CN}}^t = (1 - \beta) \hat{\mathbf{w}}_{\text{CN}}^{t-1} + \beta \hat{\mathbf{w}}_{\text{CN}}^t, \quad (24)$$

式中, β 为主模块模型的学习率, $\hat{\mathbf{w}}_{\text{HOG}}^t$ 为频域内第 t 帧学习到的关于 HOG 特征的权重系数, $\hat{\mathbf{w}}_{\text{CN}}^t$ 为频域内第 t 帧学习到的关于 CN 特征的权重系数, $\hat{\mathbf{w}}_{\text{HOG}}^t$ 为第 t 帧更新完成后关于 HOG 特征的权重系数, $\hat{\mathbf{w}}_{\text{CN}}^t$ 为第 t 帧更新完成后关于 CN 特征的权重系数。

3.3.2 辅助模块模型更新

当跟踪置信度指标低于经验阈值时,利用满足置信度条件的前一帧跟踪结果对辅助模块模型进行更新,可表示为

$$\begin{aligned} \hat{\boldsymbol{\alpha}}^t &= (1 - \eta) \hat{\boldsymbol{\alpha}}^{t-1} + \eta \hat{\boldsymbol{\alpha}}^t, \\ \hat{\mathbf{x}}^t &= (1 - \eta) \hat{\mathbf{x}}^{t-1} + \eta \hat{\mathbf{x}}^t, \end{aligned} \quad (25)$$

式中, $\hat{\boldsymbol{\alpha}}^t$ 、 $\hat{\boldsymbol{\alpha}}^{t-1}$ 分别为当前帧和上一帧频域中高维空间系数向量, $\hat{\mathbf{x}}^t$ 、 $\hat{\mathbf{x}}^{t-1}$ 分别为频域中当前帧和上一帧的目标外观模型, η 为辅助模块模型的学习率。

3.4 跟踪置信度指标

跟踪过程中如果没有干扰且跟踪准确时,相关滤波响应图大多出现单一明显的峰值;当跟踪过程存在干扰时会出现多个峰值。大量实验结果表明,根据相关滤波响应图的波动能有效预测目标的遮挡情况。图 3(a)为正常情况下的跟踪实例,跟踪目标为图中的玩具老虎,图 3(c)为其对应的相关滤波响应图;图 3(b)为目标发生遮挡情况下的跟踪实例,图 3(d)为其对应的相关滤波响应图。可以发现,跟踪正常时,响应图接近于理想的二维高斯分布,目标中心有明显的峰值,目标周围分布平缓;当跟踪目标发生遮挡时,响应图会发生剧烈振荡,不再有明显的峰值。其中, z 轴为响应值, u, v 轴为响应值对应像素点的坐标。

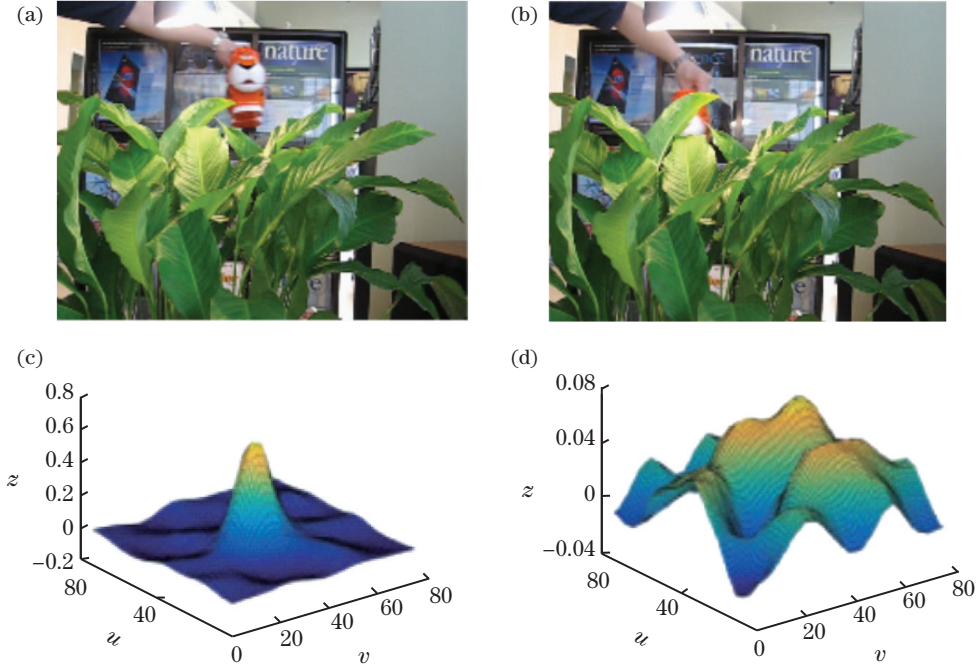


图 3 跟踪实例及其对应的响应图。(a)正常跟踪；(b)发生遮挡；(c)正常跟踪时的响应图；(d)发生遮挡时的响应图

Fig. 3 Tracking instance and its corresponding response graph. (a) Normal tracking; (b) occlusion occurs; (c) response graph when normal tracking; (d) response graph when occlusion occurs

文献[15]提出用 APCE 反映响应图的振荡程度, 图 3(b)中目标被遮挡或目标丢失情况中, $X_{APCE} = 3.2174$, 而图 3(a)中, $X_{APCE} = 41.0842$, 此时, 不更新模型, 以避免模型发生漂移。目标的剧烈形变和旋转也会引起响应图的波动, 因此, 综合考虑响应图的峰值和波动情况, 定义的置信度评价指标为

$$X_{\text{index}} = [X_{APCE}, \max(f_{\text{final}})], \quad (26)$$

式中, f_{final} 为融合响应图, $\max(f_{\text{final}})$ 为 f_{final} 的最大值, X_{APCE} 可表示为

$$X_{APCE} = \frac{f_{\text{max}} - f_{\text{min}}}{X_{\text{mean}} \left[\left(\sum_{u,v} f_{u,v} - f_{\text{min}} \right)^2 \right]}, \quad (27)$$

式中, f_{max} 、 f_{min} 和 $f_{u,v}$ 分别为融合响应图的最大响应值, 最小响应值和坐标 (u, v) 处的响应值, $\text{mean} \left[\left(\sum_{u,v} f_{u,v} - f_{\text{min}} \right)^2 \right]$ 为对 $\sum_{u,v} (f_{u,v} - f_{\text{min}})^2$ 求平均值。置信度经验阈值根据视频序列的跟踪情况设定, 如果将经验阈值设定为一个常量, 无法满足不同跟踪场景, 因此, 将经验阈值设定为跟踪置信度指标历史平均值的倍数, 可表示为

$$X_{APCE}^r > \beta_1 \frac{1}{r-1} \sum_{s=1}^{r-1} X_{APCE}^s, \quad (28)$$

$$\max(f_{\text{final}}^r) > \beta_2 \frac{1}{r-1} \sum_{s=1}^{r-1} \max(f_{\text{final}}^s)$$

式中, X_{APCE}^r 为第 r 帧响应图的 X_{APCE} , $\max(f_{\text{final}}^s)$

为第 s 帧响应图的最大值, f_{final}^r 为第 r 帧的融合响应图, β_1 和 β_2 为大于 0 小于 1 的系数, 实验中 $\beta_1 = 0.45$, $\beta_2 = 0.7$ 。 β_1 作为跟踪置信度指标 APCE 的调节系数, 其值参考文献[15]给出。相关响应的最大响应值作为一个相关滤波算法的常用跟踪置信度评价指标, 其调节系数 β_2 是一个经验值。当置信度满足(27)式时, 对主模块模型进行更新; 否则, 更新辅助模块模型, 通过辅助模块预测目标位置。

4 实验设计

在公开数据集 OTB2015 和 VOT2016 上, 从三个方面评估本算法的性能: 1) 辅助模块的有效性; 2) 与其他相关滤波算法跟踪性能和速度的对比; 3) 与当前基于深度卷积特征的跟踪算法的对比。

4.1 实验环境及算法参数

实验的硬件环境: CPU 为 Inter(R) Core(TM) i7-4790, 主频为 3.60 GHz, RAM 为 16 GB, 显卡为 NVIDIA GeForce GTX960; 操作系统为 64 位 Windows10; 软件环境为 Matlab R2016a, 网络为已经在 ImageNet 上训练好的 VGG-16, 本算法的主模块初始化参数如表 1 所示, 辅助模块初始化参数如表 2 所示。

表 1 主模块的初始化参数

Table 1 Initialization parameters of the main module

Initialization parameter	Value
λ	0.01
λ_1	0.3
λ_2	0.03
ρ	0.5
β	0.075

表 2 辅助模块的初始化参数

Table 2 Initialization parameters of auxiliary module

Initialization parameter	Value
λ_0	0.01
η	0.075
Kernel function	0.5

表 1 中滤波器正则项系数 λ 和上下文样本正则项系数 λ_1 参考文献[13]设置。HOG+CN 特征融合调节系数 ρ 参考文献[17]确定。文献[17]在数据集 OTB2015 上的实验结果表明,当 $\rho = 0.5$ 时, HOG+CN 特征的跟踪效果最佳。主模块模型学习率 β 则是一个经验值,在相关滤波跟踪算法中,一般为 $0.015 \sim 0.1$ 。跟踪模型越好、数据集难度越低,学习率越向 0.1 靠拢。大部分相关滤波算法在 OTB2015 数据集上采用的模型学习率为折中值 0.075,如 KCF^[9]、Staple 算法^[10]。流形正则项系数 λ_2 通过数据集 OTB2015 的调参实验确定,实验结果如表 3 所示,其中,DP 为距离精度, AUC 为算法成功率曲线与横轴之间的面积曲线。可以发现,当 $\lambda_2 = 0.03$ 时,算法的跟踪效果最佳。表 2 中的核相关滤波器正则项系数 λ_0 、模型学习率 η 、高斯核函

数的下降参数 Kernel function 均参考文献[9]中的 KCF 算法。

表 3 调参实验的结果

Table 3 Results of the tuning experiment

λ_2	DP/%	AUC/%
0.001	76.1	55.6
0.01	79.9	59.5
0.03	83.7	61.4
0.05	81.0	59.8
0.07	79.8	59.8
0.10	79.0	58.5

4.2 辅助模块的有效性验证

为验证辅助模块的有效性,将主模块与主模块+辅助模块在 OTB2015 数据集上进行测试实验。同时采用 VGG-16 网络的特征层 Conv5-4、Conv4-3、Conv3-3、Conv2-2 作为辅助模块进行对比。在 OTB2015 数据集上用一次成功率(OPE)与重叠成功率对结果进行评估,评估过程中用 pixel 为 20(横向或纵向相邻像素点之间的距离为 1)的 DP 作为精度评估指标,用 AUC 作为算法的成功率指标,在公开数据集 OTB2015 上的 OPE 如图 4 所示,其中,中括号中的数值分别为平均 DP 和 AUC。可以发现,添加辅助模块的网络跟踪效果明显比单独的主模块更稳定。且 VGG-16 的特征层 Conv5-4 和 Conv4-3 均取得了良好的跟踪效果,但由于网络深度的增加,Conv5-4 的相关性计算和模型更新时间的复杂度较高;且 Conv5-4 每秒能跟踪的图像帧数(FPS)仅为 5 frame,而 Conv4-3 的 FPS 能达到 17 frame,因此,实验选取 VGG-16 的 Conv4-3 作为辅助模型。

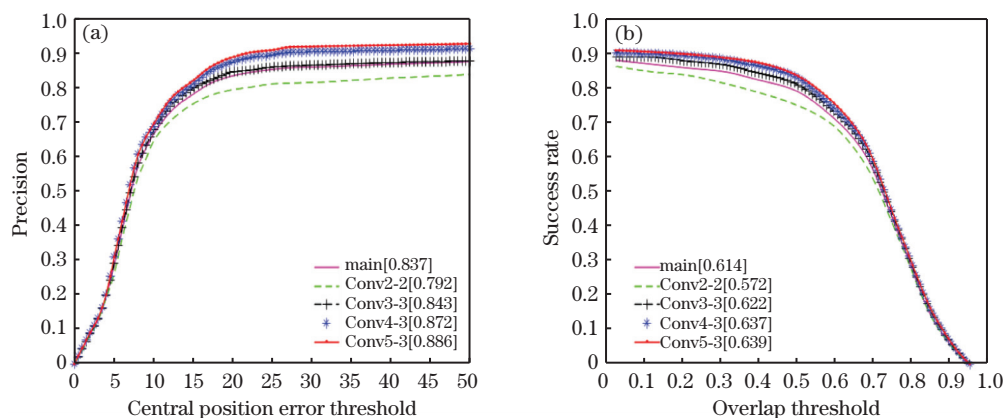


图 4 辅助模块及特征选取的对比验证。(a) DP; (b) OPE

Fig. 4 Comparison and verification of auxiliary module and feature selection. (a) DP; (b) OPE

确定辅助模块和深度特征层后,在 OTB2015 数据集上进行模块独立实验,验证主模块和辅助模块之间速度与精度的互补性,结果如表 4 所示。可

以发现,未进行流形正则处理的上下文相关滤波(CACF)算法^[13]的 DP 和 AUC 均低于主模块。辅助模块在 OTB2015 数据集上的 DP 和 AUC 都取得

了不错的结果,但平均 FPS 仅为 2.7 frame,实时性较差。主模块的 DP 和 AUC 比辅助模块稍低,但平均 FPS 为 48 frame。主模块和辅助模块相结合后,平均 FPS 为 17 frame,且 DP 和 AUC 相比主模块均有所提升。

表 4 模块的独立性验证结果

Table 4 Independence verification results of modules

Module	Feature	DP/%	AUC/%	FPS/frame
CACF	HOG+CN	76.1	58.3	54.0
Main module	HOG+CN	83.7	61.4	48.3
Auxiliary module	Conv4-3	90.4	65.1	2.7
Main+auxiliary	HOG+CN, Conv4-3	87.2	63.7	17.0

4.3 算法的总体水平评估

在 OTB2015 和 VOT2016 两个数据集上对比本算法与跟踪性及实时性较好的相关滤波跟踪

Staple 算法^[10]、最小化跟踪方法 (MEEM)^[19]、判别尺度空间跟踪 (DSST) 算法^[20]、SRDCF (Spatially regularized discriminative correlation filters) 算法^[21]、KCF^[9] 及基于深度神经网络的 LSART (Learning spatial-aware regressions for visual tracking) 算法^[22]、高效卷积算子 (ECO) 算法^[12]、基于连续域卷积操作跟踪 (CCOT) 算法^[13] 的性能。

4.3.1 OTB2015 数据集

OTB2015 数据集包含了 100 个视频序列及 11 种不同的视频属性。11 个属性分别为光照变化 (IV)、尺度变化 (SV)、遮挡 (OCC)、形变 (DEF)、运动模糊 (MB)、快速运动 (FM)、平面内旋转 (IPR)、平面外旋转 (OPR)、超出视野 (OV)、背景扰乱 (BC) 和低分辨率 (LR)。不同算法在 OTB2015 数据集上的跟踪性能如图 5 所示,可以发现,相比基于手工特征的 KCF 算法,本算法的 DP 提高了 27.7%,AUC 提高了 16%。

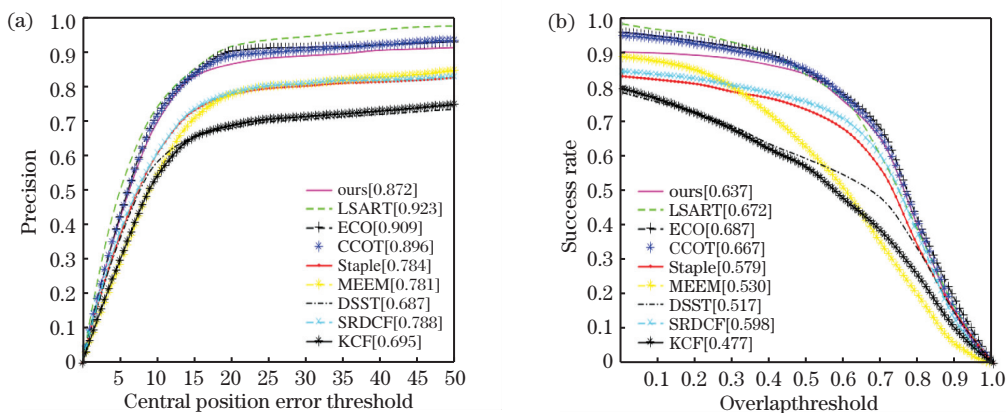


图 5 不同算法在 OTB2015 数据集的跟踪结果。(a) DP; (b) OPE

Fig. 5 Tracking results of different algorithms on the OTB2015 data set. (a) DP; (b) OPE

本算法与满足实时性要求的跟踪算法的对比结果如表 5 所示。为保证实验的可靠性,将本算法与近几年基于深度卷积神经网络的算法也进行了对比,结果如表 6 所示。可以发现,相比基于相关滤波的跟踪算法和基于深度卷积网络的算法,本算法的总体跟踪性能均更优。

表 5 本算法与实时性较好的跟踪算法的对比

Table 5 Comparison between our algorithm and the tracking algorithms with better real-time performance

Algorithm	DP/%	AUC/%	FPS/frame
Staple	78.4	57.9	57.3
MEEM	78.1	53.0	16.8
DSST	68.7	51.7	20.1
SRDCF	78.8	59.8	11.5
KCF	69.5	47.7	114.0
Ours	87.2	63.7	17.5

表 6 本算法与基于深度卷积神经网络的算法的对比

Table 6 Comparison between our algorithm and the algorithms based on deep convolutional neural network

Algorithm	DP/%	AUC/%	FPS/frame
LSART	92.3	67.2	0.1
ECO	90.9	68.7	1.4
CCOT	89.6	66.7	0.3
Ours	87.2	63.7	17.0

由于本算法的主模块部分结合了上下文相关框架与相关滤波算法,并通过对上下文相关样本的流形正则降低了模型的结构风险,对属性 MB、FM、BC 具有很好的鲁棒性,HOG 与 CN 特征的融合使算法对 IV、DEF、IPR 具有很好的鲁棒性,辅助模块的加入对 OCC、OV、OPR 具有很好的校正效果。本算法在 BC、DEF、OCC、SV、OPR、OV 六个具有

挑战性的视屏属性集中的 OPE 如图 6 所示。可以发现,本算法在 BC、DEF、OCC、SV、OPR、OV 视频属性集上的表现均超过了基于相关滤波的 Staple、

SRDCF 等算法。跟踪效果上,本算法与基于深度卷积神经网络的 ECO、CCOT、LSART 算法相当,且本算法的跟踪速度更快。

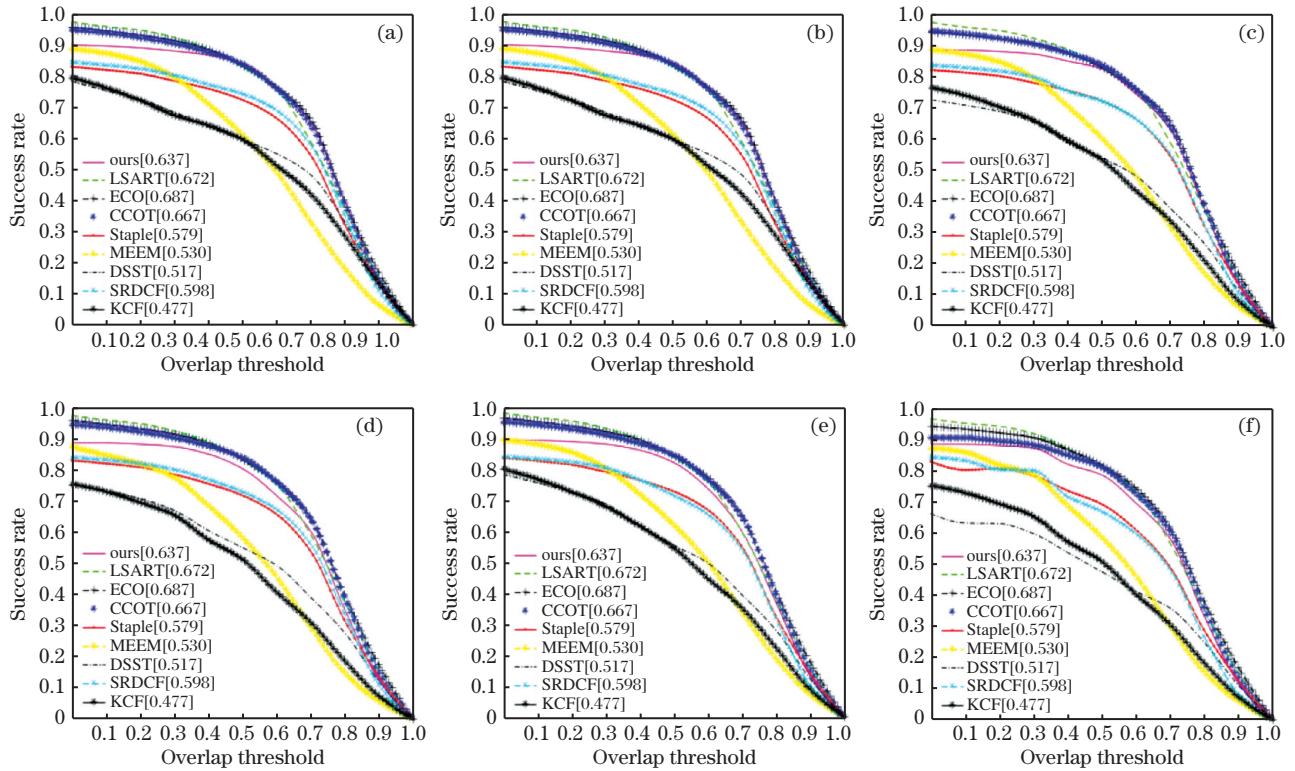


图 6 不同算法在不同场景上 OPE。(a) BC; (b) DEF; (c) OCC; (d) SV; (e) OPR; (f) OV

Fig. 6 OPE of different algorithms in different scenarios. (a) BC; (b) DEF; (c) OCC; (d) SV; (e) OPR; (f) OV

4.3.2 VOT2016 数据集

在 VOT2016 数据集上对比本算法与 EBT^[23]、CCOT、Staple、Struck^[24]、CSR-DCF^[25]、SRDCF、KCF 算法的跟踪性能,并用期望平均覆盖率(EAO)、准确率(A)、鲁棒性(R)评估算法的性

能,结果如表 7 所示。可以发现,本算法的 EAO 比 Staple 算法高 0.032,比 CCOT 算法低 0.004。总体性能高于其他基于相关滤波的跟踪算法,与基于深度卷积神经网络的 CCOT 算法也没有明显差距。

表 7 不同算法在 VOT2016 数据集上的跟踪效果

Table 7 Tracking effects of different algorithms on the VOT2016 data set

Tracker	EBT	CCOT	Staple	Struck	CSR-DCF	SRDCF	KCF	Ours
EAO	0.291	0.331	0.295	0.142	0.338	0.247	0.192	0.327
A	0.441	0.539	0.545	0.439	0.510	0.535	0.491	0.522
R	0.920	0.238	1.350	3.370	0.850	1.500	2.010	0.814

4.3.3 定性分析

本算法与 KCF、Staple、MEEM 算法在 6 个视频序列 Biker、Bird1、ClifBar、Ironman、Soccer、Swinnings 上的跟踪结果如图 7 所示。可以发现,这 6 个视频序列出现了 FM、DEF、OCC、OV、BC 等问题,其跟踪具有一定的挑战性。对 Biker 的跟踪过程中,第 70 帧图像由于目标的 FM 且伴随着 MB 和 IPR,导致 KCF、Staple、MEEM 算法跟踪失败。对 Bird1 的跟踪过程中,1~97 帧图

像 DEF 导致的误差累积使 KCF 算法的模型发生漂移,后续视频帧全部跟踪失败,其他算法则跟踪正常。97 帧图像后,视频出现 OV 问题,本算法虽然在出现 OV 的几帧图像中跟踪失败,但在辅助模块的校正下,主模块模型并未发生漂移,并在 OV 结束后成功找回目标。但其他算法的模型均发生了漂移,导致跟踪失败。对 ClifBar 和 Swinnings 的跟踪过程中,由于只出现了 IPR 和 MB 两个主要问题,除了 KCF 算法外,其他算法

都能很好地完成跟踪。Ironman、Soccer 两个视频序列由于出现了不同程度的 OCC、BC、SV，跟踪

难度较大，除本算法外，其他算法都出现了不同程度的目标丢失。

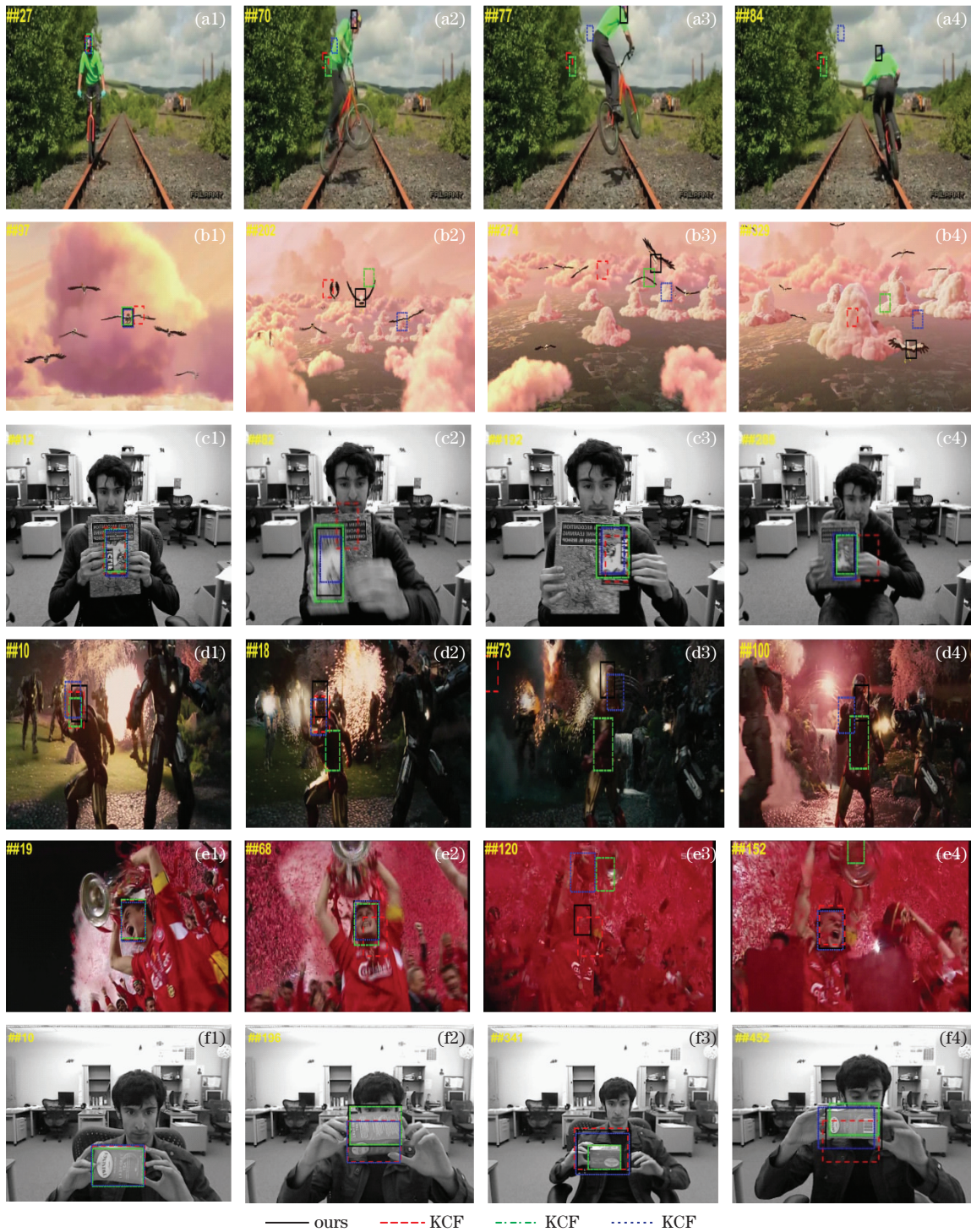


图 7 不同算法在 6 个视频序列上的跟踪结果。(a) Biker; (b) Bird1; (c) ClifBar; (d) Ironman; (e) Soccer; (f) Swinning

Fig. 7 Tracking results of different algorithms on 6 video sequences. (a) Biker; (b) Bird1;

(c) ClifBar; (d) Ironman; (e) Soccer; (f) Swinning

5 结 论

提出了一种基于双核模型上下文的流形正则相关滤波算法,将相关滤波的速度优势与深度特征的

精度优势相结合,使跟踪器同时满足跟踪精度与实时性的要求。主模块用上下文流形正则相关滤波模型,辅助模块则使用 KCF 算法。如果将辅助模型 KCF 算法改为上下文流形正则相关滤波算法,在跟

踪置信度较低时,会取得更好的跟踪效果。在 OTB2015 和 VOT2016 数据集上的实验结果表明,本算法在跟踪精度方面超过了基于相关滤波的跟踪算法,在跟踪速度方面优于基于深度卷积神经网络的跟踪算法,达到了主模型手工特征与辅助模型卷积特征优势互补的目的。

参 考 文 献

- [1] Hu S, Ge Y N, Han J L, et al. Object tracking algorithm based on dual color feature fusion with dimension reduction[J]. *Sensors*, 2018, 19(1): 73.
- [2] Tang X M, Chen Z G, Fu Y. Action prediction and scale adaptive target tracking algorithm[J]. *Laser & Optoelectronics Progress*, 2020, 57(8): 081014.
汤学猛, 陈志国, 傅毅. 行动预测及尺度自适应目标跟踪算法[J]. *激光与光电子学进展*, 2020, 57(8): 081014.
- [3] Ning X, Li W J, Tian W J, et al. Adaptive template update of discriminant KCF for visual tracking[J]. *CAAI Transactions on Intelligent Systems*, 2019, 14(1): 121-126.
宁欣, 李卫军, 田伟娟, 等. 一种自适应模板更新的判别式 KCF 跟踪方法[J]. *智能系统学报*, 2019, 14(1): 121-126.
- [4] Luo H L, Shi W. Adaptive weighted object tracking algorithm with continuous convolution operator[J]. *Journal of Image and Graphics*, 2019, 24(7): 1106-1115.
罗会兰, 石武. 结合连续卷积算子的自适应加权目标跟踪算法[J]. *中国图象图形学报*, 2019, 24(7): 1106-1115.
- [5] Shen Y L, Wu Z D, Zhao R J, et al. Long-term object tracking based on model updating and fast re-detection[J]. *Acta Optica Sinica*, 2020, 40(3): 0315002.
沈玉玲, 伍忠东, 赵汝进, 等. 基于模型更新与快速重检测的长时目标跟踪[J]. *光学学报*, 2020, 40(3): 0315002.
- [6] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 2544-2550.
- [7] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[M]//Fitzgibbon A, Lazebnik S, Perona P, et al. *Computer vision-ECCV 2012*. Lecture notes in computer science. Cham: Springer, 2012, 7575: 702-715.
- [8] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1090-1097.
- [9] 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.
- [10] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1401-1409.
- [11] Qi Y K, Zhang S P, Qin L, et al. Hedged deep tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4303-4311.
- [12] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6931-6939.
- [13] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision - ECCV 2016*. Lecture notes in computer science. Cham: Springer, 2016, 9909: 472-488.
- [14] Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1387-1395.
- [15] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4800-4808.
- [16] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[M]//Dietterich T G, Becker S, Ghahramani Z. *Advances in neural information processing systems 14*. Cambridge: The MIT Press, 2002: 585-591.
- [17] Chen F L, Ding Q H, Chang Z, et al. Multi-scale kernel correlation filter algorithm for visual tracking based on the fusion of adaptive features[J]. *Acta*

- Optica Sinica, 2020, 40(3): 0315001.
- 陈法领, 丁庆海, 常铮, 等. 自适应特征融合的多尺度核相关滤波目标跟踪 [J]. 光学学报, 2020, 40(3): 0315001.
- [18] Nguyen K, Fookes C, Sridharan S. Improving deep convolutional neural networks with unsupervised feature learning[C]//2015 IEEE International Conference on Image Processing (ICIP), September 27-30, 2015, QC, Canada. New York: IEEE Press, 2015: 2270-2274.
- [19] Zhang J M, Ma S G, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization[M]//Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8694: 188-203.
- [20] Danelljan M, Häger G, Khan F S, et al. Accurate scale estimation for robust visual tracking [C] // Proceedings of the British Machine Vision Conference 2014, September, 2014, Nottingham. British: BMVA Press, 2014: 65.1-65.11.
- [21] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4310-4318.
- [22] Sun C, Wang D, Lu H C, et al. Learning spatial-aware regressions for visual tracking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8962-8970.
- [23] Zhu G, Porikli F, Li H D. Beyond local search: tracking objects everywhere with instance-specific proposals[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 943-951.
- [24] Hare S, Golodetz S, Saffari A, et al. Struck: structured output tracking with kernels [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2096-2109.
- [25] Lukežić A, Vojir T, Zajc L C, et al. Discriminative correlation filter with channel and spatial reliability [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 4847-4856.