中国科学: 信息科学 2020 年 第 50 卷 第 1 期: 128–150
SCIENTIA SINICA Informationis
论文

学习时空一致性相关滤波的视觉跟踪

朱建章[1*], 王栋[2], 卢湖川[2]

1. 河南财经政法大学数学与信息科学学院, 郑州 450046
2. 大连理工大学信息与通信工程学院, 大连 116024
* 通信作者. E-mail: zhujianzhang@126.com

收稿日期: 2018–09–05; 修回日期: 2019–01–26; 接受日期: 2019–02–15; 网络出版日期: 2020–01–08

国家自然科学基金 (批准号: 61502070, 61725202) 和河南省自然科学基金 (批准号: 18A110013) 资助项目

摘要 判别相关滤波跟踪算法通过对中心目标块 (唯一准确正样本) 循环移位获取训练集, 依赖潜在样本周期延拓假设, 使得模型训练和检测可以通过快速傅里叶变换高效完成, 然而整个学习过程没有对真正的背景信息进行建模. 背景感知相关滤波 (BACF) 跟踪算法利用一个二进制掩码矩阵通过密集采样的方法获取真正的正、负样本对目标外观进行建模, 然而 BACF 算法在学习相关滤波器时并没有考虑滤波器的时间一致性和空间一致性信息, 当目标出现外观突变时, 学习到的相关滤波器将会偏向背景而发生漂移. 为了解决学习到的相关滤波器适应连续帧之间的外观突变问题, 本文在基准 BACF 算法框架下引入时间一致性约束项和空间一致性约束项, 提出了学习时空一致性相关滤波 (TSCF) 跟踪算法. 时间一致性约束项在时间序列意义上起到平滑多通道相关滤波的作用; 空间一致性约束项在空间分布意义上平滑多通道相关滤波, 使得学习到的相关滤波能量分布更加均匀. 本文的 TSCF 模型有闭式解, 采用共轭梯度下降法迭代逼近模型的最优解, 且优化过程利用循环矩阵性质转化到傅里叶域快速求解, 有效降低计算大型矩阵的代价. 本文的 TSCF 算法跟踪结果在 TB100 公开数据库上显示, 距离精度较基准 BACF 算法提升了 5.5%, 成功率曲线图线下面积 (AUC) 提升了 4.3%, 纯手工特征跟踪性能在 TB100 数据库上 100 个视频的跟踪距离精度达到 0.879, AUC 为 0.664, 结果展示本文的 TSCF 算法在遇到诸如短时间遮挡和面内旋转或面外旋转等挑战性问题时具有一定的鲁棒性和有效性.

关键词 视觉跟踪, 相关滤波, 时空一致性, 正则化, 共轭梯度下降

1 引言

视觉目标跟踪作为计算机视觉的研究热点问题之一, 已在自动驾驶、视频监控、人机交互、智能导航和赛事直播等方面有着广泛应用, 它融合了图像处理、信号处理、人工智能, 以及数学等诸多领

引用格式: 朱建章, 王栋, 卢湖川. 学习时空一致性相关滤波的视觉跟踪. 中国科学: 信息科学, 2020, 50: 128–150, doi: 10.1360/N112018-00232
Zhu J Z, Wang D, Lu H C. Learning temporal-spatial consistency correlation filter for visual tracking (in Chinese). Sci Sin Inform, 2020, 50: 128–150, doi: 10.1360/N112018-00232

域的先进技术和核心思想. 视觉目标跟踪的任务是在视频序列中持续精准地估计目标的位置、形状变化或所占面积, 确定目标的运动速度、方向及轨迹等运动信息, 旨在完成更高级的任务. 近年来, 随着高性能图像处理器 (GPU) 和张量处理器 (TPU) 硬件设备及并行计算软件技术的迅猛发展, 视觉跟踪陆续取得了阶段性进展, 由于被跟踪目标存在着内因 (如: 面内旋转、面外旋转、非刚体变形和尺度变化) 和外因 (如: 光照变化、背景杂乱、运动模糊和遮挡) 等多方面因素的干扰, 建立一个通用、有效且鲁棒的视觉跟踪系统仍然面临着巨大挑战.

视觉跟踪算法根据外观模型的不同主要分为生成式模型和判别式模型两大类. 生成式模型在线学习目标自身外观模型, 采用不同的描述模型来提取不同的目标特征, 构建一个紧致的目标表示, 搜索重建误差最小的图像区域完成目标定位; 而判别式模型主要考虑目标与背景的区分性, 同时提取目标前景和背景信息来训练分类器, 旨在把目标从背景中有效地分离出来. TB100 [1] (即 OTB2015) 数据库的建造者在对跟踪算法进行评估时发现背景信息对构建有效跟踪系统十分重要, 而判别式模型同时考虑目标前景和背景信息, 从近几年的跟踪算法性能比较结果来看, 判别式模型的性能已经远远超过生成式模型, 本文研究对象是判别式模型.

基于判别相关滤波 (discriminative correlation filter, DCF) 的视觉跟踪算法已取得了显著性进展且几乎占领跟踪领域半壁江山, DCF 具有较高的跟踪精度同时还具备高效的实时性能, 受到众多科学研究者的广泛关注和深入研究. DCF 在空域内通过对前景目标的循环移位密集采样大量正、负样本作为训练集, 通过回归到高斯 (Gauss) 分布的软标签, 在最小化岭回归损失函数框架下求解相关滤波器. DCF 框架巧妙地将空域中的相关运算转化为频域中的点乘运算, 此操作得益于快速傅里叶 (Fourier) 变换性质而在频域内高效地对目标外观进行建模. 早在 2010 年, Bolme 等 [2] 首次将相关滤波方法引入到自适应视觉跟踪领域, 提出了平方误差最小滤波器 (minimum output sum of squared error, MOSSE) 跟踪算法, 该算法利用多帧跟踪样本同时训练一个相关滤波器且给出了其等价的傅里叶域的求解公式, 在训练分类器和定位目标时采用快速傅里叶变换, 使得跟踪速度可达到每秒 600 多帧. 然而 MOSSE 仅采用灰度图像特征信息作为训练滤波器的输入数据, 有限的特征信息导致跟踪性能不太理想. 2012 年, Henriques 等 [3] 在相关滤波框架下引入核函数机制, 提出了基于核技术 (circulant structure kernels, CSK) 的跟踪算法. 2014 年, Danelljan 等 [4] 改进了 CSK 跟踪算法, 提出了基于颜色空间多通道相关滤波 (color name, CN) 的跟踪算法, 即把 RGB 颜色空间扩展到 11 个通道的 CN 空间. 与此同时, Henriques 等 [5] 在核技术上改进了 CSK 跟踪算法, 提出了基于方向梯度直方图 (histogram of oriented gradient, HOG) 特征的多通道核相关滤波 (kernelized correlation filters, KCF) 的跟踪算法, 采用 31 个通道的 HOG 特征, 把多通道特征融入相关滤波框架, 结合岭回归与循环矩阵将相关滤波核化, 把输入特征映射到高维非线性空间来增加模型表达能力, 该方法对运动模糊、光照变化, 以及颜色变化都具有较强的鲁棒性. 虽然多通道相关滤波跟踪算法提高了跟踪的精度, 但是由于采用多通道特征而加大了计算量. 针对尺度变化对跟踪性能带来的负面影响, 2014 年, Li 等 [6] 提出了自适应尺度变化的核相关滤波器 (scale adaptive with multiple features tracker, SAMF) 跟踪算法, 该方法使用 7 个较粗的尺度在多尺度图像块上进行检测, 选取相关滤波响应最大值所对应的平移位置和目标尺度. 与此同时, Danelljan 等 [7,8] 提出了判别尺度空间 (discriminative scale space tracker, DSST) 跟踪算法, 该算法采用 33 个较精细的尺度训练平均滤波器和尺度滤波器来估计目标位置, 然后在检测到的跟踪位置处采用尺度滤波器估计目标尺度.

DCF 框架的训练集是通过对中心目标块循环移位获取的, 基于这种操作才能使得模型训练和检测可以通过快速傅里叶变换高效完成, 而对目标块的循环移位操作的前提条件是假设样本满足周期性延拓, 即中心目标块左和右、上和下边界能够完美对接, 但通常的自然影像图片是不满足这个性质的,

因此在中心图像块满足周期延拓假设的前提下, 图像边界将出现问题, 即边界效应. 由于训练集是通过对中心目标块循环移位获取的, 因此只有唯一的中心样本是正确的, 采用不准确的负样本对目标外观模型进行建模势必会降低模型的判别力. 上面提到的大多数算法仅在特征图上加入余弦窗操作来弱化边界效应的影响, 但是该余弦窗技术可能使跟踪器仅学到部分前景信息而屏蔽了背景信息, 从而降低了跟踪器的判别力. 为了解决边界效应问题, 2015 年, Galoogahi 等 [9] 采用较大尺寸的检测图像块和较小尺寸的滤波器且动态减少训练集中样本数量来提高真实样本的比例, 在新的目标函数下构造增广拉格朗日乘子 (augmented Lagrangian method, ALM), 并采用交替方向乘子法 (alternating direction method of multipliers, ADMM) 优化求最优解. 2015 年, Danelljan 等 [10] 提出了基于空间正则化约束的相关滤波 (spatially regularized discriminative correlation filters, SRDCF) 跟踪算法, 该方法对滤波器系数进行空间约束, 越靠近边缘约束越强, 使得学习到的滤波器更加集中在中心区域, 该算法是通过高斯 – 赛德尔 (Gauss-Seidel) 方法优化求解目标函数, 虽然跟踪器的性能有所提升, 但是采用手工特征 HOG 仅有 9 fps, 相比 KCF 跟踪速度慢了近 25 倍. 2016 年, Danelljan 等 [11] 提出了连续卷积相关滤波 (continuous convolution operators for visual tracking, CCOT) 跟踪算法, 该算法在傅里叶域求解每个卷积层对应的滤波器系数, 并在频域内经过双线性插值得到连续域的滤波器响应, 该双线性插值操作不会增加模型参数数量且减少过拟合现象. 该方法融合了多分辨率的特征图, 将传统的手工特征和不同分辨的深度特征相结合, 从而提高算法的性能, 在连续空域中学习连续相关滤波器, 可以达到子像素级的定位, 提升了算法定位精度. 为了剔除 CCOT 中存在的冗余信息, 在不失精度的前提下, 2017 年, Danelljan 等 [12] 提出了有效卷积运算 (efficient convolution operators, ECO) 跟踪算法来加速 CCOT, 主要从 3 个方面改进 CCOT 的性能, 即对卷积操作进行因式分解来减少模型参数, 通过高斯混合模型 (Gaussian mixture model, GMM) 简化了训练集且保证样本多样性, 改进模型更新策略. 2018 年, Li 等 [13] 提出了空间与时间正则化 (spatial-temporal regularized correlation filters, STRCF) 跟踪算法, 该算法指出 SRDCF 破坏了相关滤波的循环移位框架而使得优化问题变得困难, 在训练过程中记录初始帧到当前帧的所有样本信息, 由于提取样本特征需要耗费一定的时间, 且采用的高斯 – 赛德尔优化求解方法, 仍具有很高的计算复杂度, 为了提高 SRDCF 的跟踪速度, 在不失精度的前提下, 引入时间正则化项, 无需保留以往样本信息, 仅用纯手工特征就能达到实时跟踪效果.

为了缓解边界效应缺陷对跟踪器的影响, 进一步提升跟踪性能, 许多基于深度学习的算法被相继提出, 比如 DeepSRDCF [14], MDNet [15], CREST [16], CFCF [17], CFNet [18], ACFN [19], DRT [20], LSART [21], MKCF [22] 和 FlowTrack [23] 等, 由于这些算法使用深度学习框架提取深度特征, 所以带来了很高的计算复杂度, 虽然跟踪性能有了大幅度的提升, 但是实时性能却有所下降, 这恰好也是视觉跟踪实际应用的一个障碍.

在相关滤波学习过程中放弃真实背景信息可能会降低学习到的跟踪器的判别力, 为了克服这一缺陷, Galoogahi [24] 提出了背景感知相关滤波 (background-aware correlation filters, BACF) 跟踪算法, 利用一个掩码矩阵通过密集采样的方法获取真正的正、负样本对目标外观进行建模, 仅采用手工特征 (HOG 特征) 就达到很高的跟踪性能, 且满足实时 (33.9 fps) 需求. 然而 BACF 算法在学习相关滤波器时并没有考虑滤波器的时间一致性和空间一致性信息, 当目标出现外观突变时学习到的相关滤波器将会偏向背景而发生漂移, 这些突变现象诸如短时间部分遮挡或完全遮挡 (比如图 1 中 Box 在 470 帧附近出现短时间完全遮挡, 在 500 帧前后目标再现时, BACF 算法发生了漂移)、面内旋转或面外旋转 (比如图 1 中 Bird2 在 48 帧前后, 被跟踪目标的身体发生了 180° 旋转, BACF 算法在 88 前后发生了漂移) 和快速运动 (比如图 1 中的 BlurOwl 在 154 帧前后因快速运动伴随运动模糊, BACF 算法在 163 帧前后发生了漂移) 等.

图 1 (网络版彩图) 4 个跟踪算法在 TB100 基准数据库中的 3 个视频序列 (Box, Bird2 和 BlurOwl) 跟踪结果展示

Figure 1 (Color online) Example tracking results of four different methods on the TB100 dataset on three video sequences (Box, Bird2, and BlurOwl)

针对基准 BACF 算法中存在的不足之处, 本文的主要工作和创新点如下:

(1) 为了解决学习到的相关滤波器适应连续帧之间的外观突变问题, 本文的 TSCF (temporal-spatial consistency correlation filter) 算法在基准 BACF 算法框架下引入时间一致性约束项和空间一致性约束项. 时间一致性约束项在时间序列意义上起到平滑多通道相关滤波的作用, 有效避免了因连续帧之间外观突变而使得学习到的相关滤波偏向背景; 空间一致性约束项在空间分布意义上平滑多通道相关滤波, 使得学习到的相关滤波能量分布更加均匀, 避免学习到的相关滤波器偏向某一不可靠背景区域.

(2) 本文的 TSCF 算法模型有闭式解, 因闭式解方程组维度过高, 在没有引入任何辅助变量的前提下, 采用共轭梯度下降法 (conjugate gradient, CG) 迭代逼近闭式解方程组的最优解, 共轭梯度优化过程利用循环矩阵性质转化到傅里叶域快速求解, 有效降低计算大型矩阵的代价.

(3) 本文的 TSCF 算法跟踪结果在 TB100 公开数据库上显示, 仅采用共轭梯度下降法优化求解 (单样本学习策略) 与 BACF 跟踪性能相近; 仅采用共轭梯度下降法优化求解 (多样本学习策略), 距离精度 (distance precision, DP) 较基准 BACF 算法提升了 3.1%, 成功率曲线图线下面积 (area under the curve, AUC) 提升了 1.3%; 引入时间一致性约束项后采用共轭梯度下降优化方法求解, 距离精度较基准 BACF 算法提升了 3.9%, AUC 提升了 3.9%; 引入时空一致性约束项后采用共轭梯度下降优化方法求解, 距离精度较基准 BACF 算法提升了 5.5%, AUC 提升了 4.3%. 纯手工特征跟踪性能在 TB100 数据库上 100 个视频的距离精度达到 0.879, AUC 为 0.664, 结果展示本文的 TSCF 算法在遇到挑战性问题时具有一定的鲁棒性和有效性.

2 相关工作

判别相关滤波跟踪算法通过对中心目标块 (唯一准确正样本) 循环移位获取训练集, 依赖潜在的样本周期延拓假设, 使得模型训练和检测可以通过快速傅里叶变换高效完成, 接下来简要回顾判别相关滤波算法的基本原理与本文基准 BACF 算法, 更加详细的阐述参见文献 [24~26].

2.1 判别相关滤波 (DCF) 算法

判别相关滤波跟踪算法在空域内通过对前景目标的循环移位密集采样大量正、负样本作为训练集, 通过回归到高斯分布的软标签, 在最小化岭回归损失函数框架下求解相关滤波器. 判别相关滤波框架巧妙地将空域中的相关运算转化为频域中的点乘运算, 此操作得益于快速傅里叶变换性质而在频域内高效地对目标外观进行建模. 判别相关滤波跟踪器的目标是在空域里从一个训练样本集 $\{(x_m, y_m)\}_{m=1}^{M}$ 中学到一个多通道的相关滤波器 h. 其中训练样本集中的样本记为 $x_m = \{x_{d,m}\}_{d=1}^{D}$, $m = 1, \ldots, M$, M 是训练集中样本的个数, $x_{d,m}$ 表示从第 m 个样本中提取的第 d 个通道的特征图, 向

量化后的特征图仍记为 $x_{d,m} \in \mathbb{R}^K$, D 是特征通道的总个数, K 为特征图 $x_{d,m}$ 向量化后的长度, 多通道的相关滤波器记为 $h = \{h_d\}_{d=1}^D$, $h_d \in \mathbb{R}^K$ 是对应于 $x_{d,m}$ 的第 d 个通道的相关滤波器, $y_m \in \mathbb{R}^K$ 是一个标量值函数对应于 $x_{d,m}$ 的高斯标签. 那么多通道相关滤波器 h 关于第 m 个训练样本 x_m 的相关响应值为

$$S_h(x_m) = \sum_{d=1}^{D} h_d * x_{d,m}, \tag{1}$$

符号 $*$ 表示空域循环卷积, 那么通过求解最优化式 (2) 中岭回归的目标函数得到多通道相关滤波器 h:

$$E(h) = \sum_{m=1}^{M} \alpha_m \|y_m - S_h(x_m)\|_2^2 + \frac{\lambda}{2} \sum_{d=1}^{D} \|h_d\|_2^2. \tag{2}$$

每个训练样本的权重为 $\alpha_m \geqslant 0$, 这里的 λ 表示正则化因子. 式 (2) 是一个线性最小二乘问题, 由帕塞瓦尔定理 (Parseval's formula) 可以转换到傅里叶域快速求解.

在检测阶段, 记 $\{z_d \in \mathbb{R}^K\}_{d=1}^D$ 表示从新一帧图像块中提取的向量化后的 D 个特征通道的特征图, 则每一个目标块位置对应的分类器响应得分值 $S_h(z)$ 可以通过式 (3) 得到.

$$S_h(z) = \mathscr{F}^{-1}\left(\sum_{d=1}^{D} \mathscr{F}(h_d) \odot \mathscr{F}(z_d)^{\mathrm{H}}\right), \tag{3}$$

其中符号 $\odot$ 表示为向量或矩阵对应元素点乘, $\mathscr{F}(h_d)$ 和 $\mathscr{F}(z_d)$ 分别表示 h_d 和 z_d 的离散傅里叶变换 (FFT), $\mathscr{F}^{-1}(\cdot)$ 表示逆离散傅里叶变换函数 (IFFT), 在频域中得益于快速傅里叶变换的计算优势, 计算复杂度仅为 $O(DK\log(K))$, 最终通过求式 (3) 的最大分类器响应得分值所对应的位置来定位被跟踪的目标.

2.2 背景感知相关滤波 (BACF) 算法

判别相关滤波跟踪算法的训练集是通过对中心目标块循环移位获取的, 这种学习方式依赖于潜在的样本周期性延拓假设, 该假设使得模型训练和检测可以通过快速傅里叶变换高效完成, 但同时带来了隐含在相关滤波框架中的容易产生过拟合现象的缺陷, 即边界效应. 由于训练集是通过对中心目标块循环移位获取的, 因此只有唯一的中心目标块是正确的, 采用不准确的负样本对外观模型进行建模势必会降低模型的判别力. 而 BACF 算法利用一个二进制掩码矩阵通过密集采样的方法获取真正的正、负样本对目标外观进行建模, 该方法甚至可以对整幅图像进行目标搜索. 学习多通道 BACF 相关滤波目标函数为

$$E(h) = \frac{1}{2} \sum_{k=1}^{K} \left\| y(k) - \sum_{d=1}^{D} h_d^{\mathrm{H}} P x_d\left[\Delta\tau_k\right] \right\|_2^2 + \frac{\lambda}{2} \sum_{d=1}^{D} \|h_d\|_2^2, \tag{4}$$

其中 $x_d \in \mathbb{R}^K$ 表示从当前帧整幅图像中提取的向量化后的第 d 个通道的特征图, K 是特征图 x_d 向量化后的长度; $x_d[\Delta\tau_k]$ 表示对信号 x_d 的第 k 步离散相关移位; P 是一个 $H \times K$ 二进制矩阵, 矩阵 P 的作用是从整幅图像的特征图 x_d 中剪切出来 H 个元素, 跟踪目标块向量化后的元素个数记为 H, 通常情况下 $H \ll K$; 上标 H 为矩阵或向量的共轭转置. BACF 采用有效的交替方向乘子方法 (ADMM) 来交替优化求解式 (4) 得到多通道相关滤波器 h, 计算复杂度仅有 $O(LDK\log(K))$, 其中 L 为 ADMM 的迭代次数.

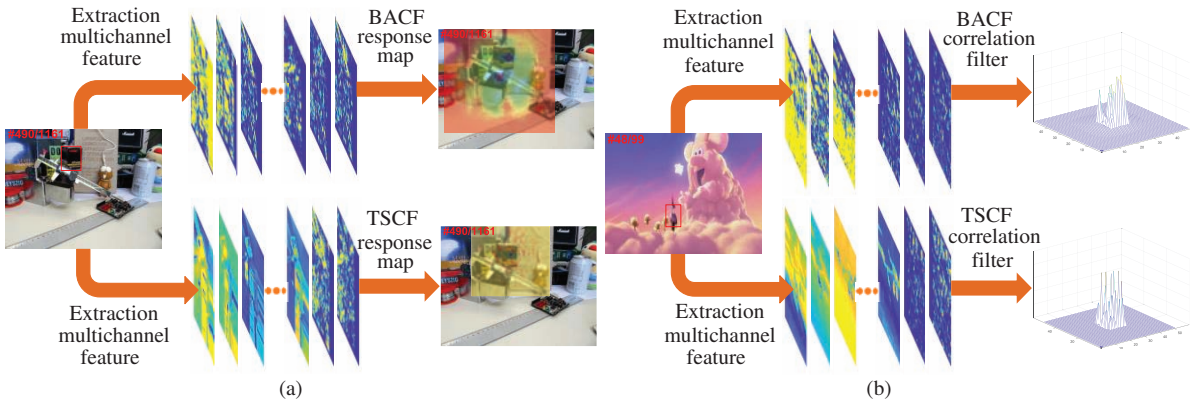


图 2 (网络版彩图) **BACF 与 TSCF 算法在 Box 和 Bird2 视频序列上的对比结果. (a) 视频序列 Box 分类器响应峰值对比; (b) 视频序列 Bird2 学习到的相关滤波器对比**

Figure 2 (Color online) Comparison between BACF and TSCF algorithms on sequences Box and Bird2. (a) Comparison of classifier response peak on sequence Box; (b) comparison of correlation filters learned on sequence Bird2

3 本文算法

3.1 时空一致性相关滤波 (TSCF) 模型

背景感知相关滤波 (BACF) 算法利用一个掩码矩阵通过密集采样的方法获取真正的正、负样本对目标外观进行建模, 已得良好的跟踪效果. 然而 BACF 算法在学习相关滤波器时并没有考虑滤波器的时间一致性和空间一致性信息, 当目标出现外观突变时, 学习到的相关滤波器将会偏向背景而发生漂移. 图 2(a) 中的视频序列 Box 在 470 帧左右出现短时间完全遮挡, 在 490 帧左右目标再次出现时, BACF 算法学习到的多通道相关滤波器的分类器响应图出现了两个峰值, 且右上方的分类器响应峰值小于左下方的, 这时 BACF 算法判定左下方为检测目标中心 (判断失误), 此时跟踪器发生了漂移. 究其原因是 BACF 算法不能很好地解决短时遮挡问题, 尤其是目标出现短时间完全遮挡, BACF 算法学习到的滤波器极可能偏向背景区域使得跟踪失败. 为了解决学习到的相关滤波器适应连续帧之间的外观突变问题, 本文的 TSCF 算法在基准 BACF 算法框架下引入时间一致性约束项 $\sum_{d=1}^{D}\|h_d-h_d^{(t-1)}\|_2^2$, 在时间序列意义上起到平滑多通道相关滤波的作用, 从图 2(a) 中的视频序列 Box 的 TSCF 算法的响应图可以看出, 引入时间一致性约束后有效避免了因连续帧之间外观突变而学习到的相关滤波偏向背景区域的情况; 图 2(b) 中的视频序列 Bird2 在 48 帧前后, 被跟踪目标出现 180° 的旋转, 从图 2 中可以看出 BACF 相关滤波器的能量分布偏向右侧, 为了解决学习到的滤波器能量分布不均的问题, 本文的 TSCF 算法在基准 BACF 算法框架下引入空间一致性约束项, 如下式:

$$\sum_{m=1}^{M}\alpha_m\left(\sum_{k=1}^{K}\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]\left(p_d^u\odot h_d\right)-\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]\left(p_d^v\odot h_d\right)\right\|_2^2\right). \tag{5}$$

在空间分布意义上平滑多通道相关滤波, 使得学习到的相关滤波器能量分布更加均匀, 从图 2(b) 中的视频序列 Bird2 的 TSCF 算法相关滤波器可以看出, 引入空间一致性约束项后可以有效避免学习到的相关滤波器偏向某一不可靠背景区域. 随后本文构造了时空一致性多通道相关滤波目标函数, 定

义如下:

$$\begin{aligned}E(h)=&\frac{1}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{k=1}^{K}\left\|y_m(k)-\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]\left(p_d\odot h_d\right)\right\|_2^2\right)+\frac{\lambda}{2}\sum_{d=1}^{D}\left\|h_d\right\|_2^2\\&+\frac{\varpi}{2}\sum_{d=1}^{D}\left\|h_d-h_d^{(t-1)}\right\|_2^2\\&+\frac{\eta}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{k=1}^{K}\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]\left(p_d^u\odot h_d\right)-\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]\left(p_d^v\odot h_d\right)\right\|_2^2\right).\end{aligned}\tag{6}$$

多通道相关滤波目标函数式 (6) 中符号 M 为训练集中样本的个数; $\alpha_m>0$ 为对应样本的权重衰减因子, 距离当前样本越近赋予权重越大; 符号 D 为特征通道的个数; $x_{d,m}\in\mathbb{R}^K$ 表示从第 m 个训练样本的整幅图像中提取的第 d 个通道向量化后的特征图, 其中 K 为特征图 $x_{d,m}$ 向量化后的长度; $x_{d,m}[\Delta\tau_k]$ 表示对信号 $x_{d,m}$ 的第 k 步离散相关移位; $y_m(k)$ 是标量值函数对应于 $x_{d,m}[\Delta\tau_k]$ 的高斯标签; $p_d\in\mathbb{R}^K$ 是对应于第 d 个通道的二进制掩码向量, 向量 p_d 的作用是从整幅图像的特征图 $x_{d,m}$ 中剪切出来 H 个元素, 跟踪目标块的元素个数记为 H, 通常情况下 $H\ll K$, 不同通道的二进制掩码向量 $p_d\in\mathbb{R}^K, 1\leqslant d\leqslant D$ 是相同的, 目的是把跟踪目标剪切出来. 上标 H 为矩阵或向量的共轭转置; $\{h_d\}_{d=1}^D$ 表示当前帧学习到的多通道相关滤波; $\{h_d^{(t-1)}\}_{d=1}^D$ 表示通过前一帧信息学习到的多通道相关滤波; λ 是避免模型退化的正则化参数; ϖ 表示时间一致性约束项因子; η 表示空间一致性约束项因子; N 表示空间一致性分块个数. 时空一致性多通道相关滤波目标函数可以改写成如下形式, 即

$$\begin{aligned}E(h)=&\frac{1}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{k=1}^{K}\left\|y_m(k)-\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]P_dh_d\right\|_2^2\right)+\frac{\lambda}{2}\left\|h\right\|_2^2+\frac{\varpi}{2}\left\|h-h^{(t-1)}\right\|_2^2\\&+\frac{\eta}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{k=1}^{K}\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]P_d^uh_d-\sum_{d=1}^{D}x_{d,m}^{\mathrm{H}}\left[\Delta\tau_k\right]P_d^vh_d\right\|_2^2\right)\\=&\frac{1}{2}\sum_{m=1}^{M}\alpha_m\left(\left\|y_m-\sum_{d=1}^{D}X_{d,m}P_dh_d\right\|_2^2\right)+\frac{\lambda}{2}\left\|h\right\|_2^2+\frac{\varpi}{2}\left\|h-h^{(t-1)}\right\|_2^2\\&+\frac{\eta}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|\sum_{d=1}^{D}X_{d,m}P_d^uh_d-\sum_{d=1}^{D}X_{d,m}P_d^vh_d\right\|_2^2\right)\\=&\frac{1}{2}\sum_{m=1}^{M}\alpha_m\left(\left\|y_m-X_mPh\right\|_2^2\right)+\frac{\lambda}{2}\left\|h\right\|_2^2+\frac{\varpi}{2}\left\|h-h^{(t-1)}\right\|_2^2\\&+\frac{\eta}{2}\sum_{m=1}^{M}\alpha_m\left(\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|X_mP^uh-X_mP^vh\right\|_2^2\right)\\=&\frac{1}{2}\left\|\Gamma^*Y-\Gamma^*XPh\right\|_2^2+\frac{\lambda}{2}\left\|h\right\|_2^2+\frac{\varpi}{2}\left\|h-h^{(t-1)}\right\|_2^2\\&+\frac{\eta}{2}\sum_{u=1}^{N}\sum_{v=1}^{N}\left\|\Gamma^*XP^uh-\Gamma^*XP^vh\right\|_2^2\\=&f_1\left(h;X\right)+f_2(h)+f_3\left(h;h^{(t-1)}\right)+f_4(h;X),\end{aligned}\tag{7}$$

其中 $P_d=\mathrm{diag}(p_d(1),p_d(2),\ldots,p_d(K))\in\mathbb{R}^{K\times K}$, $P_d^u=\mathrm{diag}(p_d^u(1),p_d^u(2),\ldots,p_d^u(K))\in\mathbb{R}^{K\times K}$ 和 $P_d^v=$

$\mathrm{diag}(p_d^v(1), p_d^v(2), \ldots, p_d^v(K)) \in \mathbb{R}^{K\times K}$ 分别为 $K\times K$ 的对角矩阵; $p_d^u \in \mathbb{R}^K; 1 \leqslant u \leqslant N$ 和 $p_d^v \in \mathbb{R}^K; 1 \leqslant v \leqslant N$ 分别对应于第 u 和 v 个分块的第 d 个通道的二进制掩码向量, 不同通道的二进制掩码向量 $p_d^u \in \mathbb{R}^K$, $1 \leqslant d \leqslant D$ 是相同的, 目的是把跟踪目标的第 u 个分块剪切出来. $P = P_1 \oplus P_2 \oplus \cdots \oplus P_D \in \mathbb{R}^{DK\times DK}$, $P^u = P_1^u \oplus P_2^u \oplus \cdots \oplus P_D^u \in \mathbb{R}^{DK\times DK}$ 和 $P^v = P_1^v \oplus P_2^v \oplus \cdots \oplus P_D^v \in \mathbb{R}^{DK\times DK}$ 分别为 $DK\times DK$ 的块对角矩阵; $X_{d,m} = [x_{d,m}[\Delta\tau_1], x_{d,m}[\Delta\tau_2], \ldots, x_{d,m}[\Delta\tau_K]]^{\mathrm{H}} \in \mathbb{R}^{K\times K}$ 是对应于第 m 个样本的第 d 个通道的循环矩阵; $X_m = [X_{1,m}, X_{2,m}, \ldots, X_{D,m}] \in \mathbb{R}^{K\times DK}$ 是第 m 个样本的 D 个通道所对应的循环矩阵的级联, M 个矩阵 $\{X_m\}_{m=1}^M$ 级联后记为 $X = [X_1^{\mathrm{H}}, X_2^{\mathrm{H}}, \ldots, X_M^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{R}^{MK\times DK}$; $\Gamma^* = \sqrt{\alpha_1}I_K \oplus \sqrt{\alpha_2}I_K \oplus \cdots \oplus \sqrt{\alpha_M}I_K \in \mathbb{R}^{MK\times MK}$ 是块对角矩阵, 第 m 个块 $\sqrt{\alpha_m}I_K$ 是一个 $K\times K$ 的矩阵, $\Gamma = \Gamma^{*\mathrm{H}}\Gamma^*$, $h = [h_1^{\mathrm{H}}, h_2^{\mathrm{H}}, \ldots, h_D^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{R}^{DK}$ 是一个 KD 维的向量; $Y = [y_1^{\mathrm{H}}, y_2^{\mathrm{H}}, \ldots, y_M^{\mathrm{H}}]^{\mathrm{H}} \in \mathbb{R}^{MK}$ 是 M 个样本所对应高斯标签的级联.

式 (7) 中 $f_1(h;X) = \frac{1}{2}\|\Gamma^*Y - \Gamma^*XPh\|_2^2$ 为数据项, 它反映了学习到的模型与期望模型之间的误差; $f_2(h) = \frac{\lambda}{2}\|h\|_2^2$ 为正则化项, 它起到有效减弱模型退化的作用; $f_3(h;h^{(t-1)}) = \frac{\varpi}{2}\|h - h^{(t-1)}\|_2^2$ 为时间一致性约束项, 是为了解决连续帧之间的外观突变问题而引入的, 它在时间序列意义上起到平滑多通道相关滤波的作用; $f_4(h;X) = \frac{\eta}{2}\sum_{u=1}^N\sum_{v=1}^N \|\Gamma^*XP^uh - \Gamma^*XP^vh\|_2^2$ 为空间一致性约束项, 是为了解决学习到的滤波器能量分布不均的问题而引入的, 它在空间分布意义上起到平滑多通道相关滤波的作用, 使得学习到的相关滤波器能量分布更加均匀.

3.2 模型求解

从式 (7) 可以看出, 目标函数 $E(h)$ 是 4 个二范数的累加, 因此目标函数 $E(h)$ 关于多通道相关滤波 h 是一个凸光滑可微函数, 因此有全局最优解, 接下来求解 $E(h)$ 关于 h 的偏导函数, 并令 $E(h) = 0$, 于是有

$$\frac{\partial E(h)}{\partial h} = \frac{\partial f_1(h;X)}{\partial h} + \frac{\partial f_2(h)}{\partial h} + \frac{\partial f_3(h;h^{(t-1)})}{\partial h} + \frac{\partial f_4(h;X)}{\partial h} = 0, \tag{8}$$

其中,

$$\begin{aligned}\frac{\partial f_1(h;X)}{\partial h} &= \frac{1}{2}\frac{\partial\|\Gamma^*Y - \Gamma^*XPh\|_2^2}{\partial h}\\ &= \frac{1}{2}\frac{\partial(h^{\mathrm{H}}P^{\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*XPh - 2h^{\mathrm{H}}P^{\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*Y - Y^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*Y)}{\partial h}\\ &= P^{\mathrm{H}}X^{\mathrm{H}}\Gamma XPh - P^{\mathrm{H}}X^{\mathrm{H}}\Gamma Y,\end{aligned} \tag{9}$$

$$\frac{\partial f_2(h)}{\partial h} = \lambda h, \tag{10}$$

$$\frac{\partial f_3\left(h;h^{(t-1)}\right)}{\partial h} = \varpi\left(h - h^{(t-1)}\right), \tag{11}$$

$$\begin{aligned}\frac{\partial f_4\left(h;X\right)}{\partial h} &= \frac{\eta}{2}\frac{\partial\sum_{u=1}^N\sum_{v=1}^N\|\Gamma^*XP^uh - \Gamma^*XP^vh\|_2^2}{\partial h}\\ &= \frac{\eta}{2}\frac{\partial\sum_{u=1}^N\sum_{v=1}^N\left(\Gamma^*XP^uh - \Gamma^*XP^vh\right)^{\mathrm{H}}\left(\Gamma^*XP^uh - \Gamma^*XP^vh\right)}{\partial h}\\ &= \frac{\eta}{2}\frac{\partial\sum\limits_{u=1}^N\sum\limits_{v=1}^N\begin{pmatrix}h^{\mathrm{H}}P^{u\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*XP^uh - h^{\mathrm{H}}P^{u\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*XP^vh\\ -h^{\mathrm{H}}P^{v\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*XP^uh + h^{\mathrm{H}}P^{v\mathrm{H}}X^{\mathrm{H}}\Gamma^{*\mathrm{H}}\Gamma^*XP^vh\end{pmatrix}}{\partial h}\end{aligned}$$

$$
\begin{aligned}
&= \frac{\eta}{2}\frac{\partial \sum\limits_{u=1}^{N}\sum\limits_{v=1}^{N}\begin{pmatrix} h^{\mathrm{H}}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - h^{\mathrm{H}}P^{u}X^{\mathrm{H}}\Gamma XP^{v}h \\ -h^{\mathrm{H}}P^{v}X^{\mathrm{H}}\Gamma XP^{u}h + h^{\mathrm{H}}P^{v}X^{\mathrm{H}}\Gamma XP^{v}h \end{pmatrix}}{\partial h} \\
&= \frac{\eta}{2}\frac{\begin{pmatrix} \partial\sum_{u=1}^{N}\sum_{v=1}^{N} h^{\mathrm{H}}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - \partial\sum_{u=1}^{N}\sum_{v=1}^{N} h^{\mathrm{H}}P^{u}X^{\mathrm{H}}\Gamma XP^{v}h \\ -\partial\sum_{u=1}^{N}\sum_{v=1}^{N} h^{\mathrm{H}}P^{v}X^{\mathrm{H}}\Gamma XP^{u}h + \partial\sum_{u=1}^{N}\sum_{v=1}^{N} h^{\mathrm{H}}P^{v}X^{\mathrm{H}}\Gamma XP^{v}h \end{pmatrix}}{\partial h} \\
&= 2\eta\left(N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - \sum_{u=1}^{N}\sum_{v=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{v}h\right) \\
&= 2\eta\left[N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - \left(\sum_{u=1}^{N}P^{u}\right)X^{\mathrm{H}}\Gamma X\left(\sum_{v=1}^{N}P^{v}\right)h\right] \\
&= 2\eta\left(N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - PX^{\mathrm{H}}\Gamma XPh\right).
\end{aligned} \tag{12}
$$

式 (8) 等价于下式

$$
\begin{aligned}
\frac{\partial E(h)}{\partial h} &= PX^{\mathrm{H}}\Gamma XPh - PX^{\mathrm{H}}\Gamma Y + \lambda h + \varpi\left(h - h^{(t-1)}\right) + 2\eta\left(N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u}h - PX^{\mathrm{H}}\Gamma XPh\right) \\
&= \left[PX^{\mathrm{H}}\Gamma XP + 2\eta\left(N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u} - PX^{\mathrm{H}}\Gamma XP\right) + (\varpi+\lambda)\,I_{DK}\right]h - PX^{\mathrm{H}}\Gamma Y - \varpi h^{(t-1)} \\
&= 0,
\end{aligned} \tag{13}
$$

整理上述方程得

$$
Ah = b, \tag{14}
$$

其中 $A = (1-2\eta)PX^{\mathrm{H}}\Gamma XP + 2\eta N\sum_{u=1}^{N}P^{u}X^{\mathrm{H}}\Gamma XP^{u} + (\varpi+\lambda)I_{DK}$, $b = PX^{\mathrm{H}}\Gamma Y + \varpi h^{(t-1)}$. 显然矩阵 A 是一个对称正定矩阵, 因此可以采用预条件共轭梯度下降法 (preconditioned conjugate gradient algorithm, PCG) 迭代逼近式 (14) 的解, 共轭梯度第 $k-1$ 步迭代搜索的方向为 $p^{(k-1)}$, 则可以通过如下迭代过程完成滤波器 h 的更新:

$$
\begin{cases}
r^{(k)} = r^{(k-1)} - \alpha_{k-1}Ap^{(k-1)}, \\
\beta_k = \dfrac{r^{(k)\mathrm{H}}r^{(k)}}{r^{(k-1)\mathrm{H}}r^{(k-1)}}, \\
p^{(k)} = r^{(k)} + \beta_k p^{(k-1)}, \\
\alpha_k = \dfrac{r^{(k)\mathrm{H}}r^{(k)}}{p^{(k)\mathrm{H}}Ap^{(k)}}, \\
h^{(k+1)} = h^{(k)} + \alpha_k p^{(k)}.
\end{cases} \tag{15}
$$

从式 (15) 可以看出, 迭代搜索的方向为 $p^{(k)}$ 后, 共轭梯度迭代的计算量主要集中在计算 $Ap^{(k)}$ 上. 由式 (14) 可知矩阵 A 的前两项具有相似的结构, 为简单起见, 本文记为 $B = PX^{\mathrm{H}}\Gamma XP$. 下面讨论

$Bp^{(k)} = PX^{\mathrm{H}}\Gamma XPp^{(k)}$ 在傅里叶域的快速计算方法:

$$Bp^{(k)} = PX^{\mathrm{H}}\Gamma XPp^{(k)} = \sum_{m=1}^{M}\alpha_m \begin{bmatrix} P_1\mathscr{F}^{-1}\left[\mathscr{F}\left(x_{1,m}\right)\odot\sum\limits_{d=1}^{D}\mathscr{F}(x_{d,m})^{\mathrm{H}}\odot\mathscr{F}\left(P_d p_d^{(k)}\right)\right] \\ P_2\mathscr{F}^{-1}\left[\mathscr{F}\left(x_{2,m}\right)\odot\sum\limits_{d=1}^{D}\mathscr{F}(x_{d,m})^{\mathrm{H}}\odot\mathscr{F}\left(P_d p_d^{(k)}\right)\right] \\ \vdots \\ P_D\mathscr{F}^{-1}\left[\mathscr{F}\left(x_{D,m}\right)\odot\sum\limits_{d=1}^{D}\mathscr{F}(x_{d,m})^{\mathrm{H}}\odot\mathscr{F}\left(P_d p_d^{(k)}\right)\right] \end{bmatrix}, \tag{16}$$

其中 $p^{(k)} = [p_1^{(k)\mathrm{H}}, p_2^{(k)\mathrm{H}}, \cdots, p_D^{(k)\mathrm{H}}]^{\mathrm{H}} \in \mathbb{R}^{DK}$, $p_d^{(k)} \in \mathbb{R}^K$ 是对应于第 d 个通道的子集; $x_{d,m} \in \mathbb{R}^K$ 表示从第 m 个训练样本的整幅图像中提取的第 d 个通道向量化后的特征图 (此处对应于没有进行循环移位的原始输入特征), $\mathscr{F}(x_{d,m})$ 表示对特征图 $x_{d,m}$ 的离散傅里叶变换, $\mathscr{F}^{-1}(x_{d,m})$ 表示对特征图 $x_{d,m}$ 的逆离散傅里叶变换, 上标 H 为矩阵或向量的共轭转置, 式 (15) 的计算复杂度为 $O(DK\log(K))$.

3.3 样本更新

大多数基于判别相关滤波的跟踪算法采用线性差值进行样本更新 (如 BACF [24]), 即 $x_t \leftarrow (1-lr)\times x_{t-1} + lr \times x_t$, 这里 x_t 为当前帧提取的特征信息, x_{t-1} 为前一帧更新后的特征信息, lr 为学习率, 这种更新策略虽然简单, 但强烈依赖于学习率参数 lr, 随时间推移训练样本很快丢掉以前的信息而导致模型退化. 另外一类常用的样本更新策略是收集离当前帧最近的一些样本作为训练集 (如 CCOT [11]), 给出当前帧附近样本集 $\{x_m\}_{m=1}^{M}$ 及指数衰减因子 $\{\alpha_m \sim (1-lr)^{M-m}\}_{m=1}^{M}$, 当训练样本集中样本个数超过最大阈值 $M_{\max}$ 后, 新的样本将取代训练样本集 $\{\alpha_m x_m\}_{m=1}^{M}$ 中权重最小值所对应的样本, 这种策略需要设置一个阈值 $M_{\max}$ (通常 $M_{\max} = 400$), 且更新后训练样本集包含了很多冗余信息. 本文采用 ECO [12] 样本更新策略, 利用高斯混合模型 (GMM) 进行样本融合, GMM 公式为

$$p(x) = \sum_{t=1}^{T_{\max}}\pi_t N(x;\mu_t;I), \tag{17}$$

其中 $T_{\max}$ 指的是高斯混合模型中高斯函数的个数, 也就是样本分组个数; π_t 为对应于第 t 个组的权重, μ_t 为对应于第 t 个组的均值, 为简单起见方差记为 I. 那么 GMM 更新方案是: x_m 为当前帧提取的特征信息, 我们构造一个新组 $\{\pi_i = lr; \mu_i = x_m\}$, 当组数大于给定的阈值 $T_{\max}$ 后, 如果所有组对应权重的最小值小于预先给定的阈值时, 那么抛弃最小的那个权重所对应的组, 让新组 $\{\pi_i = lr; \mu_i = x_m\}$ 填充这个空缺; 否则, 通过计算所有组之间的距离, 即 $\{\|\mu_k - \mu_l\|; k, l = 1, \ldots, T_{\max}\}$, 找出距离最近的两个组 k 和 l, 通过式 (18) 合并距离最近的组 k 和 l 为组 n, 把新组 $\{\pi_i = lr; \mu_i = x_m\}$ 填入空缺.

$$\begin{cases} \pi_n = \pi_k + \pi_l, \\ \mu_n = \dfrac{\pi_k\mu_k + \pi_l\mu_l}{\pi_k + \pi_l}. \end{cases} \tag{18}$$

3.4 目标定位

我们通过前一帧信息学习到的多通道相关滤波 $\{h_d^{(t-1)}\}_{d=1}^{D}$ 来定位当前帧目标位置和尺度, 通过

求分类器响应结果

$$S\left(z^{r}\right)=\mathscr{F}^{-1}\left(\sum_{d=1}^{D}\mathscr{F}\left(h_{d}^{(t-1)}\right)\odot\mathscr{F}\left(z_{d}^{r}\right)^{\mathrm{H}}\right) \tag{19}$$

的最大值在当前帧 t 上定位目标, 其中上标 $t-1$ 指的是前一帧目标. 本文和 ECO [12] 一样, 本文采用多分辨率搜索策略来估计目标变换的尺度, 即以前一帧目标为中心, 以 α^{γ} 为尺度提取

$$\{z^{r}\}_{r\in\{\lfloor\frac{1-S}{2}\rfloor,\ldots,\lfloor\frac{S-1}{2}\rfloor\}},$$

这里 S 表示为多分辨率尺度个数, α 为尺度增量因子, 并采取两步搜索策略来精确定位目标, 最后以 S 个不同尺度分类器响应值的最高得分作为最终目标位置和尺度变换的检测结果.

4 实验结果分析与性能评价

本文 TSCF 算法的性能测试是在基准数据库 TB100 和 TB50 上采用相同的参数进行的. 基准数据库 TB100 (即 OTB2015) 是视觉跟踪算法的主要测评数据库且是使用频率最高的数据库, 共包含 100 个视频序列; 数据库 TB50 指的是在数据库 TB100 中选取相对具有挑战性的 50 个视频序列, 此处 TB50 中的 50 个视频序列和 OTB2013 中的 50 个视频序列不尽相同. 为了能够客观地评测分析不同跟踪算法的优劣性, TB100 数据库的建造者按照目标的外观变化情况把 100 个视频序列标注为 11 个不同的属性, 它们分别为: 光照变化 (illumination variation, IV)、尺度变化 (scale variation, SV)、背景遮挡 (occlusion, OCC)、非刚体变形 (deformation, DEF)、运动模糊 (motion blur, MB)、快速运动 (fast motion, FM)、面内旋转 (out-of-plane rotation, OPR)、面外旋转 (in-plane rotation, IPR)、目标消失 (out-of-view, OV)、背景杂乱 (background clutters, BC)、低分辨率 (low resolution, LR). 接下来将从定性比较和定量比较两个主要方面来分析本文 TSCF 算法的鲁棒性和有效性.

4.1 实验平台

为了对本文提出的 TSCF 算法进行公平性能测评, 在 TB100 数据库上对 100 个视频序列采用相同的参数作对比实验. 本文的 TSCF 算法实验的硬件环境为 Intel i7-4790 CPU@3.6 GHz 和 32 G 的内存, 操作系统为 Ubuntu 14.04 (64 bit) 的标准个人电脑, 软件环境为 MATLAB 2017a.

4.2 实验细节及参数设置

本文求解多通道相关滤波器所使用的是 HOG 和 CN 特征, 其中 HOG 特征图通道个数为 31, CN 特征图通道个数为 11. 本文工作并没有使用任何深度特征, 使用深度特征在一定程度上会提升跟踪性能, 但同时也会带来模型参数和计算复杂度的急剧增加, 为简单起见, 本文仅采用手工特征作对比实验. 从下文的定性和定量讨论可以看出本文 TSCF 算法能够达到很高的跟踪性能. 为了验证本文 TSCF 算法的鲁棒性与有效性, 选取了当前比较优秀的 11 个基于相关滤波的视觉跟踪算法在 TB100 和 TB50 数据库上作对比实验, 它们分别是 KCF [5], SAMF [6], fDSST [8], SRDCF [10], ECOHC [12], STRCF [13], MKCFup [22], BACF [24], Staple [27], CACF [28], CFAT [29]. 为了对比实验的公平性, 本文的 TSCF 算法保留了 BACF 算法的参数设置, 这 11 个算法以及本文的 TSCF 算法采用的都是纯手工特征, 所有视频序列只有第 1 帧的初始位置为真值. 本文的部分参数设置为多分辨率尺度个数 $S=5$, 尺度增量因子 $\alpha=1.01$, 最大训练样本个数 $T_{\max}=50$, 目标区域分块个数为 4 个 (凭经验随着分块个数的增加, 跟踪性能会有所提升, 但是本文考虑到计算代价问题取最小分块数 4 来做实验), 共轭梯度迭代初始帧

图 3　(网络版彩图) 真实场景中 TSCF 和 BACF 算法对比结果展示, 红色代表 TSCF, 蓝色代表 BACF, 橙色代表 BACFPCG-S, 粉色代表 BACFPCG-M, 绿色代表 TCF

Figure 3　(Color online) Comparison between TSCF and BACF algorithms, where red denotes TSCF, blue represents BACF, orange represents BACFPCG-S, pink represents BACFPCG-M, and green refers to TCF. (a) Basketball (IV, OCC, DEF, IPR, BC); (b) Jogging-2 (OCC, DEF, IPR); (c) Diving (SV, DEF, OPR); (d) Skating2-1 (SV, OCC, DEF, FM, IPR); (e) Box (IV, SV, OCC, MB, OPR, IPR, OV, BC, LR); (f) KiteSurf (IV, OCC, OPR, IPR)

设为 150 次, 后续帧设为 5 次, 正则化因子 $\lambda = 0.01$, $\varpi = 55$, $\eta = 0.025$, 产生高斯回归函数标签的核带宽设置为 0.075. 接下来从定性比较和定量比较两个方面来分析本文算法的有效性与鲁棒性.

4.3　与 BACF 算法定性比较

TB100 数据库定性测评. 图 3 和 4 给出了 TB100 数据库中 12 个具有挑战性的视频序列的对比结果, 它们包含了 11 个不同的视频属性, 从图 3 和 4 的 12 个视频序列可以定性看出, 本文的 TSCF

图 4 (网络版彩图) 真实场景中 TSCF 和 BACF 算法对比结果展示, 红色代表 TSCF, 蓝色代表 BACF, 橙色代表 BACFPCG-S, 粉色代表 BACFPCG-M, 绿色代表 TCF

Figure 4 (Color online) Comparison between TSCF and BACF algorithms, where red denotes TSCF, blue represents BACF, orange represents BACFPCG-S, pink represents BACFPCG-M, and green refers to TCF. (a) Biker (SV, OCC, MB, FM, IPR, OV, LR); (b) ClifBar (SV, OCC, MB, FM, OPR, OV, BC); (c) Bird2 (OCC, DEF, FM, OPR, IPR); (d) DragonBaby (SV, OCC, MB, FM, OPR, IPR, OV); (e) BlurOwl (SV, MB, FM, OPR); (f) Freeman3 (SV, OPR, IPR, LR)

算法 (红色实线) 能够准确地定位目标, 结果展示明显优于基准 BACF 算法 (蓝色虚线). 接下来将从不同的视频属性来分析本文 TSCF 算法的有效性与鲁棒性. 在图 3 和 4 中 BACFPCG 算法 (粉色虚线) 指的是 BACF 算法仅采用预条件共轭梯度 (PCG) 优化方法的跟踪结果, TCF 指的是在 BACF 框架下引入时间一致性约束后采用预条件共轭梯度优化方法的跟踪结果 (绿色虚线), TSCF 指的是在 BACF 框架下引入时空一致性约束后采用预条件共轭梯度优化方法的跟踪结果, 在后面的 4.5 小节中本文还将定量讨论这 4 种算法的鲁棒性和有效性.

(1) 背景遮挡 (OCC). 前景目标在运动过程中被自身或其他物体遮挡时, 跟踪算法不可避免地学习到背景信息而导致模型快速退化, 能否有效地解决遮挡一直是视觉跟踪面临的一大难题. 本文引入时间一致性项来对多通道相关滤波进行建模, 能够有效地处理目标短暂遮挡现象. 图 3 视频序列 (a) Basketball 在 20 和 651 帧前后被其他队员短暂完全遮挡, 视频序列 (b) Jogging-2 在 56 帧前后被电线杆短暂完全遮挡, 视频序列 (d) Skating2-1 中在 84 帧前后红衣溜冰女子被同伴短暂完全遮挡, 视频序列 (e) Box 在 460 帧前后被卡尺短暂完全遮挡. 图 4 视频序列 (a) Biker, (c) Bird2 和 (d) DragonBaby 由于旋转而被自身部分遮挡. 以上列出的几个典型视频序列中都出现了短暂遮挡现象, 由于本文的模型引入了时间一致性项, 在时间意义上起到平滑滤波的作用, 使得学习到的多通道相关滤波在短时间内避免了突变的可能性, 能够有效地处理短时遮挡问题. 因此, 以上提到的几个视频序列当遇到短时完全遮挡或部分遮挡时, 本文的 TSCF 算法 (红色实线) 能够精准地锁定目标. 而 BACF 算法 (蓝色虚线) 相对来说跟踪性能有所下降, 比如图 3 视频序列 (a) Basketball 在 670 帧后发生了漂移, 视频序列 (b) Jogging-2 在 76 帧后发生了漂移, 视频序列 (e) Box 在 490 帧后发生了漂移, 图 4 视频序列 (c) Bird2 在 48 帧后发生了漂移. 究其原因是 BACF 算法遇到突变现象 (短时部分遮挡或完全遮挡) 时, 学习到的相关滤波偏向背景. 从图 2 中的相关响应图也印证了这一现象. 非刚体变形 (DEF) 通常指的是人体或动物等在进行具有较高运动自由度的运动过程中发生形状及外观的剧烈变化, 通常的跟踪算法很难准确学习到这种复杂变化. 图 3 视频序列 (a) Basketball, (b) Jogging-2, (c) Diving 和 (d) Skating2-1 是典型的非刚体变形视频序列, 本文的 TSCF 算法能够很好地跟踪目标, 而 BACF 算法发生了漂移.

(2) 面内旋转 (OPR). 类似于非刚体变形, 当目标发生面内旋转时, 目标外观模板与模型外观模板的相似度将显著降低, 从而增加了模型正确定位目标的难度. 面外旋转 (IPR), 又称垂直旋转, 通常情况下随着面外旋转一部分目标将会消失, 但另一部分新的未知目标随之出现, 用部分已知目标去预测部分未知目标, 使得学习到的相关滤波器极有可能偏向于一侧而发生漂移. 本文引入空间一致性项来对多通道相关滤波进行建模, 能够有效地处理面内旋转和面外旋转带来的不稳定性现象. 图 4 视频序列 (c) Bird2 在第 48 帧前后发生面内旋转伴随面外旋转, 图 4 视频序列 (d) DragonBaby 和图 3(f) KiteSurf 在整个运动过程中一直进行面内旋转和面外旋转. 本文的 TSCF 算法由于引入了空间一致性项, 使得学习到的多通道相关滤波器更加均匀, 防止因面外旋转而使得学习到的滤波器偏向于某一不可靠区域, 图 2 中的滤波器响应图正好核实了这一现象. 而 BACF 算法, 在图 4 视频序列 (c) Bird2 中第 48 帧后发生了漂移, 图 4 视频序列 (d) DragonBaby 中第 26 帧后发生了漂移, 图 3 视频序列 (f) KiteSurf 中第 31 帧后发生了漂移.

(3) 尺度变化 (SV) 通常指的是目标远近变化或摄像头的移动而发生的尺寸和位移上的变化. 大部分视频序列在运动过程中都面临着尺度变化现象, 尤其是图 4 视频序列 (a) Biker 和 (f) Freeman3, 在目标由远及近运动过程中目标尺寸发生了很大的变化, 本文的 TSCF 算法不但能够很好地定位目标, 而且具有很高的 AUC. BACF 算法在图 4 视频序列 (f) Freeman3 中虽然能够定位到目标, 但是当目标尺度变大后, 没能很好地调整尺度. 光照变化 (IV) 发生时很大程度上会改变目标的颜色分部信息, 而本文仅采用手工特征 (HOG + CN) 提取目标信息, 因此带来很大的挑战, 如图 3 视频序列 (a) Basketball, (e) Box 和 (f) KiteSurf, 本文的 TSCF 算法均能很好地定位目标.

(4) 快速运动 (FM) 和运动模糊 (MB) 一般同时发生, 快速运动的目标极有可能超出模型搜索范围导致跟踪算法定位目标失败, 而运动模糊使得目标外观模型发生了剧烈的变化, 尤其长时间出现运动模糊时, 如图 4 视频序列 (e) BlurOwl, 学习到的外观模型的判别力会随着在线更新而显著下降最终导致漂移. 在图 4 视频序列 (e) BlurOwl 中, BACF 算法由于快速运动伴随运动模糊在 155 帧后发生

了漂移, 在 (b) ClifBar 中由于快速运动不能很好地定位目标, 而本文的 TSCF 算法能够很好地定位目标直到视频序列最后一帧.

(5) 背景杂乱 (BC). 本文主要以矩形框来标定目标, 因此框内不可避免地混进目标前景以外的背景信息, 尤其是当目标前景和背景的纹理或颜色极其相似时, 跟踪算法很容易漂移到相似背景区域. 如图 3 视频序列 (a) Basketball, (e) Box 和图 4(b) ClifBar, 本文的 TSCF 算法较 BACF 算法能够达到很高的距离精度和 AUC. 低分辨率 (LR). 当跟踪算法遇到低分辨率目标时通常不能有效提取目标的纹理或颜色信息而带来极大的挑战, 如图 4 视频序列 (a) Biker 和 (f) Freeman3, 本文的 TSCF 算法能够有效地锁定目标. 目标短暂消失 (OV) 类似于目标遮挡, 本文引入时间一致性项后能够有效地处理这一现象, 如图 4 视频序列 (a) Biker 和 (d) DragonBaby, 本文的 TSCF 算法在目标短暂消失后没有发生漂移. 但是由于本文的跟踪算法没有加入目标检测机制, 如果将来遇到目标被长时间遮挡或消失时, 可能将会导致跟踪失败, 这也是本文接下来要研究的一个方向.

4.4 与 11 个目前最优算法的定量比较

TB100 数据库定量测评. 本文采用 TB100 数据库中的 OPE (one-pass evaluation) 评测准则, 以中心点位置误差 (center location error) 和目标矩形框的重叠率 (bounding box overlap) 为评价指标, 在 TB100 数据库上 100 个视频序列采用相同的参数与目前比较优秀的 11 个基于相关滤波的视觉跟踪算法作对比实验. 中心点位置误差指的是每一帧视频序列的跟踪结果与人工标注真值之间的欧氏距离 (以像素为单位), 如果该距离小于某个给定的阈值 (本文设置为 20 个像素) 时被认为该跟踪算法在该跟踪时刻成功地跟踪上了目标, 这是一个广泛应用于视觉跟踪领域的评估跟踪精度的方法, 本文采用距离精度曲线图 (distance percision curver) 及对应的比率来评估不同的跟踪算法; 目标矩形框的重叠率指的是跟踪结果与真值之间的交并比 (intersection-over-union, IoU), 跟踪目标框的 IoU 计算公式为

$$\mathrm{IoU} = \frac{\mathrm{area}\left(\mathrm{ROI}_T \cap \mathrm{ROI}_G\right)}{\mathrm{area}\left(\mathrm{ROI}_T \cup \mathrm{ROI}_G\right)}, \tag{20}$$

其中 ROI_T 为跟踪结果的目标框, ROI_G 为人工标注真值, 该方法同时考虑了被跟踪目标的位置和尺度, 因此这是一个不同于中心点位置误差且被广泛应用地刻画跟踪算法鲁棒性和准确性的评估方法. 本文采用成功率曲线图 (success rate curver) 和成功率曲线相对应的线下面积 (AUC) 值来综合评估不同跟踪算法的有效性. 图 5 展示的是 TB100 数据库上 100 个视频序列的成功率曲线图及对应的 AUC 和 11 个视频属性的成功率曲线图及对应的 AUC. 从图 5 可知, 本文的 TSCF 算法在 11 个视频属性上 AUC 均明显高于 BACF 算法. 图 6 为 TB100 数据库上 100 个视频序列的距离精度和 11 个视频属性的距离精度的对比实验图. 图 6 显示, 本文的 TSCF 算法在 11 个视频属性上距离精度均明显高于 BACF 算法, 这说明本文的 TSCF 算法具有一定的鲁棒性和有效性. 因此, 从图 5 和 6 可知, 本文的算法与其他 11 个优秀的算法作比较, 本文的结果排名基本上处在最优或次优, 明显优于大多数基于相关滤波的视觉跟踪算法.

TB50 数据库定量测评. 表 1 和 2 展示了在 TB50 数据库上采用相同的参数与目前最优的 11 个基于相关滤波的视觉跟踪算法的对比实验结果. 表 1 定量展示了这 12 个算法在 11 个视频属性上的 AUC, 表 2 定量展示了在 11 个视频属性上的距离精度 (阈值为 20 个像素), 两个表中上角标记 “#” 的数值代表对比结果最好, 上角标记 “*” 的数值次之. 表 1 的 AUC 数据和表 2 的距离精度比率结果展示, 与其他 11 个优秀的算法比较, 本文的 AUC 和距离精度比率排名处在最优或次优, 明显优于大多数基于相关滤波的视觉跟踪算法.

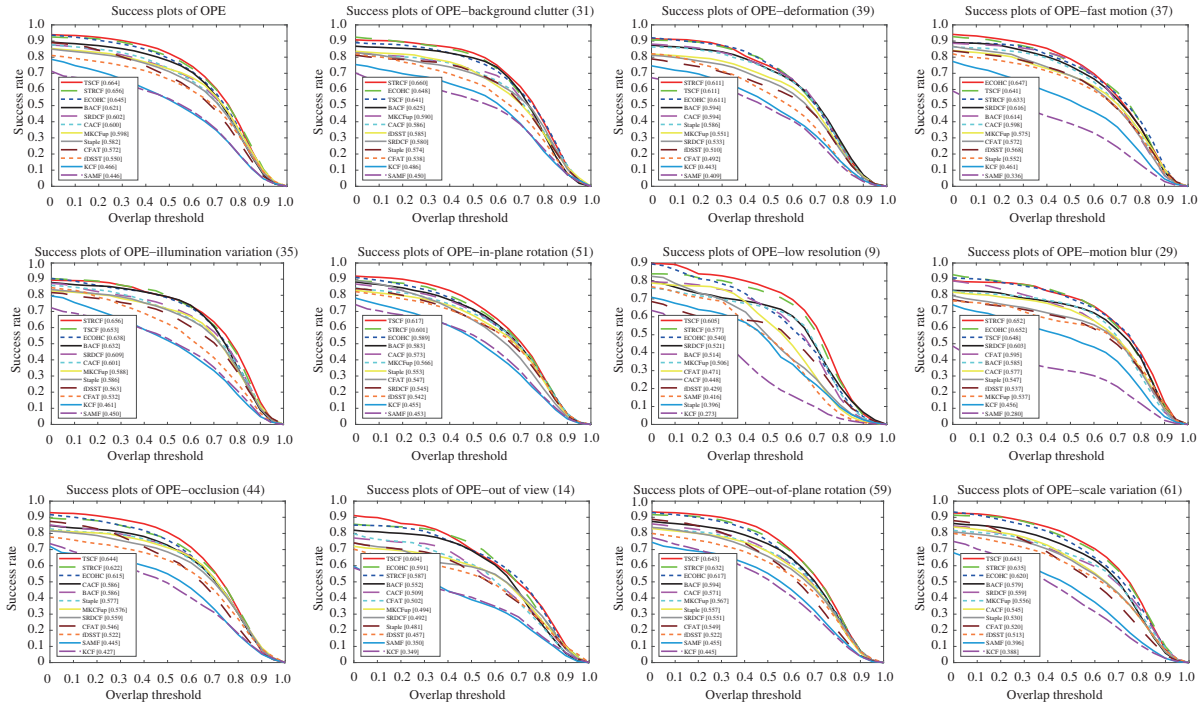


图 5 (网络版彩图) TB100 数据库上 100 个视频序列的 AUC 和 11 个视频属性的 AUC

Figure 5 (Color online) The AUC of 100 video sequences and 11 video attributes on TB100 database

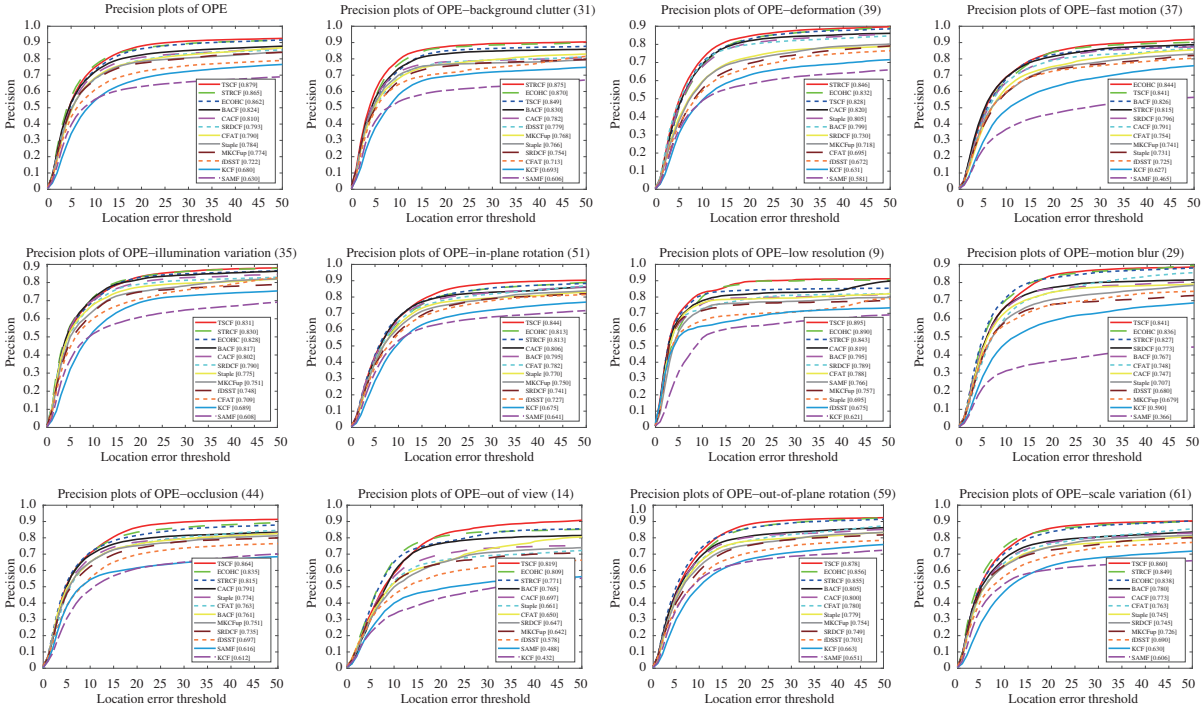


图 6 (网络版彩图) TB100 数据库上 100 个视频序列的距离精度和 11 个视频属性的距离精度

Figure 6 (Color online) The distance precision of 100 video sequences and 11 video attributes on TB100 database

表 1　在 TB50 数据库上 11 个目前最优算法和本文的 TSCF 算法在 11 个视频属性上的 AUC 定量对比结果

Table 1 Comparison of AUC of 11 state-of-the-art algorithms and our TSCF algorithm on 11 video attributes on TB50 database

	Total number of video	KCF	SAMF	Staple	CFAT	fDSST	SRDCF	CACF	MKCFup	ECOHC	STRCF	BACF	TSCF
TB50	50	0.399	0.392	0.517	0.529	0.504	0.540	0.542	0.546	0.601	0.606*	0.574	0.621#
BC	20	0.433	0.386	0.513	0.496	0.561	0.533	0.517	0.562	0.600	0.614*	0.561	0.620#
DEF	19	0.391	0.354	0.533	0.444	0.460	0.455	0.538	0.503	0.555	0.556*	0.551	0.572#
FM	22	0.389	0.282	0.495	0.533	0.554	0.573	0.541	0.530	0.611#	0.588	0.576	0.596*
IV	20	0.410	0.411	0.511	0.485	0.534	0.521	0.530	0.550	0.579	0.594#	0.560	0.583*
IPR	29	0.368	0.376	0.466	0.511	0.474	0.474	0.502	0.512	0.563	0.575#	0.546	0.573*
LR	8	0.267	0.432	0.403	0.460	0.437	0.526	0.460	0.527	0.562*	0.559	0.518	0.585#
MB	19	0.393	0.303	0.489	0.553	0.527	0.549	0.528	0.496	0.605#	0.586	0.539	0.590*
OCC	27	0.371	0.447	0.521	0.528	0.481	0.506	0.536	0.548	0.589*	0.589*	0.557	0.603#
OV	11	0.277	0.302	0.463	0.439	0.454	0.465	0.488	0.473	0.549*	0.537	0.508	0.561#
OPR	29	0.361	0.408	0.475	0.489	0.466	0.472	0.479	0.526	0.583	0.594#	0.545	0.590*
SV	34	0.344	0.370	0.470	0.502	0.491	0.509	0.493	0.523	0.593*	0.590	0.528	0.600#
FPS	–	238#	25	69	5	94	10	43	150*	55	23	34	3

表 2　在 TB50 数据库上 11 个目前最优算法和本文的 TSCF 算法在 11 个视频属性上的距离精度定量对比结果

Table 2 Comparison of distance precision of 11 state-of-the-art algorithms and our TSCF algorithm on 11 video attributes on TB50 database

	Total number of video	KCF	SAMF	Staple	CFAT	fDSST	SRDCF	CACF	MKCFup	ECOHC	STRCF	BACF	TSCF
TB50	50	0.589	0.561	0.687	0.710	0.684	0.723	0.730	0.730	0.821*	0.815	0.768	0.842#
BC	20	0.632	0.511	0.664	0.636	0.773	0.692	0.677	0.737	0.807	0.815*	0.745	0.824#
DEF	19	0.579	0.520	0.733	0.641	0.645	0.663	0.75	0.692	0.796	0.798*	0.750	0.801#
FM	22	0.555	0.408	0.660	0.702	0.733	0.767	0.733	0.692	0.819#	0.768	0.778	0.801*
IV	20	0.631	0.564	0.681	0.634	0.737	0.706	0.723	0.730	0.779*	0.783#	0.748	0.775
IPR	29	0.549	0.518	0.635	0.695	0.655	0.637	0.704	0.679	0.782*	0.772	0.737	0.786#
LR	8	0.587	0.748	0.667	0.762	0.645	0.764	0.807	0.756	0.882#	0.823*	0.770	0.882#
MB	19	0.548	0.408	0.657	0.700	0.701	0.740	0.712	0.653	0.808*	0.767	0.724	0.809#
OCC	27	0.556	0.639	0.723	0.72	0.682	0.697	0.744	0.734	0.843#	0.808*	0.757	0.843#
OV	11	0.364	0.446	0.658	0.576	0.613	0.623	0.686	0.636	0.774*	0.726	0.724	0.795#
OPR	29	0.553	0.580	0.663	0.663	0.662	0.651	0.702	0.713	0.834#	0.810	0.737	0.821*
SV	34	0.563	0.558	0.653	0.701	0.681	0.684	0.697	0.690	0.818*	0.806	0.716	0.820#
FPS	–	238#	25	69	5	94	10	43	150*	55	23	34	3

4.5　本文 TSCF 算法的有效性验证

4.5.1　TB100 数据库算法有效性验证的定性和定量测评

TB100 数据库算法有效性验证定量测评. 为了验证本文 TSCF 算法的有效性, 首先, 基准 BACF

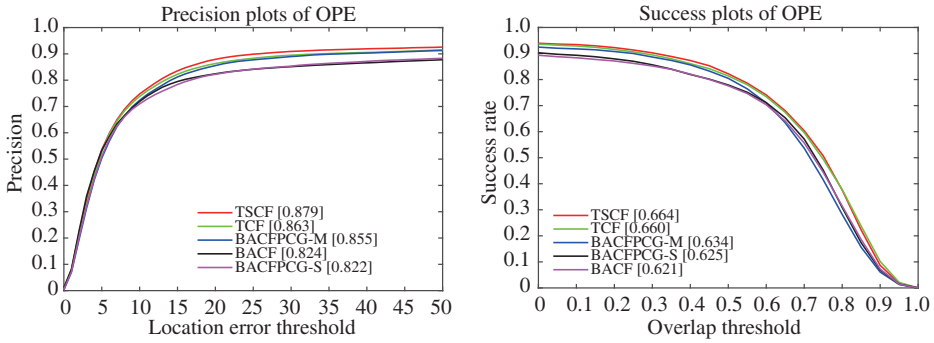


图 7 (网络版彩图) TB100 数据库上 TSCF 算法的距离精度和 AUC 对比结果

Figure 7 (Color online) Comparison of distance precision and AUC of TSCF algorithm on TB100 database

算法模型求解采用预条件共轭梯度优化方法, 基准 BACF 算法模型学习策略保持不变 (此处称之为单样本学习策略), 只是把模型更新的学习率从 lrate = 0.013 调整到 lrate = 0.025, 称为 BACFPCG-S 算法. 这里的评价准则仍是采用 TB100 数据库的 OPE 准则, 图 7 的距离精度曲线图和 AUC 显示, 跟踪性能较基准 BACF 算法有相似结果, AUC 略高于基准 BACF 算法, 而距离精度略低于基准 BACF 算法. 表 3 和 4 的数据显示 BACFPCG-S 算法的跟踪性能类似于基准 BACF 算法, 结果展示采用预条件共轭梯度方法求解 BACF 模型是有效的. 其次, 为了进一步提升跟踪性能, 受到 ECO [12] 算法中样本更新策略的启发, 使用多样本学习策略替代 BACFPCG-S 算法中的单样本学习策略, 即基准 BACF 算法仅采用预条件共轭梯度优化方法求解, 没有使用时间一致性约束和空间一致性约束条件 (即时间正则化因子 $\varpi = 0$ 且空间正则化因子 $\eta = 0$), 称为 BACFPCG-M 算法. 图 7 显示, 距离精度较基准 BACF 算法提升了 3.1%, AUC 提升了 1.3%, 结果展示多样本学习策略在一定程度上能够提升跟踪性能. 再者, 本文引入时间一致性约束项后 (即空间正则化因子 $\eta = 0$) 采用预条件共轭梯度优化方法求解称之为 TCF 算法. 图 7 显示, 距离精度较基准 BACF 算法提升了 3.9%, AUC 提升了 3.9%. 最后, 本文引入时空一致性约束项后采用预条件共轭梯度优化方法求解称之为 TSCF 算法. 从图 7 以及表 3 和 4 可知, 距离精度较基准 BACF 算法提升了 5.5%, AUC 提升了 4.3%. 这些结果展示本文的 TSCF 算法在遇到挑战性问题时具有一定的鲁棒性和有效性, 尤其是遇到短时间遮挡和面内旋转或面外旋转时, 跟踪性能较基准 BACF 算法有明显提升, 这说明本文加入时间一致性约束后, 当目标遇到诸如短时突变现象时, 能够有效地起到平滑滤波器 (时间上) 的作用, 避免学习到的多通道相关滤波偏向背景而发生漂移, 而本文的空间一致性约束在一定程度上能够有效地平滑滤波器 (空间上) 避免学习到的多通道相关滤波偏向于某一不可信的背景区域.

4.5.2 与 9 个深度学习框架算法的定量比较

为了进一步验证本文 TSCF 算法的鲁棒性与有效性, 选取了目前最优的 9 个基于深度学习框架的视觉跟踪算法在 TB100 数据库上作对比实验, 它们分别是 DRT [20], FlowTrack [23], SaSiam [30], SiamRPN [31], StructSiam [32], DasiamRPN [33], DSLT [34], MemTrack [35], ACT [36]. 此处的评价准则仍是采用 TB100 数据库的 OPE 准则, 实验对比结果如表 5. 表 5 数据显示, 本文的 TSCF 算法仅采用手工特征在跟踪性能上优于部分基于深度学习框架的视觉跟踪算法.

4.5.3 TC128 数据库算法有效性验证定性测评

最后, 为了再次验证本文 TSCF 算法的鲁棒性与有效性, 所有参数保持不变的情况下, 即保留在

表 3　在 TB100 数据库上的 AUC 定量对比结果

Table 3　Comparison of AUC for algorithm validation on TB100 database

	TB100	BC	DEF	FM	IV	IPR	LR	MB	OCC	OV	OPR	SV
TSCF	0.664	0.641	0.611	0.641	0.653	0.617	0.605	0.648	0.644	0.604	0.643	0.643
TCF	0.660	0.642	0.607	0.639	0.656	0.607	0.607	0.654	0.632	0.602	0.633	0.635
BACFPCG-M	0.634	0.609	0.600	0.617	0.627	0.586	0.605	0.640	0.606	0.584	0.604	0.626
BACFPCG-S	0.625	0.605	0.596	0.586	0.632	0.558	0.576	0.612	0.595	0.530	0.597	0.602
BACFPCG	0.621	0.625	0.594	0.614	0.632	0.583	0.514	0.585	0.586	0.552	0.594	0.579

表 4　在 TB100 数据库上的距离精度定量对比结果

Table 4　Comparison of distance precision for algorithm validation on TB100 database

	TB100	BC	DEF	FM	IV	IPR	LR	MB	OCC	OV	OPR	SV
TSCF	0.879	0.849	0.828	0.841	0.831	0.844	0.895	0.841	0.864	0.819	0.878	0.860
TCF	0.863	0.831	0.824	0.814	0.818	0.817	0.887	0.824	0.835	0.820	0.852	0.834
BACFPCG-M	0.855	0.816	0.824	0.812	0.812	0.810	0.938	0.830	0.825	0.806	0.840	0.839
BACFPCG-S	0.822	0.789	0.810	0.739	0.801	0.750	0.848	0.766	0.777	0.678	0.804	0.791
BACFPCG	0.824	0.830	0.799	0.826	0.817	0.795	0.795	0.767	0.761	0.765	0.805	0.780

表 5　在 TB100 数据库上 9 个目前最优的基于深度学习的视觉跟踪算法和本文的 TSCF 算法定量对比结果

Table 5　Comparison of 9 state-of-the-art visual object tracking algorithms based on deep learning and our TSCF algorithm on TB100 database. # and ∗ denote the first and second best results, respectively.

	DRT	Sa-Siam	SiamRPN	FlowTrack	StructSiam	DasiamRPN	DSLT	MemTrack	ACT	TSCF
AUC	0.699#	0.610	0.637	0.655	0.621	0.617	0.660	0.642	0.643	0.664*
DP	0.923#	0.823	0.851	0.881	0.851	0.880	0.909*	0.849	0.859	0.879

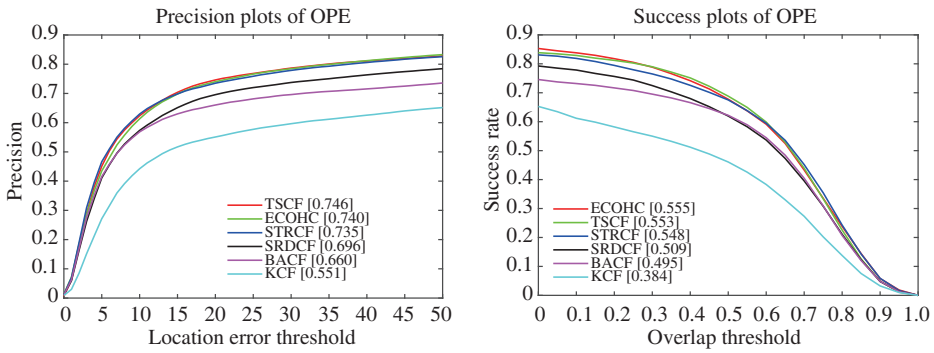


图 8　(网络版彩图) TC128 数据库上的 TSCF 算法的距离精度和 AUC 对比结果

Figure 8　(Color online) Comparison of distance precision and AUC of TSCF algorithm on TC128 database

OTB100 数据库上的参数设置, 本文的 TSCF 算法在 TC128 数据库上与 5 个基于相关滤波的视觉跟踪算法作定性对比测评, 这 5 个对比算法分别是 KCF [5], SRDCF [10], ECOHC [12], STRCF [13], BACF [24], 此处的测评准则是采用 TB100 数据库的 OPE 准则, 实验对比结果如图 8. 图 8 显示, 本文的 TSCF 算法仅采用手工特征在跟踪性能上优于大部分基于相关滤波视觉跟踪算法.

4.5.4 与相近算法的区别和联系

本文的 TSCF 模型与文献 [26] 中的 STRCF 模型都引入了时间正则项, 在时间序列意义上起到平滑多通道相关滤波的作用, 但也有很大的差别: 本文 TSCF 的基准是 BACF 模型, 而 STRCF 的基准是 SRDCF 模型; 本文的模型学习策略是采用多样本学习策略, 而 STRCF 采用的是单样本学习策略; 本文的闭式解采用共轭梯度下降法优化求解, 而 STRCF 采用 ADMM 方法优化求解; 本文还引入了空间一致性项, 目的是在空间意义上平滑多通道相关滤波, 而 STRCF 没有考虑空间一致性问题. 从图 5 和 6 以及表 1 和 2 的对比实验结果看, 本文 TSCF 算法的跟踪精度和 AUC 均明显高于 STRCF 算法.

本文的 TSCF 模型和文献 [20] 中的 DRT 模型都引入了空间一致性信息, 但是考虑的侧重点不同. 本文的 TSCF 模型在空间意义上主要侧重于平滑多通道相关滤波, 避免学习到的滤波器偏向于某一不可靠区域, 使得学习到的滤波器更加均匀, 而 DRT 则是侧重于学习各个分块区域的权重, 增加了模型参数, 这也势必会加大模型复杂度, 因此 DRT 采用共轭梯度下降方法和二次规划方法交替迭代逼近模型的最优解; 此外, DRT 模型没有考虑时间一致性信息而本文引入了时间一致性正则项. 由于 DRT 采用的是深度特征, 而本文仅采用手工特征, 如表 5 所示, DRT 的跟踪性能优于本文的手工特征的跟踪性能, 但是表 5 数据显示, 本文的 TSCF 模型仅采用手工特征优于部分基于深度学习框架的视觉跟踪算法.

5 结论

本文提出了学习时空一致性相关滤波 (TSCF) 跟踪算法, 本文的 TSCF 算法在基准 BACF 算法框架下引入时间一致性约束性项, 在时间序列意义上起到平滑多通道相关滤波的作用, 有效避免了因连续帧之间外观突变而使得学习到的相关滤波向背景偏移; 同时本文还引入空间一致性约束项, 在空间分布意义上平滑多通道相关滤波, 使得学习到的相关滤波能量分布更加均匀, 避免学习到的相关滤波器偏向某一不可靠背景区域. 本文采用共轭梯度下降法迭代逼近有闭式解方程组的最优解, 且优化过程可以利用循环矩阵性质转化到傅里叶域快速求解, 有效降低了计算大型矩阵的代价. 本文的 TSCF 算法跟踪结果在 TB100 公开数据库上显示, 距离精度较基准 BACF 算法提升了 5.5%, AUC 提升了 4.3%. 纯手工特征跟踪性能在 TB100 数据库上 100 个视频的距离精度达到 0.879, AUC 为 0.664. 本文的 TSCF 算法与 11 个目前最优的视觉跟踪算法的对比实验表明距离精度和 AUC 排名均在最优或次优, 明显优于大多数基于相关滤波的视觉跟踪算法, 结果展示本文的 TSCF 算法在遇到诸如短时间遮挡和面内旋转或面外旋转等挑战性问题时具有一定的鲁棒性和有效性. 在接下来的工作中我们将尝试将深度特征加入到 TSCF 框架中从而提升跟踪性能.

参考文献

1 Wu Y, Lim J, Yang M H. Object tracking benchmark. IEEE Trans Pattern Anal Mach Intell, 2015, 37: 1834–1848

2 Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, 2010. 2544–2550

3 Henriques J F, Rui C, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of IEEE European Conference on Computer Vision, Florence, 2012. 702–715

4 Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014. 1090–1097

5 Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters. IEEE Trans Pattern Anal Mach Intell, 2015, 37: 583–596

6 Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration. In: Proceedings of IEEE European Conference on Computer Vision Workshops, Zurich, 2014. 254–265

7 Danelljan M, Hager G, Khan F S. Accurate scale estimation for robust visual tracking. In: Proceedings of IEEE British Machine Vision Conference, Nottingham, 2014

8 Danelljan M, Hager G, Khan F S, et al. Discriminative scale space tracking. IEEE Trans Pattern Anal Mach Intell, 2017, 39: 1561–1575

9 Galoogahi H K, Sim T, Lucey S. Correlation filters with limited boundaries. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015. 4630–4638

10 Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of IEEE International Conference on Computer Vision, Santiago, 2015. 4310–4318

11 Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking. In: Proceedings of IEEE European Conference on Computer Vision, Amsterdam, 2016. 472–488

12 Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 6931–6939

13 Li F, Tian C, Zuo W M, et al. Learning spatial-temporal regularized correlation filters for visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 4904–4913

14 Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking. In: Proceedings of IEEE International Conference on Computer Vision Workshops, Santiago, 2015. 621–629

15 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 4293–4302

16 Song Y B, Ma C, Gong L J, et al. CREST: convolutional residual learning for visual tracking. In: Proceedings of IEEE International Conference on Computer Vision, Venice, 2017. 2574–2583

17 Gundogdu E, Alatan A A. Good features to correlate for visual tracking. IEEE Trans Image Process, 2018, 27: 2526–2540

18 Valmadre J, Bertinetto L, Henriques J, et al. End-to-end representation learning for correlation filter based tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 5000–5008

19 Choi J, Chang H J, Yun S, et al. Attentional correlation filter network for adaptive visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 4828–4837

20 Sun C, Wang D, Lu H C, et al. Correlation tracking via joint discrimination and reliability learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 489–497

21 Sun C, Wang D, Lu H C, et al. Learning spatial-aware regressions for visual tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 8962–8970

22 Tang M, Yu B, Zhang F, et al. High-speed tracking with multi-kernel correlation filters. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 4874–4883

23 Zhu Z, Wu W, Zou W, et al. End-to-end flow correlation tracking with spatial-temporal attention. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 548–557

24 Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking. In: Proceedings of IEEE European Conference on Computer Vision, Venice, 2017. 1144–1152

25 Lu H C, Li P X, Wang D. Visual object tracking: a survey. Pattern Recogn Artif Intell, 2018, 31: 61–76 [卢湖川, 李佩霞, 王栋. 目标跟踪算法综述. 模式识别与人工智能, 2018, 31: 61–76]

26 Li P X, Wang D, Wang L J, et al. Deep visual tracking: review and experimental comparison. Pattern Recogn, 2018, 76: 323–338

27 Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 1401–1409

28 Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 1387–1395

29 Bibi A, Mueller M, Ghanem B. Target response adaptation for correlation filter tracking. In: Proceedings of IEEE European Conference on Computer Vision, Amsterdam, 2016. 419–433

30 He A F, Luo C, Tian X M, et al. A twofold siamese network for real-time object tracking. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 4834–4843

31 Li B, Yan J J, Wu W, et al. High performance visual tracking with siamese region proposal network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018. 8971–8980
32 Zhang Y H, Wang L J, Qi J Q, et al. Structured siamese network for real-time visual tracking. In: Proceedings of IEEE European Conference on Computer Vision, Munich, 2018. 355–370
33 Zhu Z, Wang Q, Li B, et al. Distractor-aware siamese networks for visual object tracking. In: Proceedings of IEEE European Conference on Computer Vision, Munich, 2018. 103–119
34 Lu X K, Ma C, Ni B B, et al. Deep regression tracking with shrinkage Los. In: Proceedings of IEEE European Conference on Computer Vision, Munich, 2018. 369–386
35 Yang T Y, Chan A B. Learning dynamic memory networks for object tracking. In: Proceedings of IEEE European Conference on Computer Vision, Munich, 2018. 153–169
36 Chen B Y, Wang D, Li P X, et al. Real-time 'actor-critic' tracking. In: Proceedings of IEEE European Conference on Computer Vision, Munich, 2018. 328–345

Learning temporal-spatial consistency correlation filter for visual tracking

Jianzhang ZHU[1*], Dong WANG[2] & Huchuan LU[2]

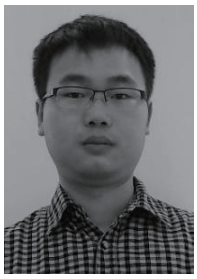
1. *School of Mathematics and Information Sciences, Henan University of Economics and Low, Zhengzhou* 450046*, China*;
2. *School of Information and Communication Engineering, Dalian University of Technology, Dalian* 116024*, China*
* Corresponding author. E-mail: zhujianzhang@126.com

Abstract Discriminant correlation filter-based tracking approaches, which adopt a circular shift operator on the tracking target object (the only accurate positive sample) to obtain training data and rely on the potential sample periodic extension hypothesis that enables model training and detection, can be efficiently accomplished through FFT. However, real background information is not modeled during the total learning process. The background-aware correlation filter (BACF) tracking algorithm uses a binary matrix to acquire real positive and negative samples using a dense sampling method to model the target's appearance. However, the BACF algorithm does not consider temporal and spatial consistency information, and when a target undergoes an abrupt change, the learned correlation filter will drift to the background. To solve this problem, in this paper, we introduce temporal and spatial consistency constraints into the baseline BACF framework and propose a learning temporal-spatial consistency correlation filter (TSCF) tracking algorithm. This enables the correlation filter to learn to adapt to the appearance of mutation between successive frames. The temporal consistency constraint smooths the multi-channel correlation filter in the time series, and the spatial consistency constraint smooths the multi-channel correlation filter in spatial distribution, thus making the energy distribution more uniform of the correlation filter learned. In this paper, the TSCF model has closed solutions and the conjugate gradient descent method is used to approximate the optimal solution of a system of closed solutions. The optimization process can then be transformed into the Fourier domain using cyclic matrix properties to quickly obtain a solution, which effectively reduces the cost of calculating large matrices. In this paper, our TSCF algorithm increases distance precision by 5.5% and raises the AUC by 4.3% compared to the baseline BACF algorithm on the TB100 public database. The distance precision achieves 0.879 and the AUC reaches 0.663 on the TB100 database making use of only hand-crafted features. The TSCF algorithm proposed in this paper can be applied to challenging conditions such as short time occlusion, out-of-plane rotation, in-plane rotation, and so on, thus demonstrating its robustness and effectiveness.

Keywords visual tracking, correlation filter, temporal-spatial consistency, regularization, conjugate gradient descent

Jianzhang ZHU was born in 1982. He received his Ph.D. degree from the School of Mathematics and Statistics, Wuhan University in 2014. He is currently a faculty member with the School of Mathematics and Information Sciences, Henan University of Economics and Law. His current research interests include computer vision and pattern recognition, with a focus on visual tracking.

Huchuan LU was born in 1972. He (SM'12) received his M.S. degree in signal and information processing and his Ph.D. degree in system engineering from Dalian University of Technology in 1998 and 2008, respectively. He joined the faculty member with Dalian University of Technology in 1998, where he is currently a full professor with the School of Information and Communication Engineering. His current research interests include computer vision and pattern recognition, with a focus on visual tracking, saliency detection, and segmentation.



Dong WANG was born in 1984. He received his B.E. degree in electronic information engineering and his Ph.D. degree in signal and information processing from Dalian University of Technology in 2008 and 2013, respectively. He is currently an associate professor with the School of Information and Communication Engineering, Dalian University of Technology. His current research interests include facial recognition, interactive image segmentation, and object tracking.