

基于自适应卷积特征的目标跟踪算法

蔡玉柱, 杨德东, 毛 宁, 杨福才

河北工业大学控制科学与工程学院, 天津 300130

摘要 针对空间正则化相关滤波(SRDCF)跟踪算法在目标跟踪中旋转变化、超出视野和严重遮挡情况下存在跟踪失败的问题,提出了一种基于自适应卷积特征的目标跟踪算法。对 VGG-Net 模型中 conv3-4 层卷积特征进行主成分分析,利用自适应降维技术将 conv3-4 层特征维数由 256 维降至 130 维。在检测区域求取分类器最大响应位置及其目标尺度信息,并对最大响应位置的目标进行置信度比较,训练在线支持向量机(SVM)分类器,以便在跟踪失败的情况下,重新检测到目标而实现长期跟踪。计算跟踪位置的峰旁比,选取可靠跟踪结果,更新模型。采用 OTB-2015 评估基准的 100 组视频序列进行测试,并与 38 种跟踪方法进行对比,验证了本文算法的有效性。实验结果表明:本文算法跟踪精度为 0.804,成功率为 0.607,排名第一,与 SRDCF 算法相比,两者分别提高了 1.9% 和 1.5%。针对目标发生旋转变化、超出视野和严重遮挡等复杂情况,本文算法均具有较强的稳健性。

关键词 机器视觉; 卷积特征; 自适应降维; 在线支持向量机分类器; 峰旁比

中图分类号 TP391 文献标识码 A

doi: 10.3788/AOS201737.0315002

Visual Tracking Algorithm Based on Adaptive Convolutional Features

Cai Yuzhu, Yang Dedong, Mao Ning, Yang Fucui

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

Abstract Focusing on the issue that spatially regularized discriminative correlation filter (SRDCF) tracking algorithm has poor performance in handling rotation, out of view and heavy occlusion, we propose a visual tracking approach based on adaptive convolutional features. First, based on the principal component analysis of conv3-4 layer features in the VGG-NET model, the dimension of conv3-4 layer features is reduced from 256 to 130 by adaptive dimension reduction technique. Then, we maximize classifier score in the detection area and get the location and scale of target. In order to redetect the target in the case of tracking failure and achieve long-term tracking, we compare the confidence of the location with maximum score and train an online support vector machine (SVM) classifier. Finally, the tracking model is updated by the reliable tracking results which are determined by peak-to-sidelobe ratio. To verify the feasibility of the proposed algorithm, the results are compared with those obtained by thirty-eight kinds of tracking algorithms in one hundred video sequences of OTB-2015 benchmark. Experimental results indicate that the precision and success rate are respectively 0.804 and 0.607. The proposed approach has a ranking of one. Compared with SRDCF tracking algorithm, the proposed approach improves the precision and the success rate by 1.9% and 1.5%, respectively. In addition, the proposed approach is robust for rotation, out of view, heavy occlusion and other complex scenes.

Key words machine vision; convolutional features; adaptive dimensionality reduction; online support vector machine classifier; peak-to-sidelobe ratio

OCIS codes 150.0155; 150.1135; 150.1488

1 引 言

目标跟踪是计算机视觉中具有挑战性的难题之一。它是首先由视频序列的初始帧给定目标的初始状态

收稿日期: 2016-09-18; 收到修改稿日期: 2016-10-29

基金项目: 国家自然科学基金(61203076)、天津市自然科学基金(13JJCQNJC03500)

作者简介: 蔡玉柱(1990—),男,硕士研究生,主要从事视觉目标跟踪方面的研究。E-mail: caiyuzhu001@sina.com

导师简介: 杨德东(1977—),男,博士,副教授,主要从事视觉目标跟踪方面的研究。E-mail: ydd12677@163.com(通信联系人)

(比如位置和尺寸),并在随后的视频序列中估计出目标的状态的过程^[1]。尽管近几十年该领域取得重大进展,但由于受到外观变化、尺度变化、形变和遮挡因素的影响,目标跟踪仍然是一个挑战性问题。目前也不存在可以成功地处理所有应用场景的单一跟踪方法。通常根据目标外观模型的表达策略,可将跟踪算法分为生成式方法^[2-3]和判别式方法^[4-8]。生成式方法利用生成模型表达目标外观,然后在搜索图像区域中找到与模型最相似的区域作为目标。判别式方法则通过利用训练数据学习到一个分类器,建立一个将目标从背景中分离出来的判别模型。在判别式方法中,基于相关滤波器的跟踪算法^[9-14]在跟踪标准数据集上取得了优秀的跟踪性能。

尽管基于相关滤波器的跟踪算法可取得较好的跟踪效果,训练样本的周期性假设有效地提高了训练和检测的计算速度,但是周期性假设将边界效应引入到样本中,严重降低了跟踪模型的判别能力。Galogahi等^[15]采用乘法器的交替方向法(ADMM),确保相关滤波器的尺寸不会过大,解决单通道相关滤波器的边界效应。Danelljan等^[16]提出空间正则化相关滤波(SRDCF)跟踪算法,在相关滤波器学习中引入空间正则化,减轻多通道相关滤波器的边界效应。但是SRDCF算法在跟踪过程中仍存在以下问题:采用的梯度类特征HoG特征无法跟踪旋转目标;短期目标跟踪算法无法处理超出视野目标;在线模型采用固定速率更新,在遮挡因素下易产生由模型更新错误而引起的目标漂移。

本文在SRDCF算法基础上,采用卷积神经网络(CNN)特征,引入在线支持向量机(SVM)分类器重定位组件,并通过峰旁比选取可靠跟踪结果更新模型,解决了SRDCF算法存在的上述问题。利用OTB-2015^[17]评估基准的100组完全标注的视频序列以及跟踪-学习-检测(TLD)^[8]算法提供的两组长期跟踪视频序列进行测试,并与38种跟踪算法进行了对比。

2 SRDCF跟踪器

SRDCF跟踪器^[16]是一种在傅里叶域利用稀疏的正则化矩阵学习得到的空间正则化相关滤波器。不但有效地减轻了多通道相关滤波器的边界效应,而且在OTB-2015评估基准中相比相关滤波跟踪算法取得了更好的跟踪性能。

2.1 空间正则化相关滤波器的获取

SRDCF跟踪器在相关滤波器学习过程中,利用空间正则化权重函数 ω 调节相关滤波器参数 f 。空间正则化权重函数 ω 根据空间位置决定相关滤波器参数 f 的重要性,即背景区域赋予较高空间权重,目标区域赋予较小的空间权重。

$$\epsilon(f) = \sum_{k=1}^t \alpha_k \|S_f(x_k) - y_k\|^2 + \sum_{l=1}^d \|\omega \cdot f^l\|^2, \quad (1)$$

式中 $\alpha_k \geq 0$,决定每个样本 x_k 对相关滤波器参数 f 的影响; $S_f(x_k) = \sum_{l=1}^d x_k^l * f^l$ 表示相关滤波器对样本 x_k 的卷积响应, l 表示特征维数; y_k 为样本 x_k 的软标签; ω 为空间正则化权重函数, k 表示样本帧数, t 表示样本总帧数, d 表示特征总维数。 $\|\cdot\|$ 表示函数、向量和矩阵的2范数; θ_a 表示目标外观阈值。

对(1)式应用帕塞瓦尔定理,并进行向量化,即

$$\tilde{\epsilon}(\hat{f}) = \sum_{k=1}^t \alpha_k \left\| \sum_{l=1}^d \mathbf{D}(\hat{x}_k^l) \hat{f}^l - \hat{y}_k \right\|^2 + \sum_{l=1}^d \left\| \frac{\mathbf{C}(\hat{\omega})}{MN} \hat{f}^l \right\|^2, \quad (2)$$

式中“ $\hat{\cdot}$ ”符号记为离散傅里叶变换, M, N 为样本尺寸; $\mathbf{D}(\hat{x}_k^l)$ 表示向量 \hat{x}_k^l 的元素在其对角线上的对角矩阵; $\mathbf{C}(\hat{\omega})$ 表示函数 $\hat{\omega}$ 的二维循环卷积,即 $\mathbf{C}(\hat{\omega}) \hat{f}^l = \text{vec}(\hat{\omega} * \hat{f}^l)$,其中 $\text{vec}(\cdot)$ 表示矢量化。

构造单位矩阵 \mathbf{B} ,获取 \hat{f}^l 的实值部分 $\tilde{f}^l = \mathbf{B} \hat{f}^l$,并将所有向量化的相关滤波器实部联结起来,即

$$\tilde{\epsilon}(\tilde{f}) = \sum_{k=1}^t \alpha_k \|\mathbf{D}_k \tilde{f} - \tilde{y}_k\|^2 + \|\mathbf{W} \tilde{f}\|^2, \quad (3)$$

式中 $\tilde{f} = [(\tilde{f}^1)^T \cdots (\tilde{f}^d)^T]^T$, $\mathbf{D}_k = (\mathbf{D}_k^1 \cdots \mathbf{D}_k^d)$, $\mathbf{D}_k^l = \mathbf{B} \mathbf{D}(\hat{x}_k^l) \mathbf{B}^H$, $\tilde{y}_k = \mathbf{B} \hat{y}_k$, $\mathbf{C} = \mathbf{B} \mathbf{C}(\hat{\omega}) \mathbf{B}^H / MN$, \mathbf{W} 为块对角矩阵,并且对角块等于 \mathbf{C} 。

最后,通过求解正规方程组 $\mathbf{A}_t \tilde{\mathbf{f}} = \tilde{\mathbf{b}}_t$ 得到正则化相关滤波器,其中

$$\mathbf{A}_t = \sum_{k=1}^t \alpha_k \mathbf{D}_k^T \mathbf{D}_k + \mathbf{W}^T \mathbf{W}, \quad (4)$$

$$\tilde{\mathbf{b}}_t = \sum_{k=1}^t \alpha_k \mathbf{D}_k^T \tilde{\mathbf{y}}_k. \quad (5)$$

2.2 模型更新

SRDCF 跟踪器利用固定学习效率 $\gamma \geq 0$,采用线性插值对新参数与之前帧的相关参数整合。(4)式中的 \mathbf{A}_t 和(5)式中的 $\tilde{\mathbf{b}}_t$ 更新为

$$\mathbf{A}_t = (1 - \gamma)\mathbf{A}_{t-1} + \gamma(\mathbf{D}_t^T \mathbf{D}_t + \mathbf{W}^T \mathbf{W}), \quad (6)$$

$$\tilde{\mathbf{b}}_t = (1 - \gamma)\tilde{\mathbf{b}}_{t-1} + \gamma \mathbf{D}_t^T \tilde{\mathbf{y}}_t. \quad (7)$$

3 自适应卷积特征目标跟踪

通过利用 VGG-Net^[18] 模型的卷积特征,设计了在线 SVM 分类器重定位组件,并通过峰旁比选取可靠跟踪结果更新模型,解决 SRDCF 跟踪器无法处理旋转、超出视野目标和严重遮挡情况下模型更新错误问题。流程框图如图 1 所示,图中 PSR 为像素支撑域。

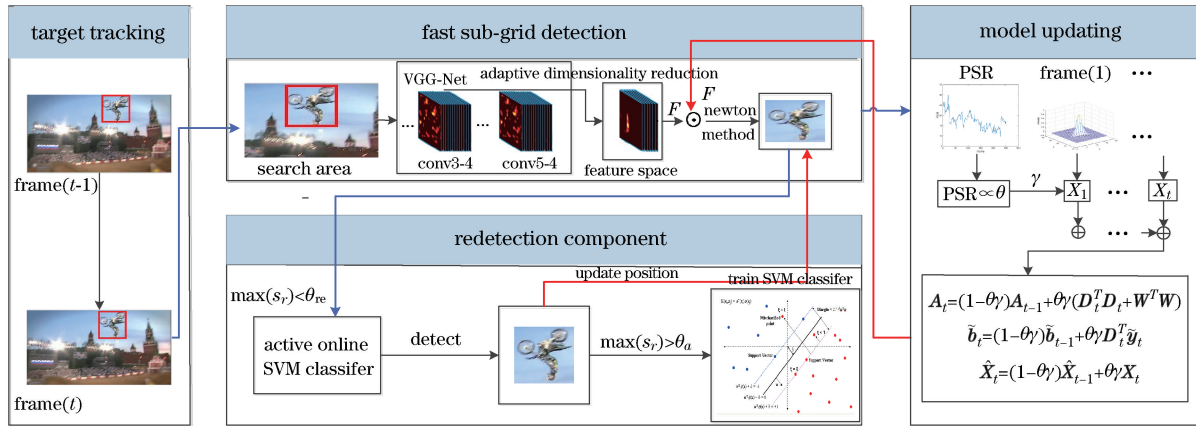


图 1 跟踪算法流程框图

Fig. 1 Flow chart of the proposed tracking algorithm

3.1 卷积特征

许多 CNN 模型已经成功应用到海量图像分类和目标检测中,如 AlexNet^[19]、R-CNN^[20]、CaffeNet^[21] 和 VGG-Net。VGG-Net 是通过利用 ImageNet 数据集的 130 万帧图片训练得到的,并在分类挑战中取得最好的结果。相比大多数只有 5 到 7 层的 CNN 模型,VGG-Net 具有更深的结构(高达 19 层,16 个卷积层和 3 个全连接层),可以提供更深层的特征。

研究表明,在同一个深度网络中,相比全连接层特征,卷积层特征具有更好的图片分类性能表现^[22],如图 2 所示。图 2(b)显示在 VGG-NET 模型中 conv3-4 层,其中卷积特征保留目标更多的空间细节,比如目标的位置信息,这对目标跟踪中准确定位非常有用;图 2(d)显示 VGG-NET 模型中 conv5-4 层,其中卷积特征包含目标更多语义信息,可以有效地处理跟踪过程中的目标形变问题,但关于目标的位置信息比较少。所以本文选择 VGG-NET 模型中 conv3-4 层卷积特征。

由于 CNN 模型中的池化操作,随着卷积层的增加,空间尺寸逐渐减小。如 VGG-Net 中的 pool5 特征映射空间为 7×7 ,这仅仅是像素大小为 224×224 的输入图片的 $1/32$ 。如此低的空间分辨率不足以正确定位目标位置,为此,利用双线性差值法将特征映射空间调整为更大的尺寸。

$$\mathbf{x}_k = \sum_i \beta_{ki} \mathbf{h}_i, \quad (8)$$

式中差值权重 β_{ki} 依赖于 k 帧位置和 i 个相邻特征向量, \mathbf{h} 为特征空间。

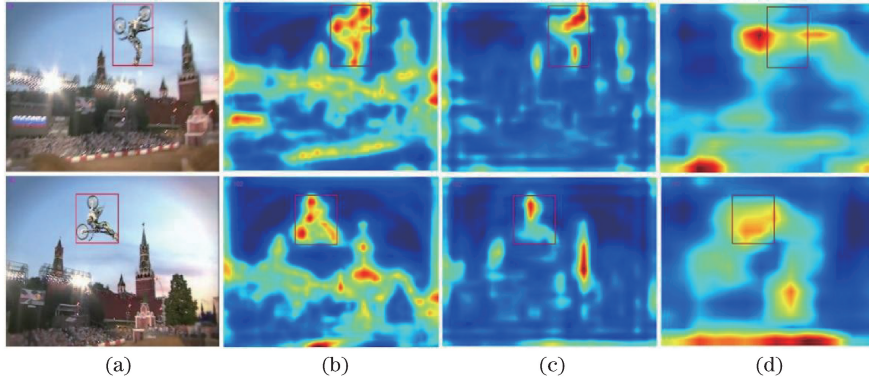


图 2 卷积特征响应图。(a) 原始图像；(b) conv3-4；(c) conv4-4；(d) conv5-4

Fig. 2 Convolutional features response. (a) Original images; (b) conv3-4; (c) conv4-4; (d) conv5-4

3.2 自适应降维

由于 VGG-Net 模型的 conv3-4 层特征空间映射为 $56 \times 56 \times 256$, 包含了一定量的次要成分, 并且较高的卷积特征维数严重影响跟踪算法的计算复杂度。通过对 conv3-4 层卷积特征进行主成分分析 (PCA), 设计了一种自适应降维技术保留 conv3-4 层卷积特征主要成分, 并将 conv3-4 层特征维数由 256 维降至 130 维。如图 3 所示, MotorRolling 数据集 conv3-4 层卷积特征前 130 维的累积贡献率已经达到 98%。

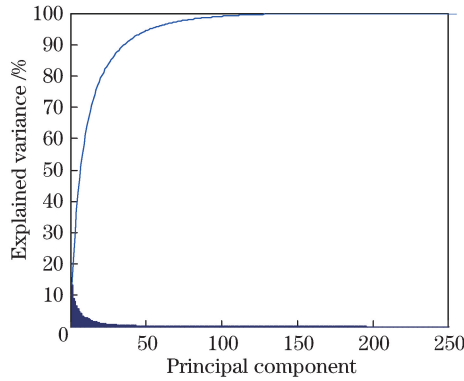


图 3 MotorRolling 的 conv3-4 卷积特征主成分分析

Fig. 3 Principal component analysis of conv3-4 convolutional features on MotorRolling

令 $\hat{\mathbf{x}}_t$ 表示 t 帧学习到的 D_1 维特征。自适应降维技术的设计思想是寻找大小为 $D_1 \times D_2$ 带有正交列向量的投影矩阵 \mathbf{P}_t , 且 $\mathbf{P}_t^T \mathbf{P}_t = \mathbf{I}$ 。利用投影矩阵 \mathbf{P}_t 得到 D_2 维新的特征空间,

$$\min \left\{ \eta_t \left[\frac{1}{MN} \sum_{m,n} \|\hat{\mathbf{x}}_t(m,n) - \mathbf{P}_t \mathbf{P}_t^T \hat{\mathbf{x}}_t(m,n)\|^2 \right] + \sum_{k=1}^{t-1} \eta_k \left[\sum_{l=1}^{D_2} \xi_k^{(l)} \|\mathbf{r}_k^{(l)} - \mathbf{P}_k \mathbf{P}_k^T \mathbf{r}_k^{(l)}\|^2 \right] \right\}, \quad (9)$$

式中 η_1, \dots, η_t 为影响权重, $\xi_k^{(l)} \geq 0$ 决定每个成分向量 $\mathbf{r}_k^{(l)}$ 的重要性, $\hat{\mathbf{x}}_t(m,n) = \mathbf{P}_t^T \hat{\mathbf{x}}_t(m,n), \forall m,n$ 。

求取(9)式等效于对矩阵 \mathbf{R}_t 作奇异值分解, 选择矩阵 \mathbf{R}_t 前 D_2 个标准的特征向量作为投影矩阵

$$\mathbf{R}_t = \eta_t \mathbf{G}_t + \sum_{k=1}^{t-1} \eta_k \mathbf{P}_k \mathbf{\Lambda}_k \mathbf{P}_k^T, \quad (10)$$

其中

$$\mathbf{G}_t = \frac{1}{MN} \sum_{m,n} [\hat{\mathbf{x}}_t(m,n) - \bar{\mathbf{x}}_t][\hat{\mathbf{x}}_t(m,n) - \bar{\mathbf{x}}_t]^T, \quad (11)$$

$$\bar{\mathbf{x}}_t = \frac{1}{MN} \sum_{m,n} \hat{\mathbf{x}}_t(m,n), \quad (12)$$

式中 \mathbf{G}_t 表示当前帧外观特征 $\hat{\mathbf{x}}_t$ 的方差矩阵; $\mathbf{\Lambda}_t$ 为 $D_2 \times D_2$ 对角矩阵, 对角上的元素为 $\xi_k^{(l)}$; $\xi_k^{(l)}$ 为成分向量 $\mathbf{r}_k^{(l)}$ 所对应矩阵 \mathbf{R}_t 的特征值。

为获得自适应投影矩阵,采用一个固定学习效率 λ ,利用线性插值对矩阵 \mathbf{R}_t 和方差矩阵 \mathbf{Q}_t 固定更新。利用固定学习效率 $\gamma \geq 0$,更新外观特征 $\hat{\mathbf{x}}_t$ 。

$$\mathbf{R}_t = (1 - \lambda)\mathbf{Q}_{t-1} + \lambda\mathbf{G}_t, \quad (13)$$

$$\mathbf{Q}_t = (1 - \lambda)\mathbf{Q}_{t-1} + \lambda\mathbf{P}_t\mathbf{A}_t\mathbf{P}_t^T, \quad (14)$$

$$\hat{\mathbf{x}}_t = (1 - \gamma)\hat{\mathbf{x}}_{t-1} + \gamma\mathbf{x}_t. \quad (15)$$

3.3 快速子网格检测

检测阶段,采用 SRDCF 跟踪算法的检测模块,在 t 帧以 $t-1$ 帧目标位置为中心,相对 $t-1$ 帧目标尺度的多个尺度 a^r 采样 $\{z_r\}$ 。其中 $r \in \{\text{ceiling}[(1-S)/2], \dots, \text{ceiling}[(S-1)/2]\}$, S 表示尺度数量, a 为尺度增量因子, $\text{ceiling}()$ 表示向下取整函数。第 t 帧目标候选区域 $(u, v) \in [0, M) \times [0, N)$ 的检测结果可以插值表示为

$$s_r(u, v) = \frac{1}{MN} \sum_0^{M-1} \sum_0^{N-1} \hat{s}_r(m, n) \exp \left[i2\pi \left(\frac{m}{M}u + \frac{n}{N}v \right) \right], \quad (16)$$

式中 i 表示虚数。

利用牛顿迭代处理(16)式,获取第 t 帧目标候选区域的最大检测值的位置和尺度:

$$(u^*, v^*, r^*) = \arg \max_{(u, v) \in [0, M) \times [0, N)} s_r(u, v). \quad (17)$$

3.4 在线 SVM 分类器重定位组件

长期跟踪中,当目标因受到长期遮挡或者超出视野而跟踪失败时,重定位组件在重新定位目标位置上至关重要。对于重定位组件,由于考虑跟踪实时性的要求,并未采用在每帧图像上训练的目标相关滤波器,而是采用独立训练的在线 SVM 分类器。

长期跟踪 TLD 算法采用对视频序列的每帧图像都执行重检测的策略,而本文算法则采用阈值 θ_{re} 激活策略,激活重定位组件。当 $\max(s_r) < \theta_{re}$ 时,即当分类器检测值最大值小于重定位激活阈值时,激活目标重检测,利用在线 SVM 分类器扫描整帧图像,获取在线 SVM 分类器的重检测结果,比较目标的置信度确定是否采用重检测结果。

采用一种半监督学习,训练在线 SVM 分类器,增强重定位组件的定位性能。利用标记集 $L_l = \{[x_{re}, y_{re}]\}$ 和未标记集 U_u ,其中 $l \ll u$, x_{re} 和 y_{re} 分别为重定位组件训练在线 SVM 分类器的样本及标签。标记集选取第一帧给定目标图像块和与第一帧给定目标框重叠率在区间(0.6, 1)的图像块作为正样本,而负样本为与第一帧给定目标框重叠率在区间(0, 0.2)的图像块。不同于正负样本来源于第一帧的标记集,未标记集在跟踪过程(帧数 $k \geq 2$)中产生,选取与跟踪的目标框重叠率在区间(0.6, 1)的图像块作为正样本,而负样本为与跟踪的目标框重叠率在区间(0, 0.2)的图像块。

1)将标签集 L_l 加入到训练集中,初始化训练在线 SVM 分类器,并获取在线 SVM 分类器的初始化参数 w_{re}^0 ; 2)将初始训练的在线 SVM 分类器在未标记集 U_u 上,经 ρ 次迭代获取在线分类器参数 w_{re}^ρ , $\forall x_{re(u)} \in U_u, y_{re(u)} = w_{re}^\rho \cdot x_{re(u)} + b$ 。利用 P-N 学习的结构化约束^[8],识别在线 SVM 分类器赋予未标记数据的标签是否与结构化约束相矛盾。只有当 $\max(s_r) > \theta_a$ 时,分类器检测值最大值大于目标外观阈值,从目标周围选取正负样本,加入并更新训练集,训练更新 SVM 分类器。

3.5 模型更新

由于 SRDCF 跟踪器采用固定学习率更新跟踪模型,一旦目标被遮挡则会出现模型更新错误,很容易导致跟踪漂移。受文献^[23]的启发,采用峰旁比(PSR, R_{PSR})计算目标位置置信度,选择性地更新模型。PSR 已经广泛应用在信号处理中,通常信号的峰值强度可表示为

$$R_{PSR, t} = \frac{\max[S_f(\mathbf{x}_t)] - \varphi_t}{\sigma_t}, \quad (18)$$

式中 $S_f(\mathbf{x}_t)$ 表示相关滤波器对样本 \mathbf{x}_t 的卷积响应, φ_t 和 σ_t 表示样本 \mathbf{x}_t 卷积响应图的均值和标准差。

图 4 显示了 David3 数据集的 PSR 分布情况,PSR 值越高目标位置的置信度越高。David3 序列在第 84 帧目标被树木完全遮挡,相应的 PSR 值下降到谷点,如图 4 中 A 点。随着遮挡程度减轻 PSR 值逐渐上升,当目标在第 188 帧再次被树木完全遮挡,相应的 PSR 值再次下降到谷点,如图 4 中 B 点。图中 A 和 B

点的跟踪结果不可靠,而此时更新模型,会出现更新错误。实验发现,当 PSR 值维持在 10~18 时跟踪结果可信度较高,即低于 9 时目标可能受到遮挡因素的影响,随着遮挡程度的增加 PSR 值逐步降低。

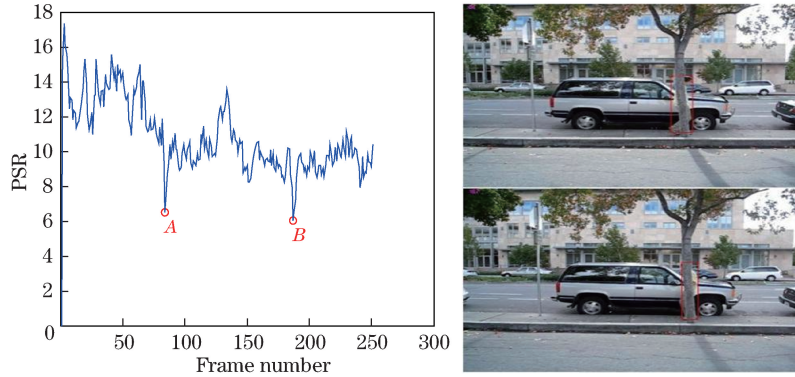


图 4 David3 数据集的 PSR 值分布及分析。(a) David3 数据集的 PSR 值分布图;
(b) David3 数据集的第 84 帧;(c) David3 数据集的第 188 帧

Fig. 4 Analysis and distribution of PSR on David3 dataset. (a) Distribution of PSR on David3 dataset;
(b) 84th frame of David3 dataset; (c) 188th frame of David3 dataset

因此,可以根据每一帧的 PSR 值判断目标是否受到遮挡影响,决定模型更新的权重:

$$\theta = \begin{cases} 0.1R_{\text{PSR}}, & \text{if } R_{\text{PSR}} \geq 10 \\ 0, & \text{if } R_{\text{PSR}} < 10 \end{cases}, \quad (19)$$

故目标的模型更新如下:

$$\mathbf{A}_i = (1 - \theta\gamma)\mathbf{A}_{i-1} + \theta\gamma(\mathbf{D}_i^T\mathbf{D}_i + \mathbf{W}^T\mathbf{W}), \quad (20)$$

$$\tilde{\mathbf{b}}_i = (1 - \theta\gamma)\tilde{\mathbf{b}}_{i-1} + \theta\gamma\mathbf{D}_i^T \tilde{\mathbf{y}}_i, \quad (21)$$

$$\hat{\mathbf{x}}_i = (1 - \theta\gamma)\hat{\mathbf{x}}_{i-1} + \theta\gamma\mathbf{x}_i. \quad (22)$$

4 实验结果与分析

为验证本文算法的有效性,利用 OTB-2015^[15]中包含 11 种属性完全标注的 100 组视频数据集进行评估,并与 SRDCF, MUSTER, LCT, RPT, SAMF, KCF, DSST, TGPR 以及 OTB-2015 代码库中所包含的 30 种跟踪算法进行了对比。OTB-2015 代码库所包含了以下跟踪算法:CSK, struck, SCM, TLD, CT, VTD, VTS, CXT, IVT, ASLA, DFT, LSK, CPF, LOT, VR-V, KMS, L1APG, MTT, MIL, OAB, SemiT, ORIA, SMS, Frag, BSBT, LSHT, LSS, CCT, FOT, PCOM。OTB-2015 视频数据集包含了以下属性: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)、复杂背景(BC)和低分辨率(LR)。

4.1 实验环境及参数设置

本实验的基本配置为 Matlab2013a, Intel(R) Core(TM) i7-4790 CPU, 主频 3.6 GHz, 8 GB 内存。空间正则化权重函数 $\omega(m, n) = \tau + \zeta\{(m/P)^2 + (n/Q)^2\}$, 最小空间正则化权重 $\omega = \tau = 0.1$, 正则化影响因数 $\zeta = 3$, $P \times Q$ 表示目标尺寸, 尺度数量 $S = 7$, 尺度增量因子 $a = 1.01$ 。自适应降维, conv3-4 层特征维数 $D_1 = 256$, 降维维数 $D_2 = 130$, 线性插值学习效率 $\lambda = 0.15$ 。重定位组件, 设置 $\theta_{re} = 0.5$ 激活训练好的在线 SVM 分类器, 设置 $\theta_a = 0.5$ 更新重定位目标模型。模型更新中, $R_{\text{PSR},i}$ 阈值设置为 10, 学习效率 $\gamma = 0.01$ 。对于实验中所有的视频数据集, 实验参数均保持固定。该算法由于采用 VGG-Net 神经网络提取目标特征比较耗时, 高维的卷积特征计算复杂度比较高, 在 CPU 模式下算法速度仅为 1.5 frame/s。

4.2 性能评估方法

为了评估跟踪算法的性能, 实验中采用 OTB-2015 中的两种评估方法: 精确度图和成功率图。

在跟踪精度评估中, 中心位置误差(CLE)被广泛应用。中心位置误差定义为跟踪目标的中心位置和手工标定的准确位置之间的平均欧氏距离。

1) 精确度图,能够显示目标中心位置误差小于阈值的帧数占视频总帧数的百分比。实验中阈值选定为 20 pixel。

2) 成功率图,能够显示重叠率 S 大于给定阈值 t_0 的帧数占视频总帧数的百分比。其中,重叠率 S 定义为 $S = |R_t \cap R_a| / |R_t \cup R_a|$, R_t 跟踪的目标框, R_a 为准确的目标框, \cap 和 \cup 分别表示两个区域的交集和并集, $|\cdot|$ 为区域内的像素点的数量。成功率给出了阈值 t_0 从 0 到 1 变化时成功帧数所占的比例,利用成功率图曲线下的面积(AUC)作为排序跟踪算法性能的准则。

4.3 性能比较

实验采用基于精确度图和成功率图的一次通过评估(OPE)方法评估跟踪算法性能。加入 SRDCF, MUSTER, LCT, RPT, SAMF, KCF, DSST, TGPR 算法,利用 OTB-2015 评估基准的 100 组视频序列进行测试,对共计 38 个跟踪算法进行了性能比较。为了显示清楚,在精确度图和成功率图上只显示了排序前 10 的跟踪算法。

4.3.1 定量比较

4.3.1.1 整体性能分析

图 5 为整体性能排名前 10 的 OPE 精确度图和成功率图。依据精确度图和成功率图,本文算法均排在第一位。在精确度图中,算法精确度为 0.804,相比于 SRDCF 算法提高了 1.9%。在成功率图中,算法成功率为 0.607,相比 SRDCF 算法提高了 1.5%。

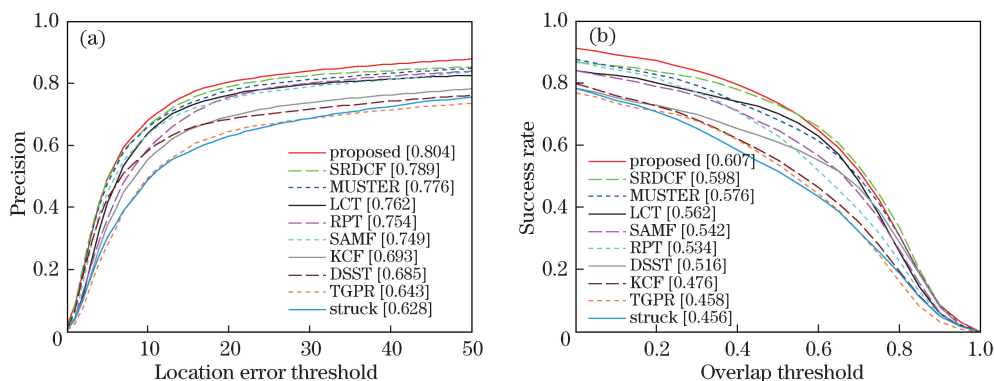


图 5 排名前 10 跟踪算法 OPE 的(a)精确度图和(b)成功率图

Fig. 5 (a) Precision plots and (b) success rate plots of OPE for the top 10 trackers

4.3.1.2 基于数据集属性的性能分析

为充分评估算法的跟踪性能,利用 OTB-2015 视频数据集的 11 个属性进一步评估算法的性能。图 6 为排名前 10 的跟踪算法在包含 11 个属性数据集上的精确度图,标题上的数字表示此种属性的数据集数量。该算法在平面内旋转、平面外旋转、超出视野、复杂背景、低分辨率和尺度变化属性精度图上排名第一。相比 SRDCF 算法,在平面内旋转、平面外旋转、超出视野属性精度图上,分别提高 6.1%、4%和 6.5%。

图 7 为排名前 10 的跟踪算法在包含 11 个属性数据集上的成功率图,标题上的数字表示此种属性的数据集数量。除了光照变化属性之外,该算法在其余 10 个属性的成功率图上都排名第一。相比 SRDCF 算法,在平面内旋转、平面外旋转、超出视野属性成功率图上,该算法分别提高了 7.27%、4.9%和 10.2%。

对于 OTB-2015 平面内旋转属性的 50 个数据集和平面外旋转属性 63 个数据集,该算法在精确度图和成功率图上均排序第一,取得不错的改进效果。本文算法采用 CNN 特征,并利用自适应降维技术选择 CNN 特征主要成分,提高了处理旋转数据集的能力。

4.3.2 定性比较

4.3.2.1 抗旋转性能分析

图 8 显示了 10 个算法在三个经历旋转属性数据集上几帧代表性跟踪结果。在 MotorRolling 数据集中,目标经历了多次旋转,对跟踪算法产生巨大挑战,实验中只有该算法能够完美地跟踪目标。对于 Bolt 数据集,目标运动员经历了平面外旋转(第 19 帧和第 77 帧)和平面内旋转(第 223 帧和第 268 帧),SRDCF 算

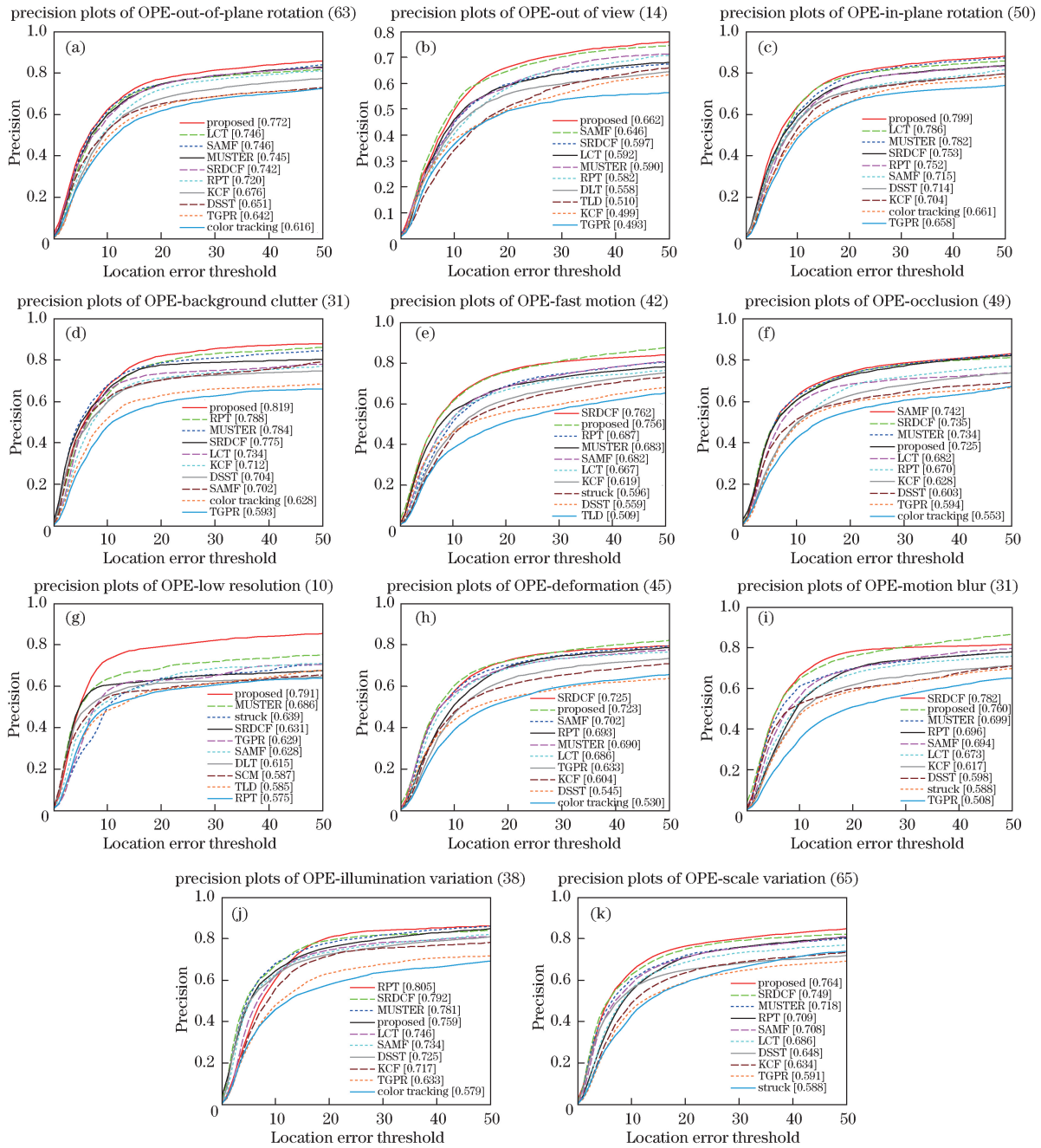


图 6 排名前 10 跟踪算法在 11 个属性的精确度图
 Fig. 6 Precision plots with 11 different attributes for the top 10 trackers

法在第 77 帧、第 223 帧和第 268 帧跟踪失败,而该算法跟踪正确。在 DragonBaby 数据集中,目标人脸经历了严重的平面内旋转和平面外旋转,产生不同角度的人脸,加大了跟踪难度。例如,第 31 帧、第 62 帧和第 93 帧目标人脸角度不同,甚至在第 48 帧目标人脸受到手臂的遮挡,导致 SRDCF 算法跟踪失败,而该算法利用包含一定语义信息的卷积特征能够有效地跟踪旋转目标。

4.3.2.2 抗遮挡性能分析

图 9 显示了 10 个算法在三个经历严重遮挡数据集上几帧代表性跟踪结果。在 Human3 数据集中,由于跟踪目标比较小,并且在第 47 帧目标被灯杆几乎完全遮挡模型更新错误,导致 SRDCF 算法跟踪失败。而该算法利用 $R_{PSR,t}$ 有效计算跟踪位置的置信度,有效地判断目标是否被遮挡,选择性地更新模型,避免在遮挡因素下产生错误模板更新导致目标漂移。在 Woman 数据集中,第 555 帧目标人的头部被大片树叶遮挡,SRDCF 算法跟踪失败,而该算法利用跟踪位置 $R_{PSR,t}$ 值判断目标是否被遮挡,自适应地更新模型,得到

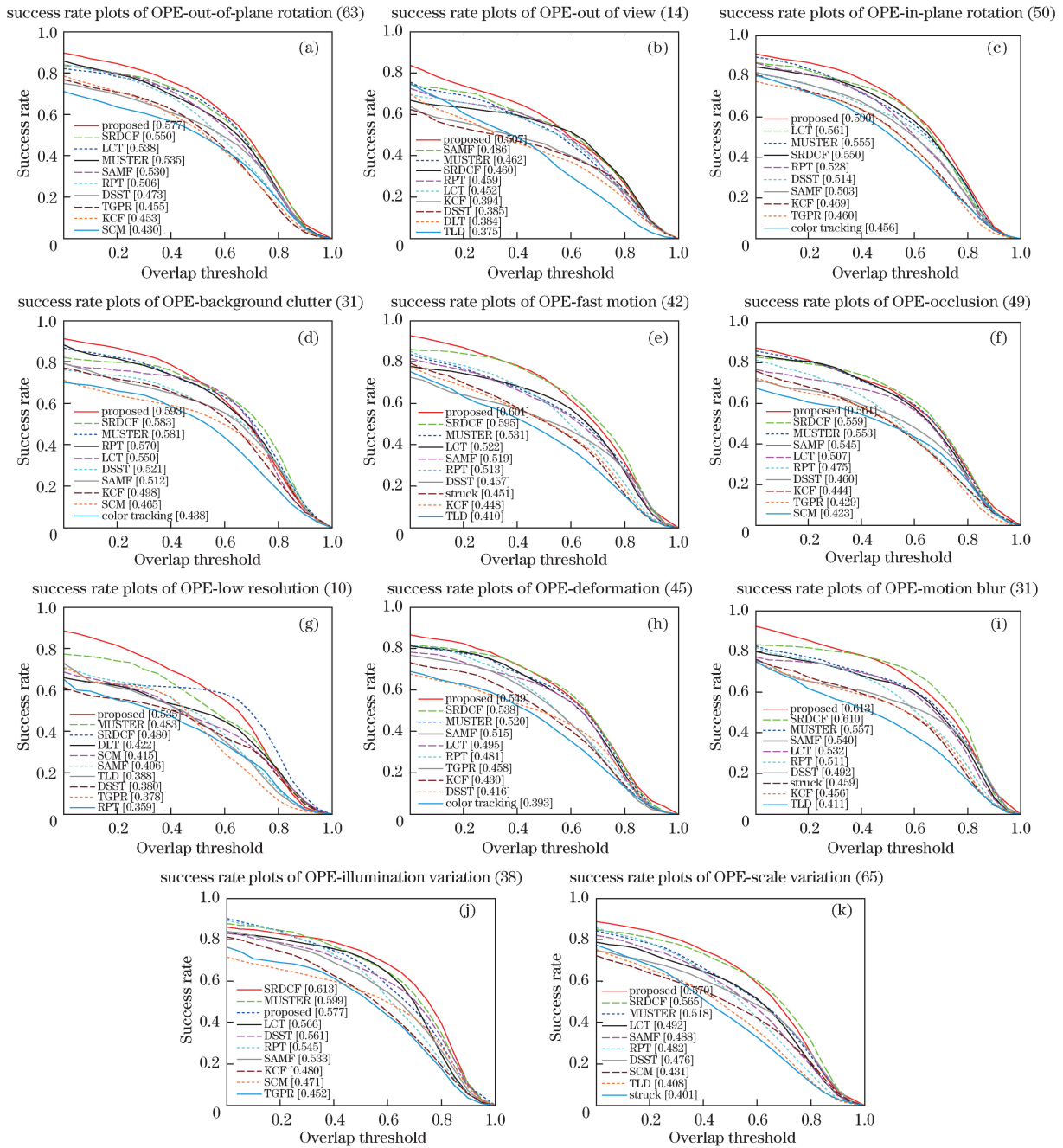


图 7 排名前 10 跟踪算法在 11 个属性的成功率图

Fig. 7 Success rate plots with 11 different attributes for the top 10 trackers

正确的跟踪结果。

4.3.2.3 超出视野性能分析

图 10 显示了长期跟踪算法在具有超出视野因素的数据集上代表性跟踪结果。由于超出视野数据集在 OTB-2015 评测数据集库中数量特别少,并且超出视野对跟踪影响不大,故选取 TLD 长期跟踪算法所提供的 Motocross 和 Carchase 数据集。在数据集 Motocross 中,目标经历了超出视野(如第 37 帧和第 491 帧)、外观严重变化(如第 422 帧和第 689 帧)影响,SRDCF 算法和 LCT 算法在第 37 帧之后就一直跟踪失败,而该算法都能正确跟踪。在 Carchase 数据集中,目标车辆经历完全超出视野一段时间又回到视野中(如第 382 帧和第 396 帧),该算法和 TLD 算法由于加入重定位组件,即使在目标完全超出视野一段时间后又回到视野中这种复杂情况下仍然能够跟踪正确,而 SRDCF 算法由于没有重定位模块才会在目标超出视野一段时间后又回到视野情况下,因此无法重新定位目标的位置而跟踪失败。

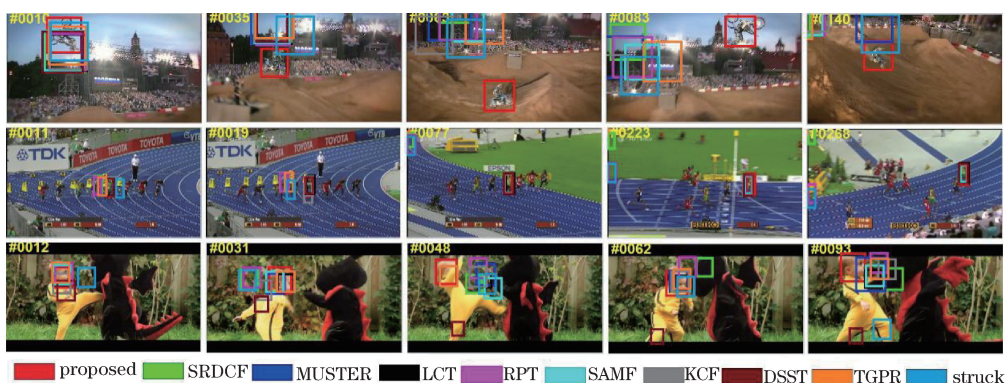


图 8 10 个跟踪算法在经历旋转的数据集 MotorRolling、Bolt、DragonBaby 上的定性结果显示
 Fig. 8 Qualitative results of the 10 trackers over rotating datasets MotorRolling, Bolt and DragonBaby

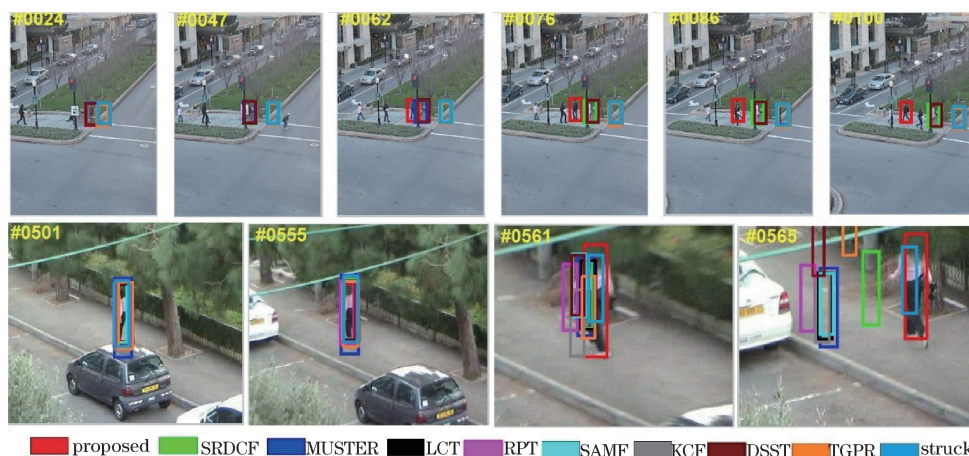


图 9 10 个跟踪算法在经历严重遮挡的数据集 Human3 和 Woman 上的定性结果显示
 Fig. 9 Qualitative results of the 10 trackers over heavily occluded datasets Human3 and Woman

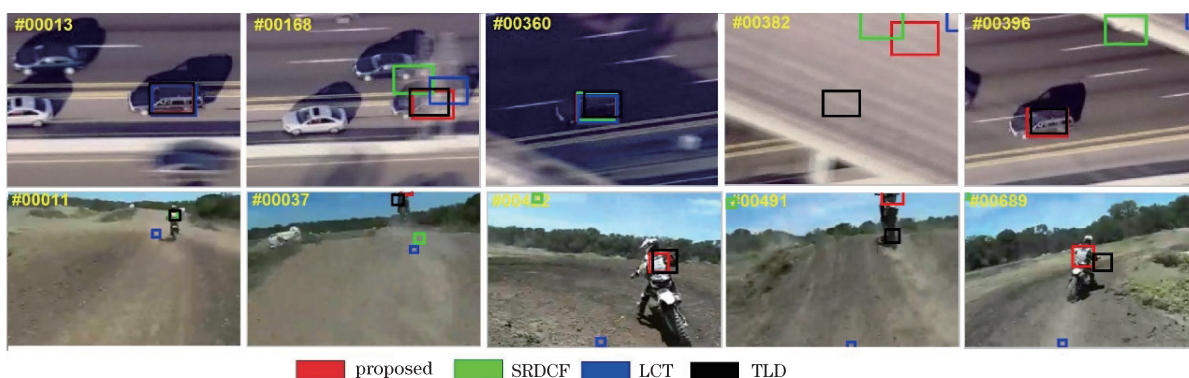


图 10 长期跟踪算法在经历超出视野的数据集 Motocross 和 Carchase 上的定性结果显示
 Fig. 10 Qualitative results of the long-term trackers over datasets Motocross and Carchase out of view

5 结 论

由于 SRDCF 算法在旋转变换、超出视野和严重遮挡复杂情况下容易跟踪失败,因而提出了一种基于自适应卷及特征的目标跟踪方法。选择 VGG-NET 模型中具有一定语义信息和丰富空间细节信息的 conv3-4 层卷积特征,并利用自适应降维技术将卷积特征由 256 维降至 130 维,有效地解决了旋转目标跟踪失败的问题。引入在线 SVM 重定位组件,解决了目标超出视野一段时间后回到视野中重新跟踪问题。另外,通过利

用 $R_{PSR,t}$ 值判断目标是否被遮挡,选取有效地跟踪结果更新模型,解决了遮挡因素下产生模型更新错误而导致跟踪漂移问题。利用 OTB-2015 的 100 个数据集进行实验,结果显示,该方法整体的精度为 0.804,成功率为 0.607,相比 SRDCF 算法分别提高了 1.9% 和 1.5%。通过基于数据集属性的定量分析和定性分析表明,在目标发生旋转变换、超出视野和严重遮挡等复杂场景下,与 38 种跟踪算法相比,该方法具有更强的稳健性。

参 考 文 献

- [1] Smeulders A W M, Chu D M, Cucchiara R, *et al.* Visual tracking: An experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1442-1468.
- [2] Ross D A, Lim J, Lin R S, *et al.* Incremental learning for robust visual tracking[J]. International Journal of Computer Vision, 2008, 77(1): 125-141.
- [3] Kwon J, Lee K M. Visual tracking decomposition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 1269-1276.
- [4] Liu Wei, Zhao Wenjie, Li Cheng, *et al.* Long-term visual tracking based on spatio-temporal context[J]. Acta Optica Sinica, 2016, 36(1): 0115001.
刘 威, 赵文杰, 李 成. 时空上下文学习长时目标跟踪[J]. 光学学报, 2016, 36(1): 0115001.
- [5] Liu Wei, Zhao Wenjie, Li Cheng, *et al.* An online learning visual tracking method based on compressive sensing[J]. Acta Optica Sinica, 2015, 35(9): 0915001.
刘 威, 赵文杰, 李 成. 一种基于压缩感知的在线学习跟踪算法[J]. 光学学报, 2015, 35(9): 0915001.
- [6] Hare S, Saffaria A, Torr P H S, *et al.* Struck: Structured output tracking with kernels [C]. IEEE International Conference on Computer Vision, 2011: 12490885.
- [7] Gao J, Ling H, Hu W, *et al.* Transfer learning based visual tracking with Gaussian processes regression[C]. European Conference on Computer Vision, 2014: 188-203.
- [8] Kalal Z, Mikolajczyk K, Matas J, *et al.* Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [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] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration[C]. European Conference on Computer Vision, 2014: 254-265.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), 2005: 886-893.
- [12] Ma C, Yang X, Zhang C, *et al.* Long-term correlation tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5388-5396.
- [13] Pan Zhenfu, Zhu Yongli. Kernelized correlation filters object tracking method with multi-scale estimation[J]. Laser & Optoelectronics Progress, 2016, 53(10): 101501.
潘振福, 朱永利. 多尺度估计的核相关滤波器目标跟踪方法[J]. 激光与光电子学进展, 2016, 53(10): 101501.
- [14] Zhang Jianwei. High-precision extended object tracking based on region feature matching[J]. Laser & Optoelectronics Progress, 2015, 52(5): 051004.
张建伟. 基于区域特征匹配的扩展目标高精度跟踪[J]. 激光与光电子学进展, 2015, 52(5): 051004.
- [15] Galoogahi H K, Sim T, Lucey S. Correlation filters with limited boundaries[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4630-4638.
- [16] Danelljan M, Häger G, Khan F, *et al.* Learning spatially regularized correlation filters for visual tracking[C]. IEEE International Conference on Computer Vision, 2015: 3074-3082.
- [17] Wu Y, Lim J, Yang M H, *et al.* Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [C/OL]. arXiv: 1409.1556, 2014. [2016-08-20]. <https://arxiv.org/pdf/1409.1556v6.pdf>.
- [19] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances

in Neural Information Processing Systems, 2012: 1097-1105.

- [20] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [21] Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM International Conference on Multimedia, 2014: 675-678.
- [22] Cimpoi M, Maji S, Vedaldi A. Deep convolutional filter banks for texture recognition and segmentation[J/OL]. arXiv preprint arXiv: 1411.6836, 2014. [2016-08-20]. <https://arxiv.org/pdf/1411.6836v2.pdf>.
- [23] Bolme D S, Beveridge J R, Draper B A, *et al.* Visual object tracking using adaptivess correlation filters[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.