

基于多层卷积特征的自适应决策融合目标跟踪算法

孙彦景 石韞开 云霄* 朱绪冉 王赛楠

(中国矿业大学信息与控制工程学院 徐州 221116)

摘要: 针对目标快速运动、遮挡等复杂视频场景中目标跟踪鲁棒性差和跟踪精度低的问题, 该文提出一种基于多层卷积特征的自适应决策融合目标跟踪算法(ASF-TT)。首先提取卷积神经网络(CNN)中帧图像的多层卷积特征, 避免网络单层特征表征目标信息不全面的缺陷, 增强算法的泛化能力; 使用多层特征计算帧图像相关性响应, 提高算法的跟踪精度; 最后该文使用自适应决策融合算法将所有响应中目标位置决策动态融合以定位目标, 融合算法综合考虑生成响应的各跟踪器的历史决策信息和当前决策信息, 以保证算法的鲁棒性。采用标准数据集 OTB2013对该文算法和6种当前主流跟踪算法进行了仿真对比, 结果表明该文算法具有更加优秀的跟踪性能。

关键词: 目标跟踪; 卷积神经网络; 相关性响应; 决策融合

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2019)10-2464-07

DOI: 10.11999/JEIT180971

Adaptive Strategy Fusion Target Tracking Based on Multi-layer Convolutional Features

SUN Yanjing SHI Yunkai YUN Xiao ZHU Xuran WANG Sainan

(School of Information and Control Engineering, China University of Mining Technology, Xuzhou 221116, China)

Abstract: To solve the problems of low robustness and tracking accuracy in target tracking when interference factors occur such as target fast motion and occlusion in complex video scenes, an Adaptive Strategy Fusion Target Tracking algorithm (ASF-TT) is proposed based on multi-layer convolutional features. Firstly, the multi-layer convolutional features of frame images in Convolutional Neural Network(CNN) are extracted, which avoids the defect that the target information of the network is not comprehensive enough, so as to increase the generalization ability of the algorithm. Secondly, in order to improve the tracking accuracy of the algorithm, the multi-layer features are performed to calculate the correlation responses, which improves the tracking accuracy. Finally, the target position strategy in all responses are dynamically merged to locate the target through the adaptive strategy fusion algorithm in this paper. It comprehensively considers the historical strategy information and current strategy information of each responsive tracker to ensure the robustness. Experiments performed on the OTB2013 evaluation benchmark show that that the performance of the proposed algorithm are better than those of the other six state-of-the-art methods.

Key words: Target tracking; Convolutional Neural Network(CNN); Correlation response; Strategy fusion

1 引言

目标跟踪是在视频序列中根据第1帧的目标状态信息估计其后续状态的过程^[1], 其在机器人技术、智能监控和人机交互等领域有广泛应用。目标

跟踪近年来取得较大发展, 但由于形变, 快速运动和遮挡等因素影响, 其仍是计算机视觉领域研究的难点。

随着卷积神经网络(Convolutional Neural Net-

收稿日期: 2018-10-17; 改回日期: 2019-02-26; 网络出版: 2019-03-16

*通信作者: 云霄 yxztong@163.com

基金项目: 江苏省自然科学基金青年基金(BK20180640, BK20150204), 江苏省重点研发计划(BE2015040), 国家重点研发计划(2016YFC0801403), 国家自然科学基金(51504214, 51504255, 51734009, 61771417)

Foundation Items: The Natural Science Foundation of Jiangsu Province (BK20180640, BK20150204), The Research Development Programme of Jiangsu Province (BE2015040), The State Key Research Development Program (2016YFC0801403), The National Natural Science Foundation of China (51504214, 51504255, 51734009, 61771417)

work, CNN)^[2]的不断发展, 其在图像分类^[3]、图像识别^[4]、显著性检测^[5]、语义分割^[6]等领域已经取得研究成果。在目标跟踪领域, 基于卷积特征的跟踪算法成为了研究的热点方向。Wang等人^[7]基于消噪自编码器原理, 使用网络深层特征提出深度学习跟踪器(DLT)算法; 文献^[8]中使用孪生神经网络的深层特征进行相关性计算, 提升了算法的稳健性。这些跟踪器与基于手工选择特征(如hog^[9]和颜色直方图^[10]等)的算法相比, 具有更高的跟踪精度。

上述基于卷积特征的跟踪算法通常使用网络深层特征来表征目标对象, 跟踪效果较好, 但是无法适应较为复杂的视频环境。网络深层特征丰富的语义信息有利于目标对象的分类, 但捕捉目标细节信息较少; 浅层特征对目标的细节捕捉较为明显, 但会引入较多的背景杂波。

利用特征的相关性计算响应主要包括相关滤波响应和相似性响应两类: 相关滤波响应^[11,12]通过训练滤波模板, 计算相关响应图的峰值定位目标。Henriques等人^[13]提出循环结构核(Circulant Structure Kernel, CSK)算法, 利用循环移位操作作为跟踪器训练构造大量负样本, 提升了跟踪器判别能力; 随后在文献^[14]中在该算法中加入核技巧, 提出了核化相关滤波(Kernel Correlation Filter, KCF)算法, 提高了跟踪器的跟踪效率。相关滤波响应仅考虑帧间相关信息, 导致目标发生遮挡、漂移时跟踪精度偏低。相似性响应^[8,15,16]采用模板匹配后的峰值定位目标, 其在孪生神经网络中通过计算视频中模板图像和待检测图像特征的相似性响应最值确定目标位置, 取得了较好的效果。相似性响应仅计算首帧图像和待检测图像的相关性, 没有考虑相邻帧图像之间的相关性致使跟踪结果较低。

由上述分析可知, 相关滤波响应仅考虑帧间相关信息, 在目标发生遮挡和突变时算法跟踪精度较低; 相似性响应仅考虑图像第1帧与待检测帧相关信息, 缺少相邻帧的相关性信息降低了算法跟踪性能。文献^[17]中提出一种比例融合算法, 通过计算当前帧中各响应所占比例, 确定下一帧各响应对应权重融合系数, 效率较高但融合精度较低; 文献^[18]中提出一种融合算法, 通过在线学习的方式对响应权重进行训练更新, 取得了较好的融合效果。上述算法在融合权重更新时仅考虑各跟踪器当前位置决策信息, 当目标移动较快, 跟踪器的当前位置决策准确性可能较低, 其在历史帧中位置决策的准确性相对较高。

综上所述, 本文提出一种基于多层卷积特征的自适应决策融合目标跟踪算法(Adaptive Strategy Fusion Target Tracking algorithm, ASFTT)。本

算法首先提取多层卷积特征表征目标信息, 提高模型泛化能力; 其次使用各层特征计算帧图像相关性响应, 以表征不同帧图像之间的相关性信息, 提高算法的跟踪精度; 最后通过本文自适应决策融合算法将所有响应中位置信息动态融合, 训练融合权重时同时考虑产生响应的各跟踪器的历史和当前决策信息, 提高算法的鲁棒性。

2 ASFTT跟踪算法

本文ASFTT算法框图如图1所示。算法包含特征提取、响应计算和决策融合3部分。在特征提取阶段, 提取孪生神经网络多层卷积特征; 在响应计算阶段, 提出一种帧图像相关性响应计算方式, 包括计算相关滤波和相似性两类响应; 在决策融合阶段, 使用自适应决策融合算法融合位置决策信息以定位目标。

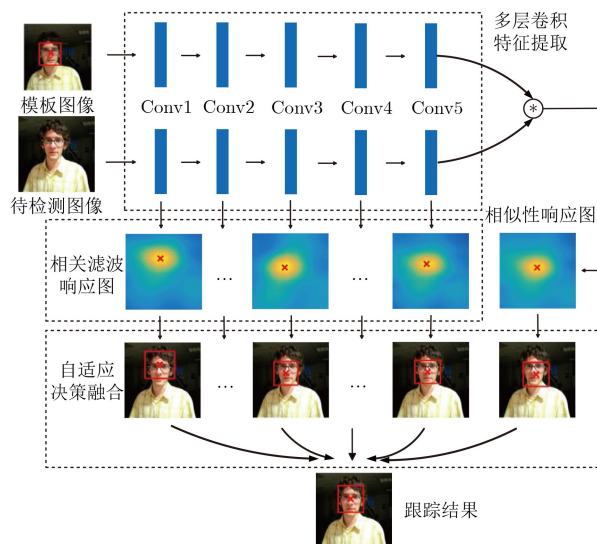


图1 ASFTT算法框图

2.1 多层卷积特征提取

卷积神经网络(CNN)作为一个多层感知器, 每个卷积层都可以表达输入图像的不同特征。孪生神经网络^[8]是典型的CNN网络结构, 其主要由Conv1~Conv5的5层卷积层组成, 其中, Conv1和Conv2层的卷积核大小分别为 11×11 和 5×5 , 其余各卷积层卷积核大小为 3×3 。孪生神经网络仅使用Conv5的深层卷积特征, 没有使用Conv1~Conv4的较浅层卷积特征表征目标信息。

不同深度的卷积层特征具有不同的表达能力, 层数越浅, 特征图中包含的细节信息越多, 但会产生背景杂波; 层数越深, 特征图中较多的语义信息有利于训练跟踪器进行跟踪, 但包含的细节信息越少。因此在算法特征提取阶段, 本文提取待检测图像通过孪生神经网络获得的多层卷积特征 o^k , 表示不同

分辨率和语义信息下的待检测目标,提取模板图像最后一层特征 u' 表征已检测目标,如式(1)所示,其中 P 、 Q 和 C 分别表示特征图的宽、高和通道数

$$o^k, u' \in \mathbb{R}^{P \times Q \times C} \quad (1)$$

由于卷积网络的池化操作,待检测图像分支中不同网络层特征图的尺寸大小不同,层级越深,特征图的尺寸越小。利用双线性插值对特征图 o^k 进行上采样操作,使得所有卷积层的特征图具有相同尺寸 x^k ,式(2) x_i^k 表示 x^k 中第 i 个位置的特征向量,其中 a_{ij} 为插值权重

$$x_i^k = \sum_j a_{ij} o_j^k \quad (2)$$

2.2 帧图像相关性响应

2.2.1 相关滤波响应

相关滤波响应值表征图像相邻帧之间的相关性。对待检测图像各层特征 x^k 在垂直和水平方向上的循环移位作为训练样本 $x_{p,q}^k$, $(p, q) \in \{0, 1, \dots, P-1\} \times \{0, 1, \dots, Q-1\}$ 。通过最小化理想输出 $y^k(p, q)$ 与 $x_{p,q}^k$ 的损失误差,对第 k 个相关滤波跟踪器进行最优建模

$$h^k = \arg \min \left[\sum_{p,q} \|y^k(p, q) - h \cdot x_{p,q}^k\|^2 + \lambda \|h\|_2^2 \right] \quad (3)$$

其中, λ 是正则化参数且 $\lambda \geq 0$; $y^k(p, q) = \exp(-(p-P/2)^2 + (q-Q/2)^2/2\sigma^2)$ 是与 $x_{p,q}^k$ 对应的2维高斯核函数, σ 为高斯核的宽度; $h \cdot x_{p,q}^k = \sum_{c=1}^C (h_{p,q,c})^T x_{p,q,c}^k$,表明各通道矩阵相乘; h^k 为需要求得的最优权重。

对式(3)在傅里叶域中求解得到第 k 个跟踪器 $c(c \in \{1, 2, \dots, C\})$ 通道的最优权重 H_c^k ,其中 \odot 表示点乘操作; Y^k 和 X^k 分别为 y^k 和 x^k 的DFT变换; \bar{X}^k 为 X^k 的复数共轭。如式(4)所示

$$H_c^k = \frac{Y^k \odot \bar{X}_c^k}{\sum_{i=1}^C X_i^k \odot \bar{X}_i^k + \lambda} \quad (4)$$

对于第 t 帧各跟踪器 $H_c^{k,t}$,令 $A_c^{k,t-1}$ 和 $B_c^{k,t-1}$ 分别表示其在第 $t-1$ 帧时的分子、分母项,则第 t 帧中,跟踪器的更新策略如式(5)所示

$$\left. \begin{aligned} A_c^{k,t} &= (1-\eta)A_c^{k,t-1} + \eta Y^k \odot \bar{X}_c^{k,t} \\ B_c^{k,t} &= (1-\eta)B_c^{k,t-1} + \eta \sum_{i=1}^C X_i^{k,t} \odot \bar{X}_i^{k,t} \\ H_c^{k,t} &= A_c^{k,t} / (B_c^{k,t} + \lambda) \end{aligned} \right\} \quad (5)$$

参数更新后的跟踪器用于预测目标的位置,对于第 k 个卷积层提取的测试图像特征 $z^k \in \mathbb{R}^{P \times Q \times C}$,通过检测得到响应输出 R^k ,如式(6)所示

$$R^k = F^{-1} \left(\sum_{c=1}^C \bar{Z}_c^k \odot H_c^k \right) \quad (6)$$

其中, \bar{Z}^k 为 z^k 的DFT的共轭复数。查找最大响应的位置即为目标的预测位置 (a^k, b^k) ,如式(7)所示

$$(a^k, b^k) = \arg \max_{a,b} R^k(a, b) \quad (7)$$

2.2.2 相似性响应

相似性响应表征待检测图像与第1帧模板图像之间的相关性,有助于提高突变,漂移等干扰性因素下跟踪器性能。本文使用离线训练的孪生神经网络最后一层输出的第1帧模板特征和待检测图像特征计算相似性响应,如式(8)所示

$$R' = u'_1 * x'_t + b \quad (8)$$

其中, u'_1, x'_t 分别表示第1帧模板图像和第 t 帧待检测图像经过网络最后一层的特征映射图,*表示卷积操作。计算特征映射 u'_1 与 x'_t 的相似性,生成响应图 R' 。响应图中得分最高的点为目标预测位置 (a', b') ,如式(9)所示

$$(a', b') = \arg \max_{a,b} R'(a, b) \quad (9)$$

2.3 自适应决策融合

文献[18]中提出了一种融合算法,以解决多决策者的决策理论问题。算法中决策者为产生响应的各个跟踪器,加权融合所有决策者生成的决策(预测位置),计算目标的最终位置。通过计算决策损失更新每个决策者对应的权重以优化决策模型。

然而,此算法只考虑了同一决策者的当前决策信息,在目标移动速度较快,发生遮挡等情况时鲁棒性较差。因此,本文使用自适应决策融合算法,通过引入稳定性度量值,在训练权重时自适应地对各决策者的历史决策信息和当前决策信息分配不同的比例。将2.2节中相关性响应产生的预测位置分别视为相应决策,预测视频中第 t 帧的目标决策位置,如式(10)所示

$$(a_t, b_t) = \sum_{m=1}^M (w_t^m \cdot a_t^m, w_t^m \cdot b_t^m) \quad (10)$$

其中, w_t^m 为每个决策者 m 对应的权重,所有权重总和 $\sum_{m=1}^M w_t^m = 1$ 。计算第 t 帧中决策者 m 的决策损失

$$L_t^m = \max(R_t^m) - R_t^m(a_t, b_t) \quad (11)$$

式(11)中 $\max(\cdot)$ 表示计算响应矩阵中的最大值, $R(a, b)$ 表示矩阵中目标位置 (a, b) 的响应值。

式(12)计算所有决策者的加权平均损失与决策者 m 决策损失的差值,即当前代价函数 p_t^m

$$p_t^m = \bar{L}_t^m - L_t^m \quad (12)$$

其中, $\bar{L}_t^m = \sum_{m=1}^M w_t^m L_t^m$ 。由于在视频序列中目标在短时间内不会发生显著的外观变化,本文使用均值为 μ_t^m ,标准方差为 σ_t^m 的高斯分布来表示每个决策者 m 在时间 Δt 内的决策损失,如式(13)所示

$$\left. \begin{aligned} \mu_t^m &= \frac{1}{\Delta t} \sum_{\gamma=t-\Delta t+1}^t \max(R_\gamma^m) - R_\gamma^m(a_\gamma, b_\gamma) \\ \sigma_t^m &= \sqrt{\frac{1}{\Delta t-1} \sum_{\gamma=t-\Delta t+1}^t (\max(R_\gamma^m) - R_\gamma^m(a_\gamma, b_\gamma) - \mu_t^m)^2} \end{aligned} \right\} \quad (13)$$

计算决策者 m 的稳定性度量值 r_t^m ,如式(14)所示

$$r_t^m = \frac{|\max(R_t^m) - R_t^m(a_t, b_t) - \mu_t^m|}{\sigma_t^m} \quad (14)$$

由式(14)可知, r_t^m 较小表明决策者 m 的决策信息准确率变化不大,较为稳定,因此应该主要相信决策者 m 在第 t 帧中的表现; r_t^m 较大说明决策者 m 的决策信息准确率变化较大,不太稳定,因此应该主要相信决策者 m 的历史表现。如式(15a)和式(15b)所示

$$S_t^m = (1 - \alpha_t^m) S_{t-1}^m + \alpha_t^m p_t^m \quad (15a)$$

$$\alpha_t^m = \min(\exp(-\beta r_t^m), g) \quad (15b)$$

其中,累积代价函数 S_t^m 由历史累积代价函数 S_{t-1}^m 和当前代价函数 p_t^m 两部分组成。 β 为比例系数, g 定义了 p_t^m 在 S_t^m 中所占的最大比例,以避免出现不考虑 S_{t-1}^m 的情况。

通过最小化 S_t^m 以优化每个决策者 m 对应的权重。本文所提自适应决策融合算法权重优化与文献[18]采用相同的方法

$$w_{t+1}^m \propto ([S_t^m]_+ / c_t) \exp\left(\frac{([S_t^m]_+)^2}{2c_t}\right) \quad (16)$$

其中, $[S_t^m]_+$ 表示 $\max\{0, S_t^m\}$, c_t 与文献[18]中比例系数相同,其值通过求解 $\frac{1}{M} \sum_{m=1}^M \exp\left(\frac{([S_t^m]_+)^2}{2c_t}\right) = e$ 来确定。

2.4 算法流程

算法流程如表1所示。

3 实验结果

为验证算法的综合性能,在标准数据集OTB2013^[19]中将本文所提ASFTT算法和SiamFC^[8], SAMF_CA^[20], fDSST^[21], Staple^[22], KCF^[14]和CNT^[23]算法进行了对比分析。其中在评估算法性能中,距离精度DP是指视频中预测目标位置与实际目标位置之间的中心误差小于阈值的帧数占总帧数的百分比,重叠成功率OS是指视频中预测目标位置与实际目标位置之间的重叠率大于阈值的帧数占总帧数的百分比。

3.1 实验条件和参数

实验条件为:PC机, Intel(R)Core(TM)i7 CPU, NVIDIA GTX Titan X GPU, Matlab R2016a。OTB2013数据集中包含11种影响因素:光照变化、尺度变化、遮挡、形变、运动模糊、快速运动、平面内旋转、平面外旋转、出视野、背景杂波和低分辨率。实验参数设置: $\lambda=10^{-4}$; $\eta=0.01$ 。

3.2 定量分析

3.2.1 整体性能分析

在一次通过评估(One Pass Evaluation, OPE)条件下使用距离精度(DP)和重叠成功率(OS)对ASFTT算法和6种算法的性能进行整体评估。精度曲线使用DP进行评估,成功率曲线使用下曲线面积(Area Under Curve, AUC)进行比较。

由图2可知,本文算法在总体的距离精度和成

表1 基于多层卷积特征的自适应决策融合目标跟踪算法

输入:视频序列第1帧的目标位置;初始各决策权重 $w_1^1, w_1^2, \dots, w_1^m$; $R_1^m = 0, L_1^m = 0$ 。

输出:每帧图像的目标位置 (a_t, b_t) 。

(1) //权重初始化。使用式(4)计算 k 个跟踪器的初始权重;

(2) for $t=2$ to T (T 是视频的总帧数):

(3) //提取网络多层特征。提取网络中待检测图像 k 层的特征 x_t^k 和模板分支最后一层特征 u_t^1 ;

(4) //响应值计算。使用式(6)和式(8)计算 k 个相关滤波响应值 R_t^k 和相似性响应值 R_t^s ;

(5) //自适应响应决策融合。计算目标位置首先使用式(7)和式(9)计算步骤(4)中每个决策者预测的目标位置 (a_t^m, b_t^m) ;通过式(10)计算最终的目标位置 (a_t, b_t) ;

(6) //更新权重值,用于下一帧检测。首先通过式(11)和式(12)计算各决策者的损失 L_t^m 和当前代价函数 p_t^m ;其次使用式(13)和式(14)更新稳定性模型并计算每个决策者的稳定性度量值 r_t^m ;使用式(15b)和式(15a)计算每个决策者当前代价函数 p_t^m 的 α_t^m 比例值和每个决策者的累积代价函数 S_t^m ;并使用式(16)更新每个决策者所对应的权重 w_{t+1}^m ;最后通过式(5)更新 k 个跟踪器的权重;

(7) end for;

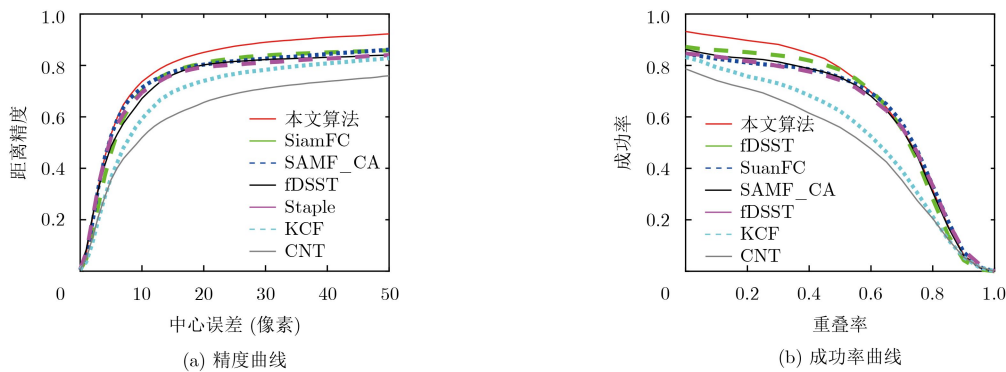


图2 算法整体的精度曲线和成功率曲线图

功率两个指标上都取得第1名，其中距离精度值DP和成功重叠率值OS高于SiamFC, SAMF_CA, fDSST等其它跟踪算法。因此本文所提ASFTT算法在整体性能上较为优秀。

3.2.2 基于视频集属性的分析

本文在OTB2013中OPE条件下对7种跟踪算法进一步比较分析，图3依次为各算法在快速运动和遮挡影响因素下的跟踪效果。精度曲线使用DP进行评估，成功率曲线使用下曲线面积(AUC)进行比较。

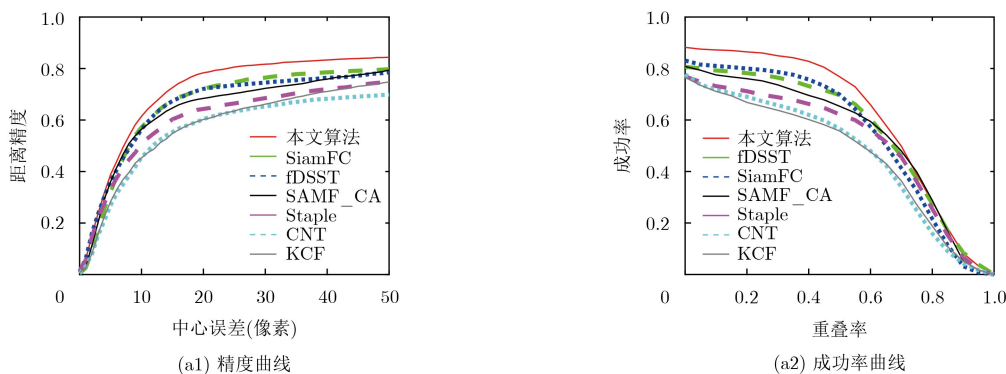
由图3可知，本文算法在快速运动和遮挡影响因素下均具优势。距离精确度DP和成功重叠率OS高于SiamFC, SAMF_CA, fDSST等其它跟踪算

法。表明本文算法在复杂视频环境下仍有良好的跟踪精度和鲁棒性。

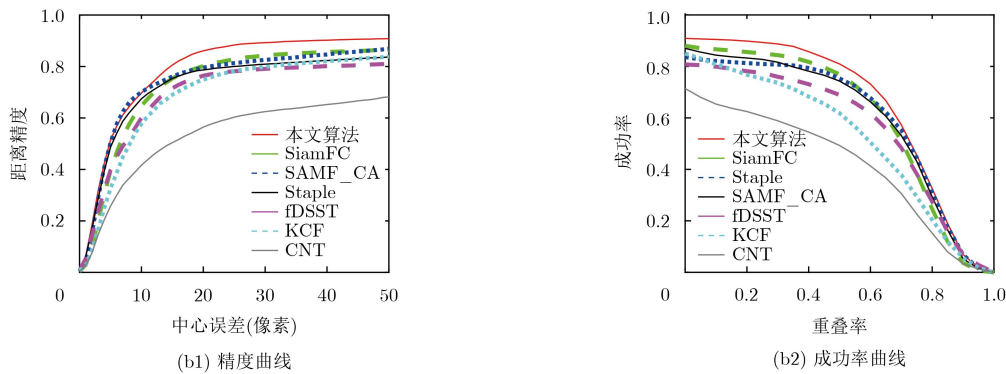
3.3 定性分析

本文从OTB2013视频库中选择4个包含多个影响因素的视频序列，比较4个跟踪器(SiamFC, fDSST, Staple和KCF)和ASFTT的跟踪性能。SiamFC是CNN跟踪器，fDSST是加入特征降维的相关滤波跟踪器，Staple是基于多特征融合的相关滤波跟踪器，KCF是基于核相关滤波的跟踪器。部分跟踪结果如图4所示。表2显示了这些序列中包含的影响因素。

(1) 图4(a)为basketball。第57帧，目标发生形



(a) 快速运动的精度曲线和成功率曲线图



(b) 遮挡的精度曲线和成功率曲线图

图3 算法各属性的精度曲线和成功率曲线图

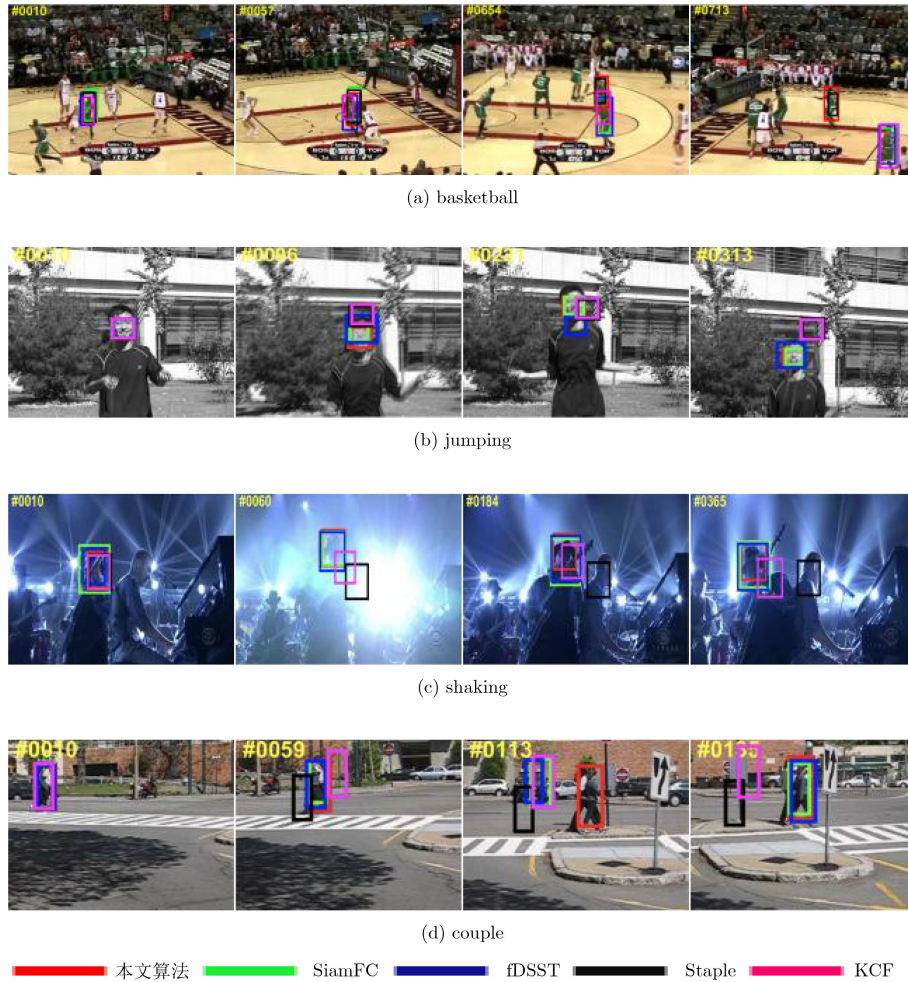


图 4 跟踪效果对比图

表 2 测试视频序列包含的影响因素

序列	帧数	影响因素
basketball	725	形变、遮挡、光照变化、背景杂波等
jumping	313	运动模糊、快速运动
shaking	365	光照变化、背景杂波、尺度变化等
couple	140	平面外旋转、尺度变化、形变等

变，fDSST和Staple出现跟踪漂移。第654帧，视频中出现相似人物目标干扰时，SiamFC, fDSST跟踪失败，KCF算法出现偏移。本文算法在视频中则始终紧跟目标。

(2) 图4(b)为jumping。第96帧，第221帧以及后续帧目标移动速度较快时，fDSST算法出现漂移，Staple和KCF跟丢物体。本文算法和SiamFC始终跟踪情况良好。

(3) 图4(c)为shaking。第60帧，本文算法，SiamFC和fDSST与目标保持一致，其它算法在背景遭遇高光时偏离目标。第184帧，出现颜色类似干扰时，SiamFC和fDSST出现漂移。

(4) 图4(d)为couple。第59帧发生形变时，

Staple, KCF发生漂移，之后丢失目标。第113帧出现相似物时，除本文算法其它跟踪器全部跟踪失败。

由上述分析可以看出，本文所提基于多层卷积特征的自适应决策融合目标跟踪算法能有效应对复杂的视频环境，在保障跟踪精度的同时兼顾鲁棒性。

4 结论

本文提出了一种基于多层卷积特征的自适应决策融合目标跟踪算法。通过提取网络多层特征，提高算法的泛化能力；使用多层特征计算帧图像相关性响应，提高算法的跟踪精度；本文使用自适应决策融合算法，将各跟踪器的历史决策信息和当前决策信息引入融合权重训练之中，提高跟踪算法在遮挡、形变等复杂环境中的鲁棒性。在标准数据集OTB2013中的实验结果表明，本文算法的跟踪精度与鲁棒性明显优于其它算法，证明了本文算法优秀的性能。

参考文献

[1] 侯志强, 张浪, 余旺盛, 等. 基于快速傅里叶变换的局部分块视觉跟踪算法[J]. 电子与信息学报, 2015, 37(10): 2397-2404.

- doi: [10.11999/JEIT150183](https://doi.org/10.11999/JEIT150183).
 HOU Zhiqiang, ZHANG Lang, YU Wangsheng, *et al.* Local patch tracking algorithm based on fast fourier transform[J]. *Journal of Electronics & Information Technology*, 2015, 37(10): 2397–2404. doi: [10.11999/JEIT150183](https://doi.org/10.11999/JEIT150183).
- [2] HUANG C, LUCEY S, and RAMANAN D. Learning policies for adaptive tracking with deep feature cascades[C]. Proceedings of IEEE International Conference on Computer Vision, Venice, Italy, 2017: 105–114.
- [3] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2012: 1097–1105.
- [4] WANG Linzhao, WANG Lijun, LU Huchuan, *et al.* Saliency detection with recurrent fully convolutional networks[C]. Proceedings of the 14th Computer Vision European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 825–841.
- [5] LONG J, SHELHAMER E, and DARRELL T. Fully convolutional networks for semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 3431–3440.
- [6] 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. Cham: Springer, 2016: 472–488.
- [7] WANG Naiyan and YEUNG D Y. Learning a deep compact image representation for visual tracking[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 2013: 809–817.
- [8] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.* Fully-convolutional siamese networks for object tracking[C]. Proceedings of the Computer Vision – ECCV 2016 Workshops, Amsterdam, The Netherlands, 2016: 850–865.
- [9] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886–893.
- [10] TIAN Gang, HU Ruimin, WANG Zhongyuan, *et al.* Improved object tracking algorithm based on new hsv color probability model[M]. YU Wen, HE Haibo, ZHANG Nian. Advances in Neural Networks – ISNN 2009. Berlin Heidelberg, Springer, 2009: 1145–1151.
- [11] 孙航, 李晶, 杜博, 等. 基于多阶段学习的相关滤波目标跟踪[J]. 电子学报, 2017, 45(10): 2337–2342. doi: [10.3969/j.issn.0372-2112.2017.10.004](https://doi.org/10.3969/j.issn.0372-2112.2017.10.004).
 SUN Hang, LI Jing, DU Bo, *et al.* Correlation filtering target tracking based on online multi-lifespan learning[J]. *Acta Electronica Sinica*, 2017, 45(10): 2337–2342. doi: [10.3969/j.issn.0372-2112.2017.10.004](https://doi.org/10.3969/j.issn.0372-2112.2017.10.004).
- [12] DANELLJAN M, HÄGER G, KHAN F S, *et al.* Accurate scale estimation for robust visual tracking[C]. Proceedings of British Machine Vision Conference, Nottingham, UK, 2014: 65.1–65.11.
- [13] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[C]. Proceedings of the 12th European Conference on Computer Vision, Florence, Italy, 2012: 702–715.
- [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. doi: [10.1109/tpami.2014.2345390](https://doi.org/10.1109/tpami.2014.2345390).
- [15] HELD D, THRUN S, and SAVARESE S. Learning to track at 100 FPS with deep regression networks[C]. Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 749–765.
- [16] TAO Ran, GAVVES E, and SMEULDERS A W M. Siamese instance search for tracking[C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1420–1429.
- [17] ZHANG Hainan, SUN Yanjing, LI Song, *et al.* Long-term tracking based on multi-feature adaptive fusion for video target[J]. *IEICE Transactions on Information and Systems*, 2018.
- [18] CHAUDHURI K, FREUND Y, and HSU D. A parameter-free hedging algorithm[C]. Proceedings of the 22nd International Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2009: 297–305.
- [19] WU Yi, LIM J, and YANG M H. Online object tracking: a benchmark[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA, 2013: 2411–2418.
- [20] MUELLER M, SMITH N, and GHANEM B. Context-aware correlation filter tracking[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 1387–1395.
- [21] DANELLJAN M, HÄGER G, KHAN F S, *et al.* Discriminative scale space tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(8): 1561–1575. doi: [10.1109/TPAMI.2016.2609928](https://doi.org/10.1109/TPAMI.2016.2609928).
- [22] BERTINETTO L, VALMADRE J, GOLODETZ S, *et al.* Staple: complementary learners for real-time tracking[C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1401–1409.
- [23] ZHANG Kaihua, LIU Qingshan, WU Yi, *et al.* Robust visual tracking via convolutional networks without training[J]. *IEEE Transactions on Image Processing*, 2016, 25(4): 1779–1792.
- 孙彦景: 男, 1977年生, 教授, 博士生导师, 研究方向为无线传感器网络、视频目标跟踪、人工智能、信息物理系统。
 石镛开: 男, 1993年生, 硕士生, 研究方向为视频目标跟踪和人工智能。
 云霄: 女, 1986年生, 讲师, 研究方向为视频目标跟踪和人工智能。
 朱绪冉: 女, 1993年生, 硕士生, 研究方向为目标检测与识别。
 王赛楠: 女, 1992年生, 硕士生, 研究方向为视频目标跟踪。