# 基于相关滤波器的视频跟踪方法研究进展

刘巧元 [1] 王玉茹 [1] 张金玲 [1] 殷明浩 [1]

**摘 要** 视频跟踪是计算机视觉的重要组成部分, 可在智能交通、医疗诊断等实际应用中发挥重要作用. 近年来, 相关滤波器凭借精度高、速度快的优势, 逐步发展为视频跟踪方法的主要研究方向之一, 可以很好地处理多种视频跟踪难题. 随着基于相关滤波器的视频跟踪系列方法被相继提出, 算法设计趋于完善, 跟踪效果也趋于精准. 本文从不同角度总结了多种具有代表性的相关滤波跟踪方法, 分析了各种方法的发展进程, 并预测了未来可能的发展方向.

**关键词** 视频跟踪, 相关滤波器, 模型训练, 岭回归

**引用格式** 刘巧元, 王玉茹, 张金玲, 殷明浩. 基于相关滤波器的视频跟踪方法研究进展. 自动化学报, 2019, **45**(2): 265−275

**DOI** 10.16383/j.aas.2018.c170394

## Research Progress of Visual Tracking Methods Based on Correlation Filter

LIU Qiao-Yuan[1] WANG Yu-Ru[1] ZHANG Jin-Ling[1] YIN Ming-Hao[1]

**Abstract** Visual tracking is an important part of computer vision, which plays a key role in practical applications such as intelligent transportation, medical diagnosis and so on. In recent years, correlation filter has been developed into a main direction of visual tracking methods due to its high precision and fast speed, as well as the ability to handle a variety of tracking challenges. With various correlation filter based tracker being proposed, the tracking algorithm design tends to be perfect, tracking effects tend to be accurate. This paper summarizes several representative correlation filter based tracking methods from different points of view, analyzes the development process of the method, and predicts its possible future development.

**Key words** Visual tracking, correlation filter, model training, ridge regression

**Citation** Liu Qiao-Yuan, Wang Yu-Ru, Zhang Jin-Ling, Yin Ming-Hao. Research progress of visual tracking methods based on correlation filter. *Acta Automatica Sinica*, 2019, **45**(2): 265−275

视频跟踪作为计算机视觉的重要研究方向, 近年来备受关注, 它的主要任务是根据已知视频序列中目标的初始状态, 通过系列算法估计出目标的运动轨迹. 视频跟踪方法在高级人机交互[1]、安全监控[2] 和行为分析[3] 等方面具有潜在的经济价值和广泛的应用前景.

视频跟踪方法从最初的差分法[4]、光流法[5] 到现在各类目标跟踪算法百花齐放的态势已有 40 多年的发展历史; 自引入机器学习算法以来, 视频跟踪算法更是得到了突飞猛进的发展. 目前视频目标跟踪主要有三大发展方向: 深度学习方向[6−8]、相关滤波方向[9−11] 和其他传统策略[12−13]. 基于深度学习的视频跟踪方法大多关注神经网络的构建与深度特征的提取, 但深度神经网络内部参数较多, 训练时间较长, 所以这类方法跟踪速度相对较慢, 很难达到实时跟踪; 而基于相关滤波器的跟踪方法却因速度快、效果好的特点吸引了众多研究者的目光, 逐步成为视频跟踪算法发展的主要方向. 由于该系列方法兴起不久, 且发展速度较快, 所以目前尚缺少相关综述性文献.

相关滤波器基于判别式框架, 与经典的支持向量机 (Support vector machine, SVM)[14] 等分类算法一样同属于监督学习. 与 SVM 等二分类算法不同的是, 基于相关滤波器的跟踪方法将训练样本标签连续化以形成置信图, 求得图中响应最大的位置即为目标. 鉴于这种方法能有效提高跟踪算法的精度和鲁棒性, 许多改进算法被相继提出, 并取得了突破性进展.

本文第 1 节介绍相关滤波器的基本理论, 第 2 节介绍近年来针对跟踪难题提出的相关滤波跟踪算

收稿日期 2017-07-19 录用日期 2017-12-23
Manuscript received July 19, 2017; accepted December 23, 2017
国家自然科学基金 (61300099), 中国博士后科学基金 (2015M570261), 吉林省科技厅科技发展计划 (20170101144JC), 教育部符号计算与知识工程重点实验室开放基金 (93K172016K14), 中央高校基础科研业务费 (2412017FZ027) 资助
Supported by National Natural Science Foundation of China (61300099), China Postdoctoral Science Foundation Funded Project (2015M570261), Science and Technology Development Plan of Jilin Province (20170101144JC), Open Fund of Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (93K172016K14), and Fundamental Research Funds for Central Universities (2412017FZ027)
本文责任编委 赖剑煌
Recommended by Associate Editor LAI Jian-Huang
1. 东北师范大学 长春 130117
1. Northeast Normal University, Changchun 130117

法, 第 3 节介绍针对跟踪策略提出的相关滤波跟踪算法, 第 4 节展示并分析所论述跟踪算法的实验结果, 第 5 节对各种方法尚存在的问题进行分析, 总结并阐述未来的发展趋势.

## 1　基本相关滤波器方法介绍

相关滤波器通常也称为判别相关滤波器 (Discriminative correlation filters, DCF), 是视频跟踪领域应用最为广泛的算法之一. 计算机方向学者把信号处理学中计算两种信号相关性的思路引入到视频跟踪的研究当中, 将目标与待检测区域比作信号, 并做相关计算, 求得相关性最大的区域, 即为目标区域.

相关滤波方法认为, 每个被良好检测的目标区域都可为跟踪提供有效信息, 且以这些目标区域作为训练样本所训练出来的模型会更可靠, 具体做法如下:

**步骤 1.** 对已跟踪出的多个目标位置提取特征, 训练出一个滤波器模板;

**步骤 2.** 用训练出的滤波器与新一帧中的待检测区域特征做相关, 相关响应最大的位置即为新一帧中目标的预测位置;

**步骤 3.** 以目标预测位置为中心提取特征, 反过来进一步训练滤波器模型, 并重复上述步骤进行后续的目标跟踪与模型训练, 进而实现模型的在线训练与目标的实时跟踪.

本节简单介绍基于相关滤波器跟踪算法的起源及发展过程中用到的经典计算策略.

### 1.1　最早的 MOSSE 方法

基于相关滤波器的目标跟踪算法最早于 2010 年[15] 提出, MOSSE (Minimum output sum of squared error) 方法开创了相关滤波器应用于目标跟踪问题的先河. 最初的相关滤波器模型相对简单, 由于使用快速傅里叶变换方法辅助计算, 所以速度较快, 可达到 669 帧/s, 虽然在处理各类跟踪问题时效果欠佳, 但具有里程碑式的意义, 为近年来相关滤波的发展奠定了坚实基础.

按照上面提到的思路, 相关滤波跟踪算法的目标就是训练一个最优的滤波模板, 使其在目标上的响应最大, 可表示为

$$g = f \otimes h \tag{1}$$

其中, $f$ 为输入图像, $h$ 为滤波模板, $g$ 为响应输出, $\otimes$ 为卷积操作, 式 (1) 可进一步展开为

$$f \otimes h(\tau) = \int_{\infty}^{-\infty} f^*(t)h(t+\tau)\mathrm{d}t \tag{2}$$

由于卷积计算耗时较大, 所以 MOSSE 方法利用快速傅里叶方法, 将 $f$ 和 $h$ 表示在傅里叶频域内, 把卷积转化为点乘, 这样可以极大地减少计算量, 则式 (1) 变为

$$G = F \times H^* \tag{3}$$

MOSSE 方法对初始跟踪框进行随机仿射变换, 生成 $m$ 个样本 $\{f_i | i \in 1, \cdots, m\}$, 再利用高斯函数生成以 $f_i$ 的中心位置为峰值的响应图 $g_i$, 最后利用如下目标函数训练出最优的相关滤波模板.

$$E = \min_{H^*} \sum_{i=1}^{m} |H^* F_i - G_i|^2 \tag{4}$$

### 1.2　经典改进方法

自 MOSSE 方法提出以后, 基于相关滤波器的跟踪算法受到了广泛的关注, 经典方法 CSK (Circulant structure of tracking-by-detection with kernels)[16] 和 KCF (Kernelized correlation filters)[17] 都是在 MOSSE 的基础上进行改进得出的, 其中用到的循环矩阵和岭回归策略巧妙有效, 有力推进了相关滤波在跟踪领域的发展, 弥补了 MOSSE 方法存在的不足, 改善了跟踪效果.

本小节主要对循环矩阵和岭回归的计算方法做简单介绍.

#### 1.2.1　循环矩阵

跟踪初始阶段的样本数量有限, 可通过对单个样本使用循环矩阵生成的新样本丰富样本库, 进而训练出更好的相关滤波模板, 该方法最早在 CSK[16] 算法中被提出, 样本转化在傅里叶频域中进行.

定义一个中心位于目标的估计位置的图像块为基础样本, 如图 1 所示.


图 1　循环采样示意图

Fig. 1　Sketch map of circular sampling

对基础样本进行循环位移操作, 以实现对目标周围的连续采样. 若将样本表示成向量形式, 可得到循环矩阵如下:

$$X = C(x) = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_n & x_1 & x_2 & \cdots & x_{n-1} \\ x_{n-1} & x_n & x_1 & \cdots & x_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_2 & x_3 & x_4 & \cdots & x_1 \end{bmatrix} \tag{5}$$

其中, 第 1 行代表基础样本, 下面各行代表经过循环位移得到的采样.

循环矩阵具备一个非常好的特性, 即可以通过离散傅里叶变换矩阵 $F$ 实现对角化.

$$X = C(x) = F \times \mathrm{diag}\{\hat{x}\} \times F^{\mathrm{H}} \tag{6}$$

其中, $x$ 为基础样本, $\hat{x}$ 为经过离散傅里叶变换的基础样本, $\hat{x} = Fx$. 通过这种方法矩阵相乘可根据循环矩阵的性质转化为元素点乘, 并在有效降低时间复杂度的同时提升跟踪速度.

#### 1.2.2 岭回归

不同于 MOSSE 以最小二乘作为目标函数, CSK[16] 和 KCF[17] 等经典相关滤波方法均采用了岭回归分类策略, 最小化目标函数

$$E = \min\left(\sum_i (f(x_i) - y_i)^2 + \lambda \left\|\omega\right\|^2\right) \tag{7}$$

其中, $f$ 为滤波器, $i$ 为样本个数, $x$ 为样本, $\lambda$ 为正则化参数, $y$ 为样本标签, $\omega$ 为滤波器系数. 岭回归目标函数在最小二乘的基础上加入了正则项 $\lambda\left\|\omega\right\|^2$, 其优点是能够放弃最小二乘法的无偏性, 以损失部分信息、降低精度为代价获得更符合实际、更可靠回归系数, 通过对式 (7) 求偏导化简目标函数, 最终可获得封闭解.

当 $f$ 为线性时, $f(z) = w^{\mathrm{T}}z$, 可得到封闭解如下:

$$w = (X^{\mathrm{T}}X + \lambda I)^{-1}X^{\mathrm{T}}y \tag{8}$$

其中, $X$ 为第 1.2.1 节提到的循环矩阵, $I$ 为单位矩阵, 根据循环矩阵的性质, 矩阵间的乘除计算可化简为元素间的点乘或点除.

当 $f$ 为非线性时, 引入核函数 $k$, 将样本从低维空间映射到高维的核空间, 则 $f(z) = w^{\mathrm{T}}z = \sum_{i=1}^{n}\alpha_i k(z, x_i)$, 在低维空间不可分的情况到高维空间之后变得线性可分了, 仍可得到封闭解为

$$w = \sum_i \alpha\varphi(x_i) \tag{9}$$

这种方法可快速检测到目标的位置.

## 2 针对跟踪难题提出的相关滤波跟踪算法

最初基于相关滤波的跟踪方法跟踪速率较快, 但在复杂背景、目标形变、长时遮挡、尺度变换等复杂情况下表现不佳, 所以近年来的相关滤波跟踪算法主要针对这几类跟踪难题提出了相应的改进策略, 跟踪效果得到了很大的改善. 本节详细介绍并分析针对这几种跟踪难题提出的相关滤波跟踪方法.

### 2.1 复杂背景

在复杂的自然环境中, 目标的机动性较大, 对目标的区分和跟踪都相对困难, 最简便有效的方法是使用颜色特征对目标进行区分. 经典的 KCF[17] 算法将 MOSSE[15] 方法的灰度特征替换为多通道的彩色特征, 使得跟踪效果得到很大提升. 如果在跟踪过程中有效利用颜色特征, 那么训练出的模型往往能够具有很强的鲁棒性.

基于自适应颜色属性的视频跟踪方法 (Color names, CN)[18] 是 2014 年 Danelljan 基于 CSK 方法提出的一种扩展方法, 该方法不仅使用了颜色特征, 还将传统的 RGB 三通道特征结合亮度细分为黑、蓝、棕、灰、绿、橙、粉、紫、红、白和黄 11 种特征, 可有效解决跟踪过程中由复杂背景导致的目标定位不准确的问题, 但由于通道过多导致计算量增大, 该方法采用自适应颜色属性策略对特征进行降维, 使整体跟踪效果得到提升, 速度可达 100 帧/s 以上. 通过最小化如下目标函数实现样本分类.

$$E = \min\sum_{j=1}^{p}\beta_j\left(\sum_{m,n}|\left\langle\phi(x_{m,n}^j, \omega^j)\right\rangle - y^j(m,n)|^2 + \lambda\left\langle\omega^j, \omega^j\right\rangle\right) \tag{10}$$

其中, $\phi$ 为核函数的投影方式, $j$ 为帧数, $y$ 为样本标签, $(m, n)$ 为样本中心位置, 每一帧的误差权重 $\beta$ 可有效增加算法的鲁棒性, 该目标函数同样需要将各参数从时域转换到傅里叶频域中进行计算.

为加快跟踪速度, 该方法利用如式 (11) 所示的 PCA (Principal component analysis) 方法实时选择当前帧中比较显著的颜色用于跟踪, 同时为当前帧 $p$ 找到一个合适的降维映射, 训练出最好的投影矩阵 $B$. 具体方法为

$$\eta_{\mathrm{tot}}^p = \alpha_p\eta_{\mathrm{data}}^p + \sum_{j=1}^{p-1}\alpha_j\eta_{\mathrm{smooth}}^j \tag{11}$$

$$\eta_{\mathrm{data}}^p = \frac{1}{MN}\sum_{m,n}\left\|\widehat{x}^p(m,n) - B_pB_p^{\mathrm{T}}\widehat{x}^p(m,n)\right\|^2 \tag{12}$$

$$\eta_{\mathrm{smooth}}^j = \sum_{k=1}^{D^2}\lambda_j^{(k)}\|b_j^{(k)} - B_pB_p^{\mathrm{T}}Tb_j^{(k)}\|^2 \tag{13}$$

其中, $\eta_{\mathrm{data}}$ 为仅由当前帧决定的重构误差, $\alpha$ 为重构误差对应权重, $\eta_{\mathrm{smooth}}$ 为新旧帧间投影矩阵的重构误差, $B$ 为降维投影矩阵, $b$ 为矩阵 $B$ 中的单个向量, $\lambda$ 为单个向量对应权重.

实验表明, 采用自适应降维方法可将初始的 11

维降低至 2 维, 可同时提高跟踪精度与计算速度, CN 方法虽然在多种跟踪难题中的表现都优于 CSK 方法, 但与目前很多算法的效果相比, 仍存在很大差距, 该算法思想仅可视为一个过渡算法.

### 2.2 目标形变

针对目标形变问题, 2016 年牛津大学的 Bertinetto 等[19] 提出了 Staple 方法, 开创了模板类特征 (HOG) 与统计类特征 (颜色直方图) 相结合的先河. HOG 特征对运动模糊和照度很鲁棒, 但是对形变不够鲁棒; 而颜色直方图特征不考虑每一个像素的位置信息, 可有效处理物体形变问题但对光照不鲁棒. 由于单独使用上述任一特征的表示模型判别能力都不够强, 且这两种特征性质互补, 所以该方法在岭回归的框架下, 结合使用 HOG 特征和颜色直方图特征, 设计目标表示方法, 并应用在跟踪方法中. 该方法优于同时期的其他复杂模型方法, 速度可以达到 80 帧/s 以上. 主要贡献在于提出了一种有效的特征融合方法, 并没有单纯用融合特征去跟踪目标得到打分, 而是从打分的角度进行融合. 当输入一张图片后, 对目标图片提取 HOG 特征, 用来训练滤波器, 然后根据相关滤波器的学习规则学习得到滤波模板并更新; 与此同时, 使用颜色直方图特征对滤波模板进行学习, 并使用同样的方式对学习到的模板进行更新.

在跟踪过程中, 首先基于上帧学习到的位置标出大致目标位置, 然后利用训练出的两个滤波器模板对目标区域做两个响应图, 最后用线性方法将得到的两个响应图融合成最终响应图, 进而最终确定目标位置. 使用的线性方法如下:

$$f(x) = \gamma_{\text{tmpl}} f_{\text{tmpl}}(x) + \gamma_{\text{hist}} f_{\text{hist}}(x) \tag{14}$$

### 2.3 目标尺度变换

大部分跟踪方法尤其是相关滤波跟踪方法在训练过程中都忽略对尺度的估计或使用统一的尺度处理不同尺度的样本, 导致在目标发生大尺度形变时较易发生目标丢失或目标偏移, 例如 KCF 方法的目标框从始至终大小未发生变化. 多数方法设计主要集中于目标定位, 也有少数方法的设计是针对尺度变化, 但跟踪速度较慢, 很难达到实时.

2014 年 Danelljan 等[20] 基于 MOSSE 框架提出 DSST (Discriminative scale space tracking) 方法, 首次在相关滤波跟踪方法中同时使用位置滤波器和尺度滤波器, 分别进行目标定位和尺度评估.

DSST 方法中的位置滤波器基于上一帧确定的目标框获取候选框, 在确定目标位置后, 尺度滤波器以当前目标框的大小为基准, 基于 33 种较精细的不同尺度候选框确定新的目标尺度. 整个联合相关滤波器基于三维尺度空间, 大小为 $M \times N \times S$, 其中 $M$, $N$ 为相关滤波器的长宽, $S$ 为相关滤波器的尺度大小; 训练样本基于特征金字塔被构造为一个大小为 $M \times N \times S$ 的立方体, 它满足以目标位置和尺度为中心的立体高斯分布, 尺度参数的更新和之前使用学习率更新相关滤波参数的方式相同, 相关响应最大的位置即为目标位置. DSST 方法虽然速度较慢, 当目标发生巨大形变时效果不佳, 但联合相关滤波器和立体化训练样本思路新颖独特, 精度方面获得了 2014 年 VOT 竞赛的冠军. 2017 年 Danelljan 等[21] 基于该文又发表了一篇扩展论文, 加入了一些加速方法后, 速度有所提升.

2015 年 Zhang 等[22] 提出了 JSSC (Tracker using joint scale-spatial correlation filters) 方法, 该方法提出了一种联合尺度空间的自适应框架, 同时考虑不同尺度的多个循环矩阵, 使用结合核函数的岭回归方法训练模型, 同时检测目标的位置和尺度信息.

该方法采用模板匹配策略, 首先假设不同尺度采样的匹配打分符合混合高斯分布, 在岭回归中最大限度的减少样本响应和匹配打分之间的差异来训练模型. 在训练阶段, 比较计算不同尺度间的样本使跟踪算法能够敏感于目标尺度的变换. 在检测阶段, 利用先验概率对样本进行线性插值计算, 进而确保对目标连续的尺度估计. 经实验分析, 使用 5 种不同尺度的采样训练出的相关滤波模板效果最优, 最终跟踪效果相对于上面提到的 DSST 方法有明显改善. 同年 Zhang 等[23] 在 JSSC 方法的基础上又延伸出了 RAJSSC (Joint scale-spatial correlation tracking with adaptive rotation estimation) 方法. 在原方法的基础上, 从目标旋转的角度对跟踪算法进行改进, 将目标模板从直角坐标系转换到极坐标系以保留旋转目标中的循环矩阵信息, 使跟踪器能够在方向空间对物体的旋转进行建模, 进而减少由于目标旋转对跟踪效果造成的影响.

### 2.4 长时遮挡

一般的相关滤波跟踪方法在处理长时间遮挡问题时较为敏感, 因为它们大都以 100～500 帧短时记忆跟踪为主, 遮挡结束后容易丢失目标或发生目标位置偏移. 针对该问题, 2015 年 Ma 等[24] 提出长时记忆相关滤波跟踪 (Long-term correlation tracking, LCT) 方法.

长时记忆跟踪即跟踪器在较长的时间内都能保持准确稳定的跟踪, 最常用的策略是给普通跟踪器搭配一个检测器, 在发现跟踪出错的时候调用自带的检测器重新检测并矫正跟踪器. LCT 方法延续了 DSST 方法中联合使用位置滤波器和尺度滤波器的

思想, 但与 DSST 不同的是, 在位置滤波器中, 通过考虑目标周围临时的上下文信息训练回归模型, 即在提取特征后加入检测区域内目标和背景的空间权重关系, 以此来对抗严重形变、长时遮挡等跟踪难题, 有效地缓解了可塑性–稳定性窘境, 这样就可以保证在学习新知识的同时, 还能保持对旧知识的记忆; 而在尺度相关滤波器中, 该方法使用大小相同但尺度不同的图像块, 提取 HOG 特征构造尺度特征金字塔, 与目标回归模型做相关, 响应最大的图像块的尺度即为最优尺度.

为防止目标丢失导致的跟踪失败, 该方法通过比较响应最大值与指定阈值, 决定是否使用 K 近邻在线分类器进行再检测并矫正跟踪器. LCT 方法可有效处理遮挡和目标移出视野的情况, 在保证精度的同时, 速度可达到 27 帧/s.

## 3 针对跟踪策略提出的相关滤波跟踪算法

### 3.1 边缘效应

由于标准的 DCF 方法在利用循环矩阵生成多样化样本时, 会不可避免地引发边缘效应进而导致过拟合, 所以 2015 年 Danelljan 等[25] 对 DCF 方法进行了改进, 提出了考虑空间信息的 SRDCF (Spatially regularized correlation filters) 方法, 在目标函数中将普通正则项改为空间惩罚正则项, 期望抑制离中心较远的特征对跟踪的影响. 通过此种方式依然采用 HOG 特征用于目标跟踪, 很好地解决了边缘效应问题, 并成为当年效果最好的跟踪方法之一, 但速度较慢, 4 帧/s 左右.

目标函数相对于式 (7) 仅第 2 项发生了改变, 即加入了空间惩罚正则项.

$$E = \min\left(\sum_{k=1}^{t}\alpha_k \left\|S_f(x_k) - y_k\right\|_2 + \sum_{l=1}^{d}\left\|\omega f^l\right\|_2\right) \tag{15}$$

其中, 惩罚权重 $\omega$ 由空间位置决定, 并满足高斯分布, 是决定相关滤波系数的一个重要参数.

图 2 是加入惩罚正则项前后相关滤波系数对比示意图. 从图 2 可以看出, 越接近边界的位置, 惩罚越大, 越接近中心的位置, 惩罚越小, 以此来更加突显目标, 同时减小边界对跟踪的影响.

最小化目标函数 (15) 的求解过程仍是在傅里叶域中进行, 由于加入了空间正则项, 破坏了 DCF 中的矩阵块对角结构, 故该方法迭代使用 Gauss-Seidel 方法进行在线学习优化, 进而得到新的相关滤波系数.

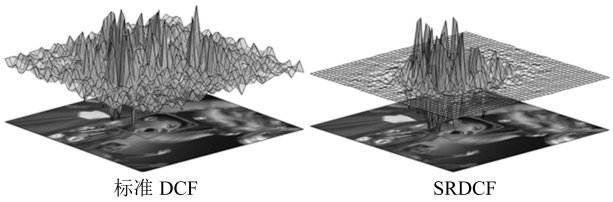


图 2 加入惩罚正则项前后相关滤波系数对比示意图

Fig. 2 Schematic diagram of correlation filtering coefficients before and after adding penalty regular

### 3.2 深度特征的采用

2015 年以前的相关滤波方法使用的特征主要集中于 HOG 梯度直方图、颜色直方图及边界等手工特征. 提取手工特征的方法是人为规定的, 仅适用于指定情况, 模型适应性较差. 而提取深度特征仿照的是人脑对信息的多层逐步递增的处理模式, 可以通过学习大量数据得到更有效、泛化能力更强的信息表达, 采用深度特征训练的模型适应性更强, 已成功应用于行为识别、图像分割等问题的求解中, 并获得了很好的效果.

DeepSRDCF (Convolutional features for spatially regularized correlation filter)[26] 方法分别在标准 DCF 和 SRDCF 框架下证明了深度特征的有效性, 详细论证了多种人工特征及各层深度特征对跟踪效果的影响. 大多数深度学习方法更偏重于使用预训练深度神经网络中全连接层提取的特征, 然而, 各卷积层提取的特征包含更多的结构和语义信息, 判别能力更强, 训练出的模型更适用于图像分类问题, 其中较浅层次的卷积层提取的特征包含更多的视觉信息, 因而更适用于跟踪问题.

该方法使用 ImageNet 数据集预训练的 CNN 深度神经网络来提取特征, 以待检测区域 RGB 图像作为输入, 逐层输出卷积特征, 通过实验对比, 在相关滤波跟踪框架中单独使用各层深度特征得到的结果. 可以看出, 使用 CNN 深度特征尤其是底层特征 (第一层特征) 解决跟踪问题的效果更好, 远远超过应用手工特征的相关滤波方法. DeepSRDCF 方法首次将深度特征引入视频跟踪领域, 虽然深度特征参数较多, 导致在提高跟踪精度的同时速度有所下降, 但仍具有里程碑式的意义.

虽然深度特征使得视频跟踪方法的效果得到很大提升, 但大多数方法采用的都是预训练好的深度特征. 2017 年 Wang 等[27] 提出 DCFNet (Discriminant correlation filters network) 方法, 认为预训练网络训练出来的深度特征或手工特征是独立于整个相关滤波跟踪过程的, 会对所训练模型的适应性造成影响, 故该方法自己构造出一种端到端的轻型网络框架, 通过将相关滤波看成一层网络加入到孪

生网络中, 训练出最适合相关滤波跟踪的特征, 在优化整个网络的过程中, 谨慎地反向传播误差并输出目标位置的概率热图, 进而实现在学习卷积特征的同时跟踪目标.

DCFNet 网络在 DCF 框架下构造级联特征提取过程, 最小化目标函数为

$$L(\theta) = \|g - \tilde{g}\|^2 + \gamma \|\theta\|^2 \tag{16}$$

其中, $g$ 为通过 DCF 得到的响应, $\tilde{g}$ 为期望响应, 该响应图在真实目标位置响应最大. 然后对目标函数进行一系列的求导过程, 来实现误差的反向传播, 进而训练整个网络, 详细推导过程见文献 [27].

DCFNet 将相关滤波器融入到训练网络中的想法新颖独特, 通过端到端的方法, 针对性地学习出适合相关滤波跟踪的特征, 所学出特征的性能能够与广泛使用的 HOG 特征相媲美, 在维持跟踪精度的同时, 大幅度提高跟踪速度, 跟踪速度可达到 89 帧/s.

### 3.3　训练样本的选择

追踪过程中通常会遇到样本损坏的问题, 例如错误的跟踪预测、扰动、局部或全遮挡都会导致样本不同程度的受损, 受损样本进入训练集, 必然会使模型泛化能力和判别能力下降. SRDCFdecon (Spatially regularized correlation filters with decontaminated training set)[28] 方法针对这一问题在 SRDCF 框架的基础上, 对训练样本集进行了改进, 通过评估样本的质量来动态管理训练集, 可有效增强模板的泛化能力. 该方法第一次提出联合优化外观模型参数和样本权重, 最小化损失函数, 其中目标函数如下:

$$E = \min(\sum_{k=1}^{t} \alpha \sum_{j=1}^{n^k} L(\theta; x_{jk}, y_{jk}) + \frac{1}{\mu} \sum_{k=1}^{t} \frac{\alpha_k^2}{\rho_k} + \lambda R(\theta)) \tag{17}$$

其中, $\alpha_k$ 为每个样本的权重, $L$ 为针对训练样本的损失函数, $\theta$ 是外观模型参数, $x$ 为训练样本, $y$ 为预期标签, 第 2 项控制变化速度与第 3 项一起都属于正则项. 值得一提的是, 在该目标函数中, 权重为一个连续的数值, 因为在目标发生轻微遮挡或微小变化时, 训练样本没有完全受损, 仍含有有用信息, 这时让权重连续化能够更准确地定义训练样本的性质.

仅凭前一帧的信息决定一个样本的重要性往往不够客观. 对于泛化能力较强的模板, 在更新样本时, 应该考虑到包括更早帧中的信息在内的所有有用信息. 该方法利用所有先前帧目标的信息, 将样本的权重规定为指定变化趋势, 即当前帧样本权重最大, 向前逐渐减小直至不. 具体方法为

$$\rho_k = \begin{cases} a, & k = 1, \cdots, t-K-1 \\ a(1-\eta)^{t-K-k}, & k = t-K, \cdots, t \end{cases} \tag{18}$$

其中, $\rho_k$ 为第 $k$ 帧的权重, 在每一帧迭代中, 都重新决定样本的权重, 进而纠正跟踪错误. 实验证明, 该方法可有效提高跟踪精度, 但速度较慢, 为 3 帧/s.

### 3.4　特征融合的应用

在验证了深度特征可有效提升模型适应性以后, Danelljan 等[29] 进一步改进了 DeepSRDCF 方法并提出 C-COT (Continuous convolution operators for visual tracking) 方法, 并获得了 2016 年 VOT 竞赛的冠军.

不同于使用单一分辨率特征的常用方法, Danelljan 等发现由不同卷积层得到的特征图分辨率大小不同, 即高层分辨率较小特征和低层分辨率较高特征, 因此这两种特征可在跟踪中发挥不同的作用. 该方法尝试使用插值运算, 将离散的特征图转化到连续空间域中进行计算, 有效结合不同层次的深度特征训练模型, 再次提高了模型的适应性. 其目标函数在基础 SRDCF 的目标函数式 (15) 的基础上改进为

$$E = \min\left(\sum_{j=1}^{m} \alpha_j \|S_f\{x_j\} - y_j\|^2 + \sum_{d=1}^{D} \|\omega f^d\|^2\right) \tag{19}$$

与基本公式不同的是, 式 (19) 中的 $S_f$ 为滤波模板与插值后样本做相关计算后的得分, 即使用从不同卷积层训练得到的滤波模板进行运算, 得到不同的置信图, 对所有的置信图进行加权求和, 得到最终的置信图. 最终置信图中最大值所在的位置即为要跟踪的目标所在的位置.

该方法对深度神经网络的各层及各层的不同组合提取的特征进行了逐一试验. 经测试, 融合第 0 层、第 1 层和最后一层提取的深度特征, 应用在视频跟踪问题中效果最好, 并且离散特征连续化的策略对跟踪算法效果的提升也起到了重要作用.

事实上, 直接在 DCF 跟踪器中融合多维特征导致表示模型参数增多, 例如 C-COT 就需连续更新 800 000 个参数, 模板泛化能力较差, 极易导致过拟合, 在增加计算复杂度的同时, 减慢了跟踪速度. DCF 跟踪器虽使用大量训练样本集, 但实际上可使用的样本数量有限, 通常的做法是丢弃最老的样本, 这很容易使跟踪结果拟合于最近的变化, 导致跟踪

偏移. DCF 跟踪器逐帧更新模型, 受孪生网络 (无需更新模型) 启发, 文献 [29] 认为逐帧更新为过渡更新模型, 反而会导致模板泛化能力下降, 敏感于目标的突然改变, 导致跟踪速度和算法鲁棒性的降低.

2017 年 ECO (Efficient convolution operators)[9] 方法在 C-COT 方法的基础上, 主要提出三个策略, 有效解决了上述问题: 1) 提出多项式卷积计算, 用 PCA 方法训练投影矩阵对融合特征进行降维, 仅考虑能量值最大的特征, 减少了参数个数; 2) 提出生成样本空间模型, 用混合高斯的方法合并样本集中最相似的两个训练样本, 减少了训练样本的个数, 减轻了重复计算相似样本带来的计算负担; 3) 抛弃逐帧更新模型的策略, 每隔固定帧数更新一次, 节省了无效更新浪费的时间.

ECO 方法通过上述策略训练的模型泛化能力较强, 在有效提高目标跟踪速度的同时, 精度也有所提高.

## 4 实验结果对比

本节主要从精度和速度两方面分析相关滤波系列跟踪算法; 首先展示文中所述方法在两大常用数据集 (OTB (Online object tracking: a benchmark)[30] 和 VOT (Visual object tracking)[31]) 上的测试结果, 然后着重讨论并分析该系列算法在不断完善的过程中在性能和速度方面发生的改变.

### 4.1 OTB 数据集上的方法结果对比

OTB 数据集是评价视频跟踪算法的重要公测数据集之一, 于 2013 年被首次提出, 包含 50 个涉及背景复杂、目标旋转、尺度变换、目标快速移动、目标变形、目标遮挡等多种跟踪难题的视频序列, 可对目标跟踪方法进行全面系统的评价.

该数据集主要有以下三种评价方式: 1) 一次性鲁棒评估 (One pass evaluation, OPE). 传统评估方式, 即从头到尾跑一遍视频序列, 以第 1 帧的真实目标位置作为初始位置. 2) 时间鲁棒评估 (Temporal robustness evaluation, TRE). 从不同的视频帧开始跟踪, 或随机跟踪视频序列的一个片段. 3) 空间鲁棒评估 (Spatial robustness evaluation, SRE). 以不同的目标框做初始开始跟踪, 通过对初始真实目标框采用中心转移、角度变换、尺度大小变换等不同方式得到不同的目标框.

SRE 对算法的鲁棒性要求最高, 为增强说服力, 本文在该数据集上采用 SRE 的方式对上述跟踪算法进行对比; 同时, 以成功率 $S$ 做为评价指标, 从不同角度分析实验结果, 具体做法为: 计算跟踪框和真实框的重叠率, 对重叠率大于阈值的帧进行计数, 由于使用指定阈值, 不能公平地对比不同的跟踪器, 成功率展示的是阈值为 0 ∼ 1 的成功帧数的比率, 比率曲线下覆盖面积的大小可用来对不同的跟踪器进行性能排序.

本文在 OTB 数据集上对比分析了上述提到的 ECO, CSK, KCF, CN, Staple, DSST, LCT, SRDCF, DeepSRDCF, CCOT, DCFNet, SRDCFdecon 等 12 种跟踪方法. 虽然上述方法目前仅从有限的角度对相关滤波跟踪算法进行了改进, 为方便分析总体性能及未来改进方向, 本文分别从尺度变换、目标旋转、低分辨率、光照变化、运动模糊、复杂背景、快速移动、目标形变、移除视野等角度综合展示了各种跟踪方法的 SRE 成功率排序, 对比图如图 3 所示.

### 4.2 VOT 数据集上的方法结果对比

VOT 数据集是评价目标跟踪方法的又一重要数据集, 包含 60 多个视频片段, 涵盖了尺度变换、相机移动、光照变化、运动变换、遮挡等多种跟踪问题, 虽然涉及到的跟踪问题与 OTB 略有重叠, 但评价方式不同, 该数据集通过期望平均覆盖率 (Expected average overlap, EAO) 评估跟踪算法的精确度和鲁棒性, 为 OTB 数据集提供了有效的补充评估.

上述基于相关滤波器的视频跟踪算法在 VOT 数据集上以 EAO 为评价标准的对比与排序如图 4 所示. 由于 CN, LCT, DCFNet 和 CSK 方法没有提供跟踪结果, 而 RAJSSC 仅提供了 VOT 结果, 故本文分别仅对其他 9 种方法的结果做了逐一比较, 另外比较了目前深度学习类跟踪方法中极具代表性的 TCNN[32] 方法.

### 4.3 关于跟踪性能的讨论与分析

从 OTB 数据集上的测试结果可以看出, 近年来基于相关滤波的各种跟踪方法效果被逐步改善, 虽然早期的 CSK 方法效果不尽如人意, 却是所有相关滤波跟踪方法的基础. 从结果对比图可以看出, 之后的算法每加入一种改进策略, 跟踪效果就提升一点: CN 方法加入了对颜色特征的改进策略, 不仅在复杂背景的情况下优于 CSK 方法, 在尺度变换、目标快速移动等情况下跟踪效果都有一定的改善, 但影响不是十分显著, 仅提高了 0.1 的成功率; 相比而言, 针对目标形变提出的 Staple 方法则带来了很大程度的改进, 在目标形变、复杂背景等情况下, 效果甚至强于使用深度特征的跟踪算法, 相对于最初的相关滤波方法, 提高了近 0.3 的成功率; DSST 方法针对目标尺度变化提出了相应的改进策略, 效果有所改善, 但不是很多; 旨在处理长时跟踪的 LCT 方法在目标旋转方面表现不错, 但在处理低分辨率目标时

效果不佳; 总体来说, Danelljan 等提出的系列方法均处于排名上游, 目前 ECO 算法不仅在相关滤波系列跟踪方法中综合效果最好, 在所有目标跟踪算法综合效果也是最好, 在大部分的视频跟踪问题中都能得到很好的效果, 平均成功率可达到 0.7 以上, 但在目标快速移动、光照变化、目标形变等视频难题中虽然速度占优势, 但效果不如 C-COT 或其他跟踪算法, 还有一定的改进空间.

VOT 数据集上的测试结果再次确认了各相关滤波算法性能的排名, 在整个 VOT 2016 竞赛中, 结

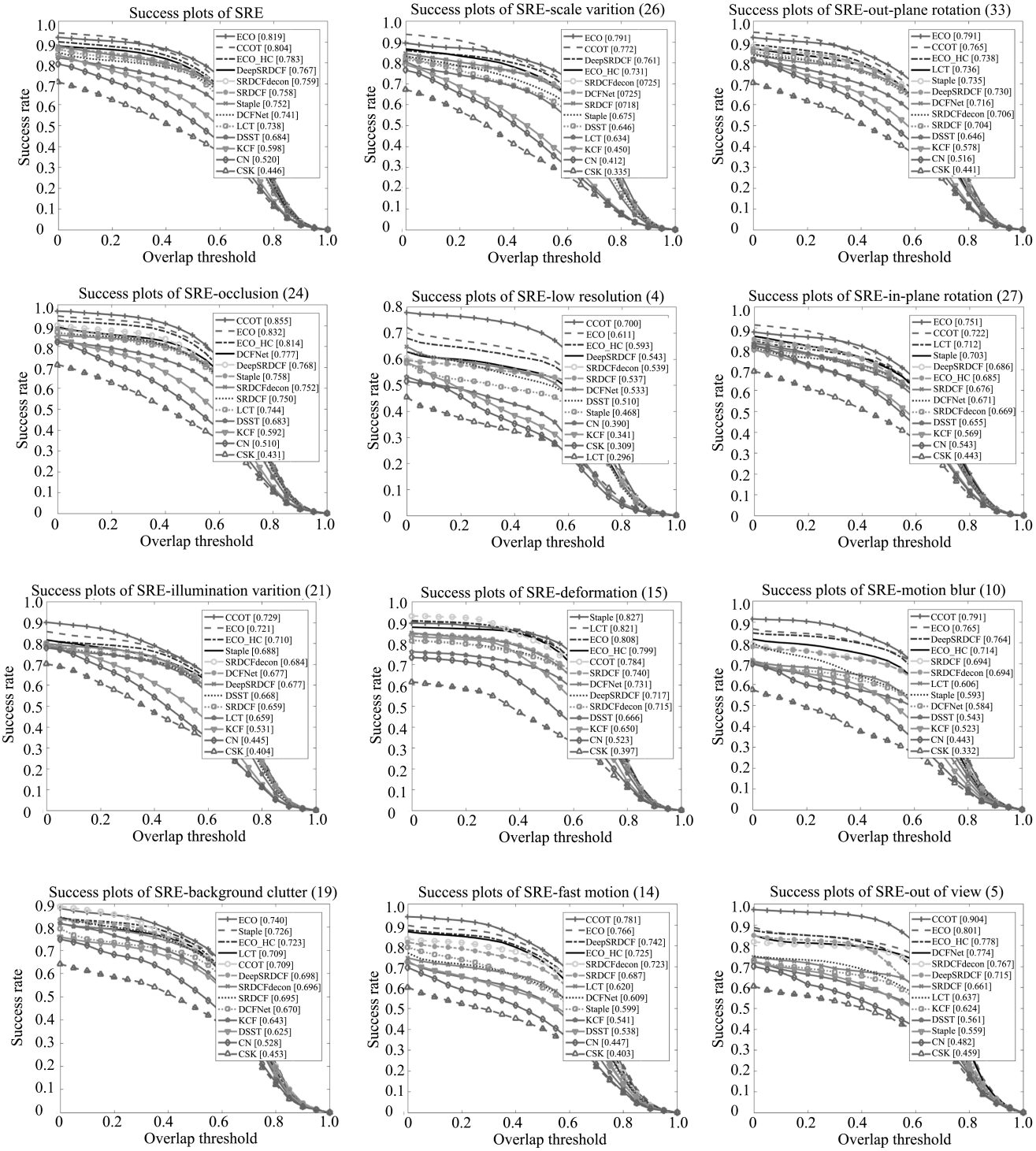


图 3　各种基于相关滤波跟踪方法成功率对比曲线图

Fig. 3　Various success ratio comparison curve based on correlation filter tracking methods

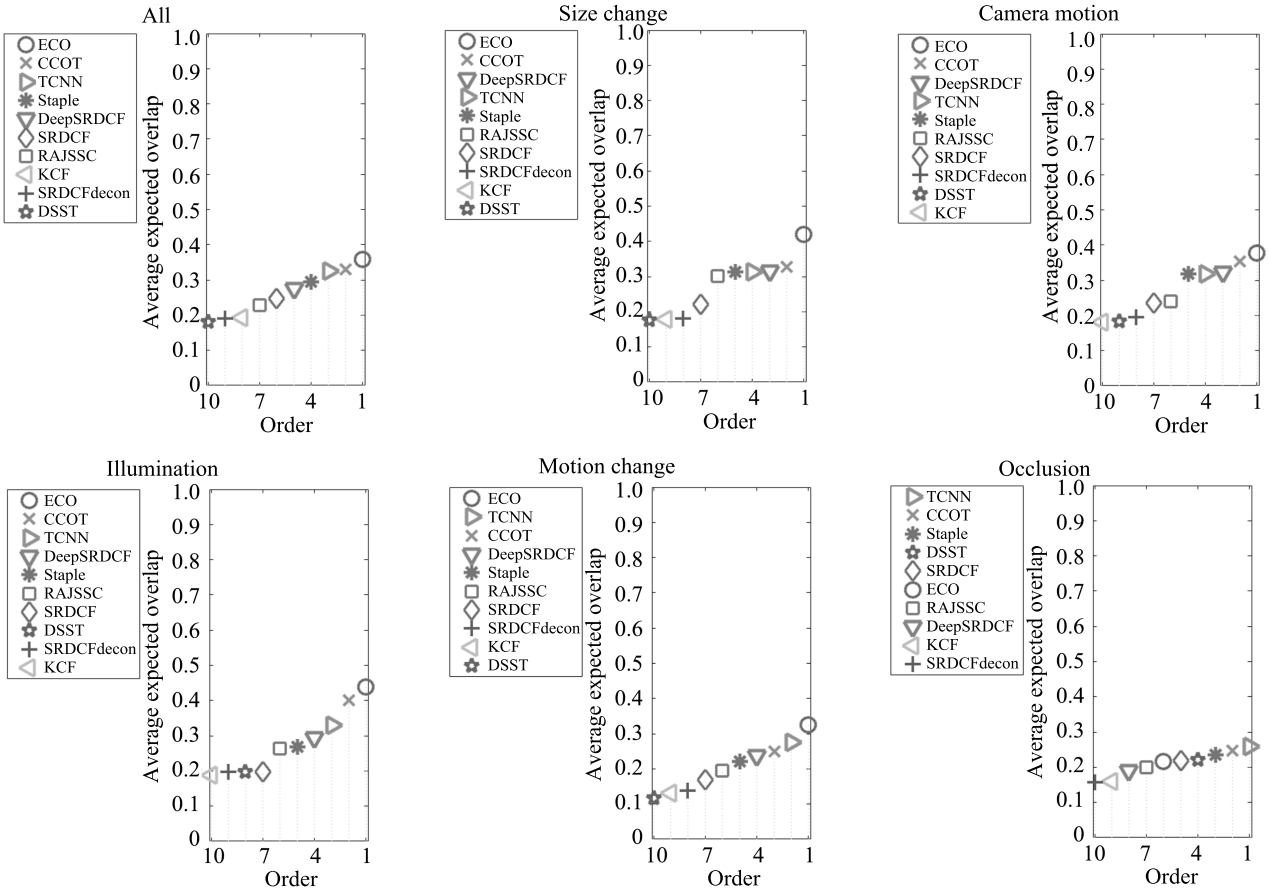


图 4 各种基于相关滤波跟踪方法的 EAO 等级图

Fig. 4 Various EAO level maps based on correlation filtering tracking methods

合相关滤波器的 CCOT 方法的 EAO 值为 0.331, 排名第 1, 优于仅结合深度学习的第 2 名方法 TCNN[32] (EAO 为 0.325). 如图 4 所示, 其他相关滤波跟踪方法排名也比较靠前. 深度特征的影响力毋庸置疑, 在相关滤波跟踪方法中直接使用深度特征的 DeepSRDCF 方法要明显优于大部分跟踪方法; 除此之外, 针对尺度变换的改进 RAJSSC 方法的效果要明显优于 DSST 方法, 可知模板匹配策略的有效性; 利用不同分辨率深度特征的 CCOT 使得跟踪算法效果又得到了很大的提升; 在此基础上, 采用优化训练集和模型更新次数策略的 ECO 方法的跟踪性能评价指标 EAO 被提升至 0.358, 优于 2016 最好方法 CCOT, 尽管在目标遮挡方面还存在一定的改进空间, 却是 2017 年度跟踪方法中效果最好的方法.

### 4.4 关于跟踪速度的讨论与分析

从基于相关滤波的跟踪算法提出至今, 随着跟踪策略的不断改进和完善, 算法的跟踪速度也发生了很大改变.

最初的 MOSSE 算法仅使用单通道的灰度特征, 相对简单, 跟踪精度不高但跟踪速度较快, 可达到 669 帧/s; 之后的算法在 MOSSE 的基础上逐渐演变得更复杂, 跟踪速度也相应降低, CSK 方法和 KCF 方法在 MOSSE 方法的基础上, 引入了循环矩阵和岭回归策略, 并使用多通道的彩色特征, 使跟踪精度得到了显著提高, 但跟踪速度分别降至 362 帧/s 和 172 帧/s; CN 方法和 Staple 方法对跟踪使用的手工特征分别做出了相应改进, 改善跟踪效果的同时速度分别降至 152 帧/s 和 80 帧/s; SRDCF 方法修改岭回归目标函数, 有效解决了边缘效应导致的过拟合现象, 却导致跟踪速度大幅下降至 4 帧/s; 随着将深度特征引入跟踪算法, 跟踪速度越来越慢, 演化到 C-COT 方法时跟踪速度已经降至 0.3 帧/s, 研究人员也终于对跟踪速度引起了关注, 2017 年 Danelljan 等提出的 ECO 算法旨在改善跟踪效果的同时提高跟踪速度, 简化特征并减少模型更新的次数, 使采用深度特征的跟踪算法速度提升至 6 帧/s, 采用手工特征的跟踪算法速度提升至 60 帧/s; 同年提出的 DCFNet 方法将相关滤波最为卷积网络的最后一层, 跟踪速度可达到 89 帧/s.

从相关滤波系列跟踪方法的演变过程可以看出,

这是一个先从简到繁, 又从繁到简的过程, 不变的是跟踪精度始终在持续提高, 大多数跟踪难题都已得到很好的解决.

## 5 结论

相关滤波器因在傅里叶域计算速度快、效果好等优点, 已被成功应用于各种计算机视觉问题中. 事实证明, 将相关滤波器引入跟踪方法可更好地应对跟踪问题中的各种挑战, 提高跟踪的准确性和鲁棒性, 进而实现长时在线跟踪. 综合目前基于相关滤波跟踪方法的发展现状, 本文认为该方法未来研究方向如下: 1) 分析各层深度特征的作用及重要性, 有效结合最优深度特征和人工特征来弥补彼此的不足, 进而提高模型对特征的表达能力; 2) 自适应更新相关滤波模型, 增强模型对目标变化的适应能力; 3) 优化相关滤波目标函数, 从根本上提升算法的检测性能.

## References


1 Zhang Tie, Ma Qiong-Xiong. Human object tracking algorithm for human-robot interaction. *Journal of Shanghai Jiao Tong University*, 2015, **49**(8): 1213−1219
(张铁, 马琼雄. 人机交互中的人体目标跟踪算法. 上海交通大学学报, 2015, **49**(8): 1213−1219)

2 Pantrigo J J, Hernández J, Sánchez A. Multiple and variable target visual tracking for video-surveillance applications. *Pattern Recognition Letters*, 2010, **31**(12): 1577−1590

3 Quan Yi-Ping, Yang Dao-Ye. Kalman filter vehicle tracking algorithm and behaviour analysis based on video detection. *Journal of Beijing University of Technology*, 2014, **40**(7): 1110−1113
(权义萍, 杨道业. 基于视频检测的卡尔曼滤波车辆跟踪算法及行为分析. 北京工业大学学报, 2014, **40**(7): 1110−1113)

4 Yang G, Zhao J S, Zheng C H, Fan Y. An approach based on mean shift and background difference for moving object tracking. In: Proceedings of the 6th International Conference on Wireless Communications Networking and Mobile Computing. Chengdu, China: IEEE, 2010. 1−4

5 Horn B K P, Schunck B G. Determining optical flow. *Artificial Intelligence*, 1981, **17**(1−3): 185−203

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

7 Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional siamese networks for object tracking. In: Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016. 850−865

8 Wang L J, Ouyang W L, Wang X G, Lu H C. STCT: sequentially training convolutional networks for visual tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1373−1381

9 Danelljan M, Bhat G, Khan F S, Felsberg M. ECO: efficient convolution operators for tracking. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 6931−6939

10 Hong Z B, Chen Z, Wang C H, Mei X, Prokhorov D, Tao D C. MUlti-store tracker (MUSTer): a cognitive psychology inspired approach to object tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 749−758

11 Valmadre J, Bertinetto L, Henriques J, Vedaldi A, Torr P H S. End-to-end representation learning for correlation filter based tracking. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017. 5000−5008

12 Zhang J M, Ma S G, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 188−203

13 Possegger H, Mauthner T, Bischof H. In defense of color-based model-free tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 2113−2120

14 Adankon M M, Cheriet M. Support vector machine. *Computer Science*, 2002, **1**(4): 1−28

15 Bolme D S, Beveridge J R, Draper B A, Lui Y M. Visual object tracking using adaptive correlation filters. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 2544−2550

16 Henriques J F, Caseiro R, Martins P, Batista J. Exploiting the circulant structure of tracking-by-detection with kernels. In: Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012. 702−715

17 Henriques J F, Caseiro R, Martins P, Batista J. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **37**(3): 583−596

18 Danelljan M, Khan F S, Felsberg M, van de Weijer J. Adaptive color attributes for real-time visual tracking. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014. 1090−1097

19 Bertinetto L, Valmadre J, Golodetz S, Miksik O, Torr P H S. Staple: complementary learners for real-time tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1401−1409

20 Danelljan M, Häger G, Khan F, Felsberg M. Accurate scale estimation for robust visual tracking. In: Proceedings of the 2014 British Machine Vision Conference. Michel, Canada: BMVA Press, 2014. 1−65

21 Danelljan M, Häger G, Khan F S, Felsberg M. Discriminative scale space tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(8): 1561−1575

22 Zhang M D, Xing J L, Gao J, Hu W M. Robust visual tracking using joint scale-spatial correlation filters. In: Proceedings of the 2015 IEEE International Conference on Image Processing. Quebec City, QC, Canada: IEEE, 2015. 1468−1472

23 Zhang M D, Xing J L, Gao J, Shi X C, Wang Q, Hu W M. Joint scale-spatial correlation tracking with adaptive rotation estimation. In: Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE, 2015. 595−603

24 Ma C, Yang X K, Zhang C Y, Yang M H. Long-term correlation tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 5388−5396

25 Danelljan M, Häger G, Khan F S, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 4310−4318

26 Danelljan M, Häger G, Khan F S, Felsberg M. Convolutional features for correlation filter based visual tracking. In: Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE, 2015. 621−629

27 Wang Q, Gao J, Xing J L, Zhang M D, Hu W M. DCFNet: discriminant correlation filters network for visual tracking. arXiv: 1704.04057. 2017.

28 Danelljan M, Häger G, Khan F S, Felsberg M. Adaptive decontamination of the training set: a unified formulation for discriminative visual tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and pattern Recognition. Las Vegas, NV, USA: IEEE, 2016. 1430−1438

29 Danelljan M, Robinson A, Khan F S, Felsberg M. Beyond correlation filters: learning continuous convolution operators for visual tracking. In: Proceedings of the 14th Computer Vision. Amsterdam, Netherlands: Springer, 2016. 472−488

30 Yi W, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 2411−2418

31 Kristan M, Leonardis A, Matas J, Felsberg M, Pflugfelder R, Ćuehovin L, et al. The visual object tracking vot2016 challenge results. In: Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016. 777−823

32 Nam H, Baek M, Han B. Modeling and propagating CNNs in a tree structure for visual tracking. arXiv: 1608.07242. 2016.

**刘巧元** 东北师范大学博士研究生. 2014 和 2016 年获得东北大学学士学位和硕士学位. 主要研究方向为视频目标跟踪, 模式识别.
E-mail: liuqy558@nenu.edu.cn
(**LIU Qiao-Yuan** Ph. D. candidate at Northeast Normal University. She received her bachelor and master degrees from Northeast University in 2014 and 2016, respectively. Her research interest covers visual tracking and pattern recognition.)

**王玉茹** 东北师范大学副教授. 2010 年获得哈尔滨工业大学博士学位. 主要研究方向为计算机视觉, 模式识别. 本文通信作者.
E-mail: wangyr915@nenu.edu.cn
(**WANG Yu-Ru** Associate professor at Northeast Normal University. She received her Ph. D. degree from Harbin Institute of Technology in 2010. Her research interest covers computer vision and pattern recognition. Corresponding author of this paper.)

**张金玲** 东北师范大学硕士研究生. 2016 年获得东北师范大学学士学位. 主要研究方向为计算机视觉, 模式识别.
E-mail: zhangjl575@nenu.edu.cn
(**ZHANG Jin-Ling** Master student at Northeast Normal University. She received her bachelor degree from Northeast Normal University in 2016. Her research interest covers computer vision and pattern recognition.)

**殷明浩** 东北师范大学教授. 2008 年获得吉林大学博士学位. 主要研究方向为自动规划, 自动推理, 语义网和近似推理.
E-mail: ymh@nenu.edu.cn
(**YIN Ming-Hao** Professor at Northeast Normal University. He received his Ph. D. degree from Jilin University in 2008. His research interest covers automated planning, automated reasoning, semantic web, and approximate reasoning.)