

# 课程论文题目

Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking (CVPR 2018)

## 摘要

判别相关滤波器 (DCF) 在视觉跟踪方面很有效, 但会受到不需要的边界效应的影响。已建议空间正则化 DCF (SRDCF) 通过对 DCF 系数实施空间惩罚来解决此问题, 这不可避免地以增加复杂性为代价来提高跟踪性能。为了解决在线更新问题, SRDCF 在多个训练图像上制定模型, 进一步增加了提高效率的困难。在这项工作中, 通过将时间正则化引入具有单个样本的 SRDCF, 我们展示了我们的时空正则化相关滤波器 (STRCF)。STRCF 公式不仅可以作为具有多个训练样本的 SRDCF 的合理近似, 而且在外观变化较大的情况下提供比 SRDCF 更稳健的外观模型。此外, 它可以通过乘法器的交替方向法 (ADMM) 有效地解决。通过结合时间和空间正则化, 我们的 STRCF 可以在不损失太多效率的情况下处理边界效应, 并在准确性和速度方面实现优于 SRDCF 的性能。与 SRDCF 相比, 具有手工制作特征的 STRCF 提供了 5 倍的加速, 并且在 OTB-2015 和 TempleColor 上分别获得了 5.4% 和 3.6% 的 AUC 分数。此外, 具有深度特征的 STRCF 在 OTB-2015 上的 AUC 得分也达到了 68.3%。

**关键词:** 判别相关滤波器; 时间正则化; 空间正则化;

## 1 引言

近年来, 鉴别相关滤波器 (DCF) 在视觉跟踪中的快速发展。受益于训练样本的周期性假设, 可以通过快速傅里叶变换 (FFT) 在频域中非常有效地学习 DCF。例如, 最早的基于 DCF 的跟踪器, 即 MOSSE<sup>[1]</sup> 的跟踪速度可以达到每秒 700 帧 (FPS)。随着特征表示<sup>[2][3]</sup>、非线性核<sup>[4]</sup>、尺度估计<sup>[5][6][7]</sup>、最大边缘分类器<sup>[8]</sup>、空间正则化<sup>[9]</sup>和连续卷积<sup>[10]</sup>的引入, 基于 DCF 的跟踪器得到了极大改进, 显著提高了最先进的跟踪精度。然而, 在没有任何额外成本的情况下, 无法获得这种性能改进。大多数排名靠前的跟踪器, 例如 SRDCF<sup>[9]</sup>和 C-COT<sup>[10]</sup>, 已经逐渐失去了早期基于 DCF 的跟踪器的特征速度和实时能力。例如, 使用手工制作的 HOG 功能的 SRDCF 的速度 $\sim 6$  FPS, 而基线 KCF<sup>[19]</sup> 的 FPS $\sim 170$  帧/秒。

为了更好地理解这个问题, 我们剖析了 SRDCF 中精度和速度之间的权衡。一般来说, SRDCF 的效率低下可归因于三个因素: (i) 尺度估计, (ii) 空间正则化, 以及 (iii) 在大型训练集上的公式化。图 1b 列出了 SRDCF 及其变体在两个流行基准 (包括 SRDCF) 上的跟踪速度和精度 ( $-M$ ) (即, 移除 (iii)), SRDCF( $-MS$ ) (即, 去除 (ii) 和 (iii)) 和 KCF (即, 除去 (i) (ii) (iii))。我们注意到, 当删除 (iii) 时, 采用线性插值作为在线模型更新的替代策略。从图 1b 中可以看出, 跟踪器仍然保持其实时能力 ( $\sim 33$  FPS)。但是, 随着在大训练集上进一步引入空间正则化和公式化, 跟踪速度显著降低。因此, 开发一种利用 (ii) 和 (iii) 而不损失效率的解决方案是很有价值的。

在本文中, 我们研究了在大型训练集上利用空间正则化和公式化而不损失效率的解决方案。一方面, SRDCF 的高复杂性主要来自于对多个训练图像的公式化。通过去除约束, 可以通过 ADMM 有效地求解具有单个图像的 SRDCF。由于 SRDCF 的凸性, ADMM 也可以保证收敛到全局最优。另一方

面，在 SRDCF 中，空间正则化被集成到多个训练图像的公式中，用于 DCF 学习和模型更新的耦合，这确实有利于跟踪精度。受在线被动攻击性（PA）学习<sup>[11]</sup>的启发，我们在具有单个图像的 SRDCF 中引入了时间正则化，从而产生了我们的时空正则化相关滤波器（STRCF）。STRCF 是多个训练图像上完整 SRDCF 公式的合理近似，也可用于同时 DCF 学习和模型更新。此外，ADMM 算法也可以直接用于求解 STRCF。因此，我们的 STRCF 将空间和时间正则化结合到 DCF 中，并可用于加速 SRDCF。

此外，作为在线 PA 算法的扩展，STRCF 还可以在外观变化显着的情况下提供比 SRDCF 更鲁棒的外观模型。与 SRDCF 相比，我们可以看到，随着时间正则化的引入，STRCF 对遮挡的表现更加稳健，同时能够很好地适应较大的外观变化。STRCF 不仅以实时跟踪速度（ $\sim 30\text{FPS}$ ）运行，而且在两个数据集上的平均 OP 比 SRDCF 的性能提升了 +5.7%。综上所述，STRCF 可以在所有数据集上实现比基线 SRDCF 的显著改进，并且以快 5 倍以上的跟踪速度运行。我们对几个基准进行了比较实验，包括 OTB-2015、Temple-Color 和 VOT-2016。与最先进的基于 CF 和 CNN 的跟踪器相比，STRCF 在准确性、稳健性和速度方面表现出色。

我们对几个基准进行了比较实验，包括 OTB-2015<sup>[12]</sup>、Temple Color<sup>[13]</sup>和 VOT-2016<sup>[14]</sup>。与最先进的基于 CF 和 CNN 的跟踪器相比，STRCF 在精度、鲁棒性和速度方面表现良好。本文的贡献如下：

(1) 通过将空间和时间正则化结合到 DCF 框架中来呈现 STRCF 模型。基于在线 PA，STRCF 不仅可以作为 SRDCF 公式在多个训练图像上的有理逼近，而且在外观变化较大的情况下提供比 SRDCF 更鲁棒的外观模型。

(2) 开发了一种 ADMM 算法来有效地解决 STRCF，其中每个子问题都有封闭形式的解决方案。而且我们的算法可以根据经验在极少数迭代内收敛。

(3) 我们具有手工制作功能的 STRCF 可以实时运行，通过跟踪精度实现了对 SRDCF 的显著改进。此外，与最先进的跟踪器相比，我们具有深层特征的 STRCF 表现出色。

## 2 相关工作

本节首先简要介绍 DCF 跟踪器，然后重点介绍与我们的 STRCF 最相关的大型训练集的空间正则化和公式化。

### 2.1 判别相关滤波器

使用 DCF 进行自适应跟踪始于 MOSSE，它在频域中学习样本较少的 CF。为了解决几个限制性问题，对这一流行的跟踪器进行了显著改进。例如，Henriques 等人通过内核技巧学习内核化 CF（KCF）。<sup>[15]</sup>中还研究了 MOSSE 的多通道版本。更具区别性的特征被广泛使用，例如 HOG<sup>[16]</sup>、颜色名称（CN）<sup>[2]</sup>和深度 CNN 特征<sup>[17]</sup>。为了应对尺寸变化和遮挡，进一步研究了几种尺度自适应<sup>[5][6][7]</sup>和基于部分的跟踪器<sup>[18][19]</sup>。此外，还开发了长期跟踪<sup>[20]</sup>、连续卷积<sup>[10]</sup>和基于粒子滤波的方法<sup>[21]</sup>，以提高跟踪精度和鲁棒性。由于空间的限制，这里我们只回顾了空间正则化和大训练集公式化的方法，这些方法与我们的算法非常接近。

### 2.2 空间正则化

基于 DCF 的跟踪器中的循环移位样本总是在边界位置上遭受周期性重复，从而显著降低了跟踪性能。已经提出了几种空间正则化方法来减轻不期望的边界效应。Galoogahi 等人使用包含目标区域

的固定掩模矩阵预乘图像块，然后通过 ADMM 解决约束优化问题。然而，它们的方法只能应用于单信道 DCF。Danelljan 等人提出了一个空间正则化项，根据 DCF 系数的空间位置对其进行惩罚，并建议使用 Gauss-Seidel 算法来求解得到的正规方程。工作<sup>[22]</sup>也使用了类似的空间正则化项，但使用多方向 RNN 预测空间正则化矩阵，以识别可靠的分量。然而，这两种方法无法利用学习中的循环结构，导致较高的计算成本。最近，Galoogahi 等人<sup>[23]</sup>将方法扩展到多个信道，并进一步加快跟踪器的实时速度。与这些方法相比，我们的 STRCF 具有几个优点：（1）虽然 STRCF 在多个训练样本上用作的近似值，但使用所提出的 ADMM 算法可以更有效地解决该问题。（2）随着时间正则化的引入，STRRC 可以学习比<sup>[9][23]</sup>更稳健的外观模型，从而获得更好的跟踪性能。

### 2.3 大型训练集的公司

视觉跟踪中最关键的挑战之一是在外观变化较大的情况下学习并保持鲁棒和快速的外观模型。为此，MOSSE 通过从历史跟踪结果中学习具有多个训练样本的 CF 来实现同时的 DCF 学习和模型更新。在<sup>[24][25][10][15]</sup>中也可以找到将大型训练集纳入公式的类似策略。在实践中，可以通过考虑不同时间点的样本来学习稳健的 CF。然而，这导致了以更高的计算负担为代价的优异性能。与这些方法相比，KCF 及其变体<sup>[26][5]</sup>将 DCF 学习和模型更新解耦，并进一步利用循环结构以实现高效率。因此，具有 HOG 功能的 KCF 可以在单个 CPU 上以超过 150 FPS 的速度运行。在这项工作之后，还存在几种启发式方法<sup>[19][27]</sup>来解决天真模型更新问题。然而，这些方法获得的性能不如具有大训练集的基于 DCF 的跟踪器。与这些跟踪器相比，STRCF 不仅可以通过避免部署大型训练集来有效地解决，而且还可以通过引入时间正则化来同时进行 DCF 学习和模型更新。

## 3 本文方法

### 3.1 回顾 SRDCF

用

$$\mathcal{D} = \{(\mathbf{x}_k \mathbf{y}_k)\}_{k=1}^T \quad (1)$$

表示多个图像的训练集。每个样本

$$\mathbf{x}_k = [\mathbf{x}_k^1, \dots, \mathbf{x}_k^D] \quad (2)$$

由大小为  $M \times N$  的  $D$  个特征图组成。其中  $y_k$  是预定义的高斯形状标签。SRDCF 是通过最小化以下目标函数来得到的，

$$\arg \min_{\mathbf{f}} \sum_{k=1}^T \alpha_k \left\| \sum_{d=1}^D \mathbf{x}_k^d * \mathbf{f}^d - \mathbf{y}_k \right\|^2 + \sum_{d=1}^D \|\mathbf{w} \cdot \mathbf{f}^d\|^2 \quad (3)$$

其中  $\cdot$  表示 Hadamard 乘积， $*$  表示卷积算子， $\mathbf{w}$  和  $\mathbf{f}$  分别是空间正则化矩阵和相关滤波器。 $\alpha_k$  表示每个样本  $x_k$  的权重，并设置为更强调最近的样本。Danelljan 等人使用 Gauss-Seidel 方法迭代更新滤波器  $\mathbf{f}$ 。

然而，尽管 SRDCF 可以有效抑制不利的边界效应，但由于以下两个原因，它也增加了计算负担：

(i) 利用循环矩阵结构的失败。为了学习稳健的相关滤波器  $\mathbf{f}$ ，DCF 跟踪器合并了几个历史样本用于训练。然而，与其他仅使用当前帧样本学习的基于 CF 的跟踪器不同，多幅图像上的公式打破了循环矩阵结构，导致计算负担高。至于 SRDCF，由于空间正则化项，优化变得更加困难。

(ii) 大型线性方程组和高斯-赛德尔求解器。Eqn(2) 导致  $DMN \times DMN$  大稀疏线性方程组。虽然建议使用 Gauss-Seidel 方法来求解 Eqn. (1) 利用稀疏矩阵的特性，仍然存在较高的计算复杂度。此外，由于 Gauss-Seidel 方法的收敛速度较慢，SRDCF 跟踪器也需要较长的启动时间来学习第一帧中的判别相关滤波器。

多幅图像的空间正则化和公式化都会破坏循环矩阵结构。幸运的是，可以规避这两个问题以提高跟踪速度。通过引入时间正则化，可以将多幅图像上的公式放宽到单幅图像上的 STRCF 模型。此外，空间正则化的引入可以通过利用 ADMM 有效求解的等效重构来解决。

### 3.2 STRCF

在在线分类中，当每一轮出现一个新实例时，算法首先预测其标签，然后根据新实例标签对更新分类器。一方面，学习算法应该是被动的，以使更新的分类器与之前的分类器相似。另一方面，学习算法应该是积极的，以确保新的实例被正确分类。因此，Crammer 等人通过引入时间正则化，提出了一种在线被动攻击 (PA) 算法，并导出了 PA 的累积损失与最佳固定预测器之间的界限。

受 PA 的启发，我们引入了时间正则化项，从而产生了我们的时空正则化 CF (STRCF) 模型，

$$\arg \min_{\mathbf{f}} \frac{1}{2} \left\| \sum_{d=1}^D \mathbf{x}_t^d * \mathbf{f}^d - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w} \cdot \mathbf{f}^d\|^2 + \frac{\mu}{2} \|\mathbf{f} - \mathbf{f}_{t-1}\|^2, \quad (4)$$

其中  $\mathbf{f}_{t-1}$  是第  $t-1$  帧得到的滤波器，Eqn(4) 中第二项是空间正则项，第三项是时间正则项。

从两个方面来看，STRCF 也可以看作是在线 PA 的扩展：(i) STRCF 不是分类，而是线性回归的在线学习；(ii) STRCF 中的样本不是按实例更新，而是在每一轮中以批处理级别（即图像的所有移位版本）出现。因此，STRCF 自然地继承了在线 PA 在主动和被动模型学习之间自适应平衡的优点，从而在外观变化较大的情况下产生更鲁棒的模型。在图 2 中，我们将 STRCF 与 SRDCF 在序列 Lemming 上进行比较，以突出它们在 CF 模型学习上的关系。从中我们可以做出以下观察：(i) 与 SRDCF 类似，STRCF 也通过引入时间正则化器实现了同步的 DCF 学习和模型更新，因此可以作为具有多个训练样本的 SRDCF 的有理逼近；(ii) 在遮挡的情况下，虽然 SRDCF 对最近损坏的样本过度拟合，但 STRCF 可以通过被动更新 CF 使其与之前的样本保持接近来缓解这种情况。

### 3.3 STRCF 的优化算法

对于 Eqn(4)，我们首先引入辅助变量  $\mathbf{g}$  和步长参数  $\gamma$ ，其中  $\mathbf{f} = \mathbf{g}$  引，然后得到 Eqn(4) 的增广拉格朗日形式为

$$\mathcal{L}(\mathbf{w}, \mathbf{g}, \mathbf{s}) = \frac{1}{2} \left\| \sum_{d=1}^D \mathbf{x}_t^d * \mathbf{f}^d - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w} \cdot \mathbf{g}^d\|^2 + \sum_{d=1}^D (\mathbf{f}^d - \mathbf{g}^d)^T \mathbf{s}^d + \frac{\gamma}{2} \sum_{d=1}^D \|\mathbf{f}^d - \mathbf{g}^d\|^2 + \frac{\mu}{2} \|\mathbf{f} - \mathbf{f}_{t-1}\|^2, \quad (5)$$

其中  $\mathbf{s}$  和  $\mu$  分别为拉格朗日算子和惩罚因子。引入  $\mathbf{h} = \frac{\mathbf{s}}{\gamma}$ 。

$$\mathcal{L}(\mathbf{w}, \mathbf{g}, \mathbf{h}) = \frac{1}{2} \left\| \sum_{d=1}^D \mathbf{x}_t^d * \mathbf{f}^d - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w} \cdot \mathbf{g}^d\|^2 + \frac{\gamma}{2} \sum_{d=1}^D \|\mathbf{f}^d - \mathbf{g}^d + \mathbf{h}^d\|^2 + \frac{\mu}{2} \|\mathbf{f} - \mathbf{f}_{t-1}\|^2 \quad (6)$$

然后采用 ADMM 算法，交替求解以下子问题，

$$\begin{cases} \mathbf{f}^{(i+1)} = \underset{\mathbf{f}}{\operatorname{argmin}} \left\| \sum_{d=1}^D \mathbf{x}_t^d * \mathbf{f}^d - \mathbf{y} \right\|^2 + \gamma \|\mathbf{f} - \mathbf{g} + \mathbf{h}\|^2 + \mu \|\mathbf{f} - \mathbf{f}_{t-1}\|^2 \\ \mathbf{g}^{(i+1)} = \underset{\mathbf{g}}{\operatorname{argmin}} \sum_{d=1}^D \|\mathbf{w} \cdot \mathbf{g}^d\|^2 + \gamma \|\mathbf{f} - \mathbf{g} + \mathbf{h}\|^2 \\ \mathbf{h}^{(i+1)} = \mathbf{h}^{(i)} + \mathbf{f}^{(i+1)} - \mathbf{g}^{(i+1)}. \end{cases} \quad (7)$$

子问题 f:

$$\mathcal{V}_j(\hat{\mathbf{f}}) = \frac{1}{\mu + \gamma} \left( I - \frac{\mathcal{V}_j(\hat{\mathbf{x}}_t) \mathcal{V}_j(\hat{\mathbf{x}}_t)^\top}{\mu + \gamma + \mathcal{V}_j(\hat{\mathbf{x}}_t)^\top \mathcal{V}_j(\hat{\mathbf{x}}_t)} \right) \mathbf{q} \quad (8)$$

其中

$$\mathbf{q} = \mathcal{V}_j(\hat{\mathbf{x}}_t) \hat{y}_j + \gamma \mathcal{V}_j(\hat{\mathbf{g}}) - \gamma \mathcal{V}_j(\hat{\mathbf{h}}) + \mu \mathcal{V}_j(\hat{\mathbf{f}}_{t-1}) \quad (9)$$

子问题 g:

$$\mathbf{g} = (\mathbf{W}^\top \mathbf{W} + \gamma I)^{-1} (\gamma \mathbf{f} + \gamma \mathbf{h}) \quad (10)$$

其中 W 表示由 D 个对角矩阵  $\operatorname{Diag}(\mathbf{w})$  串联而成的  $\text{DMN} \times \text{DMN}$  对角矩阵。

更新步长参数:

$$\gamma^{(i+1)} = \min(\gamma^{\max}, \rho \gamma^{(i)}) \quad (11)$$

## 4 复现细节

### 4.1 与已有开源代码对比

复现过程参考了作者提供的源码，链接如下：<https://github.com/lifeng9472/STRCF>

与原论文相比，本文复现基于其目标函数进行了改进，即对时间正则项和空间正则项进行了改进。

以下是改进后的迭代公式:

子问题 f:

$$\mathcal{V}_j(\hat{\mathbf{f}}) = \frac{1}{\mu + \gamma} \left( I - \frac{\mathcal{V}_j(\hat{\mathbf{x}}_t) \mathcal{V}_j(\hat{\mathbf{x}}_t)^\top}{\mu + \gamma + \mathcal{V}_j(\hat{\mathbf{x}}_t)^\top \mathcal{V}_j(\hat{\mathbf{x}}_t)} \right) \mathbf{q} \quad (12)$$

其中

$$\mathbf{q} = \mathcal{V}_j(\hat{\mathbf{x}}_t) \hat{y}_j + \gamma \mathcal{V}_j(\hat{\mathbf{g}}) - \gamma \mathcal{V}_j(\hat{\mathbf{h}}) + \mu \mathcal{V}_j(\hat{\mathbf{f}}_{t-1}) \quad (13)$$

子问题 g:

$$\mathbf{g}^j = \max \left( 0, 1 - \frac{\lambda_1}{\mu \|k^j\|_2} \right) k^j, \text{ 其中 } k^j = \mathbf{f} + \frac{\eta^j}{\mu} \quad (14)$$

子问题 h:

$$\mathbf{H} = \mathbf{H} + \mu (\mathbf{f} - \mathbf{g}) \quad (15)$$

### 4.2 实验环境搭建

按照 STRCF 中的设置，我们裁剪以目标为中心的正方形区域，其中该区域的边长为  $\sqrt{5WH}$  其中 (W 和 H 分别代表目标的宽度和高度)。然后我们提取图像区域的 HOG、CN 和 CNN 特征。这些特征由余弦窗口进一步加权以减少边界不连续性。至于 ADMM 算法，我们在实验中设置了 Eqn(7) 中的所有超参数。其中  $\mu = 16$ 。初始步长参数  $\gamma^{(0)}$ 、最大  $\gamma^{\max}$  和比例因子  $\rho$  分别设置为 10、100 和 1.2。我们的 STRCF 使用 Matlab 2017a 实现，所有实验都在配备 Intel i9-12900H CPU、32GB RAM 和单个

NVIDIA GTX 3080Ti GPU 的 PC 上运行。

4.3 创新点

在模型更新阶段，STRCF 只使用了当前帧的样本和上一帧的学习得到的滤波器并强调最近的样本，因此可能会过拟合最近不准确的样本，并因为遮挡导致跟踪失败。且 STRCF 在对快速移动或受到长时间遮挡的目标进行跟踪时的效果不够理想。故对其对滤波器的约束进行更改，并将第  $t-2$  帧学习得到的滤波器加入到模型更新过程中，以形成更鲁棒的模型。

其中更新后的空间正则项已在 4.1 中给出。

更新后的时间正则项为

$$\frac{\mu}{2} \|\mathbf{f} - \alpha \mathbf{f}_{t-1} - (1 - \alpha) \mathbf{f}_{t-2}\|^2$$

(16)

其中  $0 < \alpha < 1$ 。

5 实验结果分析

本部分对实验所得结果进行分析，详细对实验内容进行说明，实验结果进行描述并分析。

本部分展示了 STRCF 源码与进行创新后的在 OTB 数据集上的部分实验效果对比，如下：

	STRCF	STRCF	STRCF_OURS	STRCF_OURS
		op		op
Basketball	FPS: 32.8217	36.979%	FPS: 30.2236	76.328%
Biker	FPS: 49.1034	37.659%	FPS: 48.4814	37.726%
Bird1	FPS: 42.9777	19.153%	FPS: 41.7778	25.093%
BlurBody	FPS: 21.9549	68.748%	FPS: 19.9582	69.247%
BlurCar2	FPS: 25.228	87.163%	FPS: 23.9577	87.310%
BlurFace	FPS: 26.3824	85.898%	FPS: 18.7564	85.994%
Box	FPS: 27.763	74.328%	FPS: 25.7068	74.825%
Carl	FPS: 67.5524	85.243%	FPS: 55.6328	80.317%
CarDark	FPS: 50.4075	84.006%	FPS: 42.8755	85.181%
CarScale	FPS: 38.5105	69.822%	FPS: 33.8839	73.658%
Couple	FPS: 32.4876	67.925%	FPS: 33.8888	69.354%
Deer	FPS: 29.628	79.343%	FPS: 30.1114	80.282%
Diving	FPS: 31.4312	21.949%	FPS: 32.4698	27.309%
DragonBaby	FPS: 29.7498	50.611%	FPS: 31.0313	66.372%
Football	FPS: 66.6882	68.535%	FPS: 66.9491	71.797%
Human3	FPS: 31.0372	69.012%	FPS: 29.1148	69.251%
Boy	FPS: 31.5256	82.171%	FPS: 35.5482	82.503%
Car24	FPS: 43.0043	58.733%	FPS: 42.7917	67.762%
Coke	FPS: 23.0073	55.916%	FPS: 31.1218	57.683%
Coupon	FPS: 53.921	35.765%	FPS: 65.2358	35.984%
Dancer	FPS: 54.8647	72.085%	FPS: 64.8305	73.841%
Dudek	FPS: 47.8624	80.869%	FPS: 51.6904	81.044%
Fish	FPS: 56.3194	83.874%	FPS: 64.2818	84.294%
FleetFace	FPS: 41.5667	69.206%	FPS: 55.8795	72.095%
Girl	FPS: 31.5042	70.181%	FPS: 33.5014	73.505%
Gym	FPS: 31.8087	41.137%	FPS: 33.2404	46.235%
Jumping	FPS: 85.9759	67.534%	FPS: 87.894	67.808%
KiteSurf	FPS: 49.946	35.941%	FPS: 52.443	35.998%
Lemming	FPS: 27.5097	74.551%	FPS: 28.6493	75.788%

图 1: 实验结果示意

从上面数据可以看出，创新后的 STRCF 显然是优于原文提出的 STRCF 的，尤其是在处理长视频和遮挡视频的效果远优于 STRCF。本方法在遵循原 STRCF 的滤波器更新策略的同时，对空间正则项与时间正则项进行改进。从视频序列 Basketball 和 Bird1 的 op 分别达到：76.328% 和 25.093% 可以看出在模型更新过程中在时间正则项中引入第  $t-2$  帧的滤波器，能够产生更鲁棒滤波器模型。

6 总结与展望

这次的复现工作使我获益良多，我对基于相关滤波的目标跟踪领域有了进一步的认识，并且首次尝试了对顶会论文进行改进，这是一次很宝贵的体验。从上述的实验结果分析可见，引入更稳健的空间约束项和第  $t-2$  帧的滤波器进行模型更新是能够获得更鲁棒的跟踪器的。但是是否可以引入更多的时间信息，以及如何权衡当前输入与以往的时间信息仍然是一个值得研究的课题。并且在特征提取的过程中，本文虽提取了多种特征，但是如何判断特征是否是有效的，对跟踪定位和模型更新是有益的，以及如何权衡有益的特征的权重仍是值得探讨的问题。

## 参考文献

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//2010 IEEE computer society conference on computer vision and pattern recognition. 2010: 2544-2550.
- [2] DANELLJAN M, SHAHBAZ KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1090-1097.
- [3] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking[C]//Proceedings of the IEEE international conference on computer vision. 2015: 3074-3082.
- [4] 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, 2014, 37(3): 583-596.
- [5] DANELLJAN M, HÄGER G, KHAN F S, et al. Discriminative scale space tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(8): 1561-1575.
- [6] LI F, YAO Y, LI P, et al. Integrating boundary and center correlation filters for visual tracking with aspect ratio variation[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017: 2001-2009.
- [7] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]//European conference on computer vision. 2014: 254-265.
- [8] ZUO W, WU X, LIN L, et al. Learning support correlation filters for visual tracking[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 41(5): 1158-1172.
- [9] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Learning spatially regularized correlation filters for visual tracking[C]//Proceedings of the IEEE international conference on computer vision. 2015: 4310-4318.
- [10] DANELLJAN M, ROBINSON A, SHAHBAZ KHAN F, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking[C]//European conference on computer vision. 2016: 472-488.
- [11] CRAMMER K, DEKEL O, KESHET J, et al. Online passive aggressive algorithms[J]., 2006.
- [12] ZHANG J, MA S, SCLAROFF S. MEEM: robust tracking via multiple experts using entropy minimization[C]//European conference on computer vision. 2014: 188-203.
- [13] LIANG P, BLASCH E, LING H. Encoding color information for visual tracking: Algorithms and benchmark[J]. IEEE transactions on image processing, 2015, 24(12): 5630-5644.
- [14] GÜNDOĞDU E, ALATAN A A. The visual object tracking VOT2016 challenge results[J]., 2016.

- [15] GALOOGAHI H K, SIM T, LUCEY S. Multi-channel correlation filters[C]//Proceedings of the IEEE international conference on computer vision. 2013: 3072-3079.
- [16] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05): vol. 1. 2005: 886-893.
- [17] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4303-4311.
- [18] LIU S, ZHANG T, CAO X, et al. Structural correlation filter for robust visual tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 4312-4320.
- [19] LIU T, WANG G, YANG Q. Real-time part-based visual tracking via adaptive correlation filters[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 4902-4912.
- [20] MA C, YANG X, ZHANG C, et al. Long-term correlation tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 5388-5396.
- [21] ZHANG T, XU C, YANG M H. Multi-task correlation particle filter for robust object tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4335-4343.
- [22] CUI Z, XIAO S, FENG J, et al. Recurrently target-attending tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1449-1458.
- [23] KIANI GALOOGAHI H, FAGG A, LUCEY S. Learning background-aware correlation filters for visual tracking[C]//Proceedings of the IEEE international conference on computer vision. 2017: 1135-1143.
- [24] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1430-1438.
- [25] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE international conference on computer vision workshops. 2015: 58-66.
- [26] BIBI A, MUELLER M, GHANEM B. Target response adaptation for correlation filter tracking[C]//European conference on computer vision. 2016: 419-433.
- [27] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4021-4029.